# Flash-flood forecasting in two Spanish Mediterranean catchments: a comparison of distinct hydrometeorological ensemble prediction strategies

VINCENDON Béatrice[1] and AMENGUAL Arnau[2]

[1]CNRM (Météo-France, CNRS),42 av. Coriolis, 31057 Toulouse Cedex, Toulouse, France
[2]Grup de Meteorologia, Dep. De Física, Universitat de les Illes Balears. Palma, Mallorca, Spain

*Correspondence to:* Vincendon B.(beatrice.vincendon@meteo.fr)

**Abstract.** Hydrological Ensemble Prediction Systems (HEPSs) are becoming more and more popular methods to deal with the meteorological and hydrological uncertainties that affect discharge forecasts. These uncertainties are particularly difficult to handle when dealing with Mediterranean flash-flood forecasting as many hydrological and meteorological factors take place and precipitation comes from small scale convective systems. In this work, the performances of distinct HEPS are

5   compared for two heavy precipitation events that affected two different semi-arid Spanish Mediterranean catchments : the cases of the 03 November 2011 on the Llobregat River in Catalonia, and the 28 September 2012 on the Guadalentín River near in Murcia. The latter case corresponds to the eighth Intense Observing Period (IOP8) of HYMEX field campaign. The uncertainty on quantitative precipitation forecasting is sampled by using two different convection-permitting meteorological ensemble generation strategies. The first EPS strategy consists in dynamically downscaling the ECMWF-EPS directly by means

10   of the WRF model, whereas the second is based on the AROME-WMED model. Its deterministic QPFs are perturbed based on a previous rainfall forecast error climatology and by using the probability density functions of the errors, in term of total amounts and location of the heaviest rainfalls. The population of both ensembles is of 50 members, which are used to drive the semi-distributed and conceptual HEC-HMS and the fully distributed and physically-based ISBA-TOP hydrological models. For each HEPS, the performance is assessed in term of the quantitative discharge forecasts. The results point out the benefits

15   of using (i) a hydrological model when evaluating highly-variable and convective-driven precipitation fields and (ii) an EPS to better encompass these uncertainties arising from the different elements of the HEPS. Issues about the optimal number of ensemble members and impact of the ensemble forecasting lead time are addressed for optimal flash-flood forecasting purposes as well.

Natural Hazards
and Earth System
Sciences
Open Access
EGU
Discussions

# 1 Introduction

Flash floods are among the worst hazards worldwide (Doocy et al., 2013). These hydrometeorological episodes can result in substantial human, social and economic losses. The HYdrological cycle in the Mediterranean EXperiment (HyMeX, http://www.hymex.or programme fosters a better understanding, quantification and modelling of precipitation and flood extremes in the Mediter-

5 ranean (Drobinski et al., 2014). The Spanish Mediterranean region is affected recurrently by extreme precipitations resulting in flash-flooding, mainly during the end of the warm season. The early-autumn intrusion of upper-level cold air masses in the Western Mediterranean region and its comparatively large sea surface temperature boost the convective available potential energy. Besides, the prominent orography of Mediterranean Spain results in the lifting of the warm and moist sea and the subsequent generation of deep moist convection (Fig. 1).

10 Flash floods result from the persistence for several hours of high precipitation rates over specific hydrographic catchments. In the Spanish Mediterranean region, quasi-stationary precipitation is often a consequence of long-lived Mesoscale Convective Systems (MCSs) which remain anchored by the prominent orography (Romero et al., 2000; Ducrocq et al., 2014). In addition, many semi-arid river basins are small-to-medium sized, highly urbanized and contain coastal steep streams. All these factors shorten even more the hydrological response times. As many small rivers are ephemeral, large and rapid flows carrying exten-

15 sive quantities of debris can exacerbate extensive flood damage. Therefore, the development and evaluation of state-of-the-art hydrometeorological tools is an issue of major relevance. These tools can contribute to a better understanding and forecasting of flash floods in order to implement more reliable forecasting and warning systems over the Mediterranean Spain.

For this purpose, we have selected the 03 November 2011 and the 28 September 2012 flash floods in the Llobregat and Guadalentín basins, respectively. The Guadalentín and Llobregat basins are an archetype of Spanish Mediterranean catchments

20 that are recurrently affected by flooding (for further details, see Amengual et al. (2007, 2009, 2015) ). The Guadalentín River basin is located in Murcia while the Llobregat watershed is located in Catalonia, south and north-eastern Spain respectively (Fig. 1). For both catchments, we have implemented two event-based models : (i) the semi-distributed and conceptual-based HEC-HMS model (USACE-HEC, 2000) and; (ii) the fully-distributed and physically-based ISBA-TOP model (Bouilloud et al., 2010; Vincendon et al., 2010, 2016).

25 The first step of the present work is to explore the impact on the simulated flood hydrographs of these distinct hydrological model set-ups that account for different levels of variability in the rainfall fields and basin properties. The use of distinct hydrological models for both extreme events allows investigating more comprehensively the potential impacts of the aforementioned modelling issues in real-time.

Regarding the aim of the present work, the use of two models with different formulations, especially for the modelling of the

30 soil infiltration mechanism, may result beneficial to better understand and describe the rainfall-runoff transformation processes, according to the nature of the rainfall episode which occur over the catchment in question. Specifically, we use two distinct model structures and physical parameterizations to simulate the complex rainfall-runoff transformation of intense precipitations over semi-arid basins. In fact, the different soil infiltration and routing schemes are determining factors modulating peak discharges, timings and runoff volumes. As a matter of fact, the characteristics of the rainfall event (i.e. spatial-temporal distri-

Natural Hazards
and Earth System
Sciences
Discussions

bution and intensity) may influence the simulated catchment's response depending on the modelled surface runoff generating mechanism (Hearman and Hinz, 2007).

Quantitative precipitation forecasts (QPFs) by short-range and high-resolution numerical weather prediction (NWP) models are a primary tool to further extend flood forecasting lead times beyond the basin response times. Nowadays, high-resolution

5  NWP models capture the initiation and intensification of convectively-driven rainfalls realistically, with similar spatial and temporal scales to the flash flood-prone catchments (Leoncini et al., 2013; Fiori et al., 2014). That is, QPFs can be employed to force rainfall-runoff models without the need of implementing additional downscaling procedures (Verbunt et al., 2007; Amengual et al., 2008; Addor et al., 2011; Vincendon et al., 2011; Ravazzani et al., 2016). However, hydrometeorological forecasts are impacted by different types of uncertainty that degrade the final outputs. Uncertainties arise from the physical schemes

10  used in each particular hydrological and meteorological model as well as from their initial and lateral boundary conditions (IC/LBCs). Furthermore, QPFs of small-scale convective rainfall are challenging as many factors are involved in their determination. When mesoscale convective developments are dominant, errors can become larger during the forecasting process and produce inaccurate QPFs.

Short-range ensemble prediction systems (SREPSs) aim at forecasting the probability of extreme weather as accurately as

15  possible. Uncertainties in the atmospheric state are most often encompassed by running NWP models with perturbed IC/LBCs (Buizza, 2003; Grimit and Mass, 2007). Next, 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 (*i.e.* external to the hydrological models) 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. For instance, the HEPEX (Hydrological Ensemble

20  Prediction EXperiment, Schaake et al. (2007); Thielen et al. (2008)) initiative investigated the issues linked to HEPS development, especially focussing on medium range daily streamflow forecasts over large to medium size watersheds (Thirel et al. (2008); Randrianasolo (2010) among others). However, hourly streamflow short range are less documented, and the identification of the most suitable methods for their generation and the quantification of their added value are still under investigation (Cloke et al., 2013).

25  The objective of the present work is to evaluate the predictive skill of several distinct HEPSs for the 03 November 2011 and the 28 September 2012 floods. To this end, we have used the AROME and WRF meteorological models to build two SREPSs. The WRF-SREPS has been generated by dynamically downscaling the operational ECMWF-EPS (European Centre for Medium Range Weather Forecasts – global Ensemble Predictions System) forecasts, analogously as the method implemented in Amengual et al. (2015). However, the WRF model has been run with an enhanced horizontal and vertical spatial

30  resolutions in order to resolve deep convection explicitly, without the use of any cumulus parameterization. The so-called AROME ensemble has been built on the perturbation of the rainfall outputs of AROME-WMED (Fourrié et al., 2015) model by the Vincendon et al. (2011) methodology. Aftetrwards, the performance of these two distinct SREPS generation strategies when forcing both hydrological models has been analysed through a twofold approach. First, QPFs have been evaluated by comparing them to the interpolated raingauge data. Second, QDFs have been examined by using classical probabilistic objec-

tive scores. The impact of the forecasting lead time of the SREPSs as well as the size of the ensemble have been investigated also.

Next, section 2 presents a short overview of the flash-floods, the study areas and observational networks; section 3 provides the hydrological tools, the atmospheric models and ensemble generation strategies; results are presented in section 4. The last
5    section summarizes main conclusions and provides further remarks.

## 2   Study regions, databases and flash-flood episodes

### 2.1   The Llobregat River basin and the first select episode

The Llobregat basin is the largest internal hydrographic catchment of Catalonia (Fig. 2 a). Altitudes range from above 3,000 $m$ in the Pyrenees to between 200-750 $m$ in the pre-coastal and littoral ranges. The Llobregat River basin extends covering an
10    area of 5,040 km$^2$, with a maximum length around 170 $km$. The Llobregat covers a wide spectrum of annual rainfall amounts depending on altitude. Precipitation is broadly beyond 1000 $mm$ where the altitude is higher than 1000 $m$ in the Pyrenees. Annual rainfall amounts are rougly of 700 $mm$ over the pre-Pyrenees, with altitudes ranging from 600 to 1000 $m$, and barely reach 500 $mm$ for lower heights. The Llobregat basin exhibits a mild and rainy cold season and a hot and dry warm season, characteristic of the Mediterranean climate. Furthermore, extreme precipitations affect the Spanish Mediterranean region every
15    year, which represent a substantial part of the total amounts. Two dams are found in the montainous regions of the Llobregat (Fig. 2 a).

Available raw precipitation consists of 5-minute rainfall data recorded at 81 stations inside or very close to the Llobregat basin (Fig. 2 a). These pluviometric stations belong either to the Catalan Agency of Water – Automatic Hydrological Information System (ACA-SAIH) or the Spanish Agency of Meteorology (AEMET) networks. Five-minute streamflow data are
20    recorded at four hydrometric sections in the Llobregat River. These stream-gauges are included in the ACA-SAIH network and are deployed in: (i) Súria town, with a dranaige area of 940 km$^2$ (labelled as Súria); (ii) Sant Sadurní d'Anoia city, with a drainage area of 736 km$^2$ (Sadurní); and (iii) Castellbell (3340 km$^2$) and (iv) Sant Joan Despí (4915 km$^2$) towns (labelled as Castellbell and Despí, respectively). A more detailed description of the Llobregat River basin and databases can be found in Amengual et al. (2007).
25    The first selected event occurred in autumn 2011, which has been loaded with heavy precipitation events (HPE) in the north-western Mediterranean (Silvestro et al., 2012; Hally et al., 2013; Rebora et al., 2013). The northern part of Catalonia was particularly affected at the beginning of November, 2011. On November 2nd, the large-scale situation was dominated by a deep, cold upper-level trough approaching from the North Atlantic Ocean while a southeasterly low-level flow was strengthening over the Catalonia region. Those atmospheric conditions favoured convection and heavy rainfall occurred. A maximum amount of
30    daily precipitation of 202.9 $mm$ was recorded on November 3rd over Catalonia and slightly above 150 $mm$ over the Llobregat River basin. Most of the rivers of Catalonia were flooded even if minor damage was produced in Catalonia (Llasat et al., 2014).

## 2.2 The Guadalentín River basin and the second selected episode

The Segura is one of the most important Spanish rivers running into the Mediterranean Sea. This catchment spans over an area of 18 208 km$^2$ and the maximum length of the Segura River is about 325 $km$. The Guadalentín is the main affluent of the Segura River. The river basin extends over a region of 3343.1 km$^2$ and the river length is close to 121 $km$ (Fig. 2 b). The

5  Guadalentín river basin comprises altitudes ranging from above 2,000 m (in the Baetic system) to 1,200 m in the Murcian pre-litoral depression, and to barely 110 m in its mouth to the Segura. The Guadalentín basin is found in a particularly arid region of Mediterranean Spain, owing to its particular settlement. This Baetic range shelters this area from the frequent passage of cold fronts coming from the Atlantic in the wet season and bringing copious precipitations to other Spanish regions. Thus, precipitation in the catchment relies on the cyclogenesis of Mediterranean systems and the subsequent impinging of humid

10  low level flow coming from south west. But these disturbances are sparse in time and small in space. Rainfall amounts are altitude-dependent and range from 50 to 300 $mm$.

108 automatic rain-gauges are located within the Confederación Hidrológica del Segura (CHS) boundaries (Fig. 2 b). Raw precipitation accumulations are recorded with a time frequency of 5 minutes. These automatic stations belong either to the Confederación Hidrológica del Segura – Automatic Hydrological Information System (CHS-SAIH) or to AEMET networks.

15  Almost 40 of these 108 stations are deployed over the Guadalentín River watershed or in the vicinity. Streamflow is recorded at three hydrometric sections along the Guadalentín River: in Lorca and Paretón de Totana cities (labelled as Lorca and Paretón, respectively) and in Salabosque -outskirts of Murcia city-. Their respective drainage areas are of: 1827.1 km$^2$, 2384.7 km$^2$ and 3170.4 km$^2$ (Fig. 2 b). Runoff is registered in the CHS-SAIH network with a temporal frequency of 5 minutes as well.

The hydraulic section of the CHS is well aware of the short recurrence of catastrophic flash-floods affecting the whole region.

20  Therefore, numerous structural measures have been implemented over the Segura River basin along the years. Specifically, four reservoirs are located within the Guadalentín. Furthermore, a channel was constructed downstream of Paretón so as to link directly the river with the Mediterranean Sea (Fig. 2 b). This channel was designed specially so as to prevent hazardous flooding impacting Murcia city. Therefore, this channel partially diverts large runoff discharges from the Guadalentín River into the Mediterranean Sea. Further details on the Guadalentín and databases can be found in Amengual et al. (2015).

25  The second study case is a classical Spanish Mediterranean HPE that affected Andalusia, Murcia and later on Valencia and Catalonia (even if less intense there) from 27 to 29 September 2012. This case has been documented during the HyMex campaign within an Intensive Observing Period, IOP8 (Amengual et al., 2015). About ten casualties have been deplored and damage has been estimated to more than 120 Meuros. On 27 September 2012, at upper-level, the synoptic atmospheric situation was controled by a cut-off low centered on the South-West of the Iberian Peninsula moving slowly north-eastward and a low-

30  level depression centered upon inner Andalusia. The upwards forcing associated with the north-east flank of the cut-off low favoured the trigerring of deep convection and the development of a low-level convergence zone. Those ingredients reinforce convection and heavy precipitation across Murcia, Valencia, the Balearics and Catalonia, that was affected on 30 September 2012. The amounts of precipitation locally reached 214 $mm$ in 24 hours in Andalusia and 240 $mm$ in Murcia. They caused the flash-flooding around Murcia, especially the one of the Guadalentín River (Ducrocq et al., 2014; Amengual et al., 2015).

Natural Hazards
and Earth System
Sciences
Discussions

## 3   Models and methodology

### 3.1   Meteorological models

#### 3.1.1   The WRF model

The Weather and Research Forecasting (WRF) model version 3.5 (Skamarock, 2008) has been implemented with a single
5  computational domain of 767 x 575 grid-points that is centered in the Western Mediterranean and spans the entire Mediter-
ranean Spanish coast (Fig. 1). We have used a horizontal resolution of 2.5 km, 50 vertical levels and an integration time-step of
12 s. This model set-up allows resolving explicitly deep moist convective systems with relevant entity (Weisman and Klemp,
1997; Bryan et al., 2003; Roberts and Lean, 2008; Zheng et al., 2016). With this model configuration, the aim is to improve
the prediction of the convective precipitations through a more accurate representation of the physics of the convective systems
10  (Done et al., 2004).

The operational European Centre for Medium Range Weather Forecasts – global Ensemble Predictions System (ECMWF-
EPS) aims at coping with uncertainty on the initial atmospheric conditions, taking all the observed and modelled information
available into account. In particular, the global ECMWF-EPS consist of 50 integrations with a spatial resolution of $\sim$ 20 km,
after applying singular vector perturbation to an initial forecast (Palmer et al., 1998; Leutbecher, 2005; Hoskins and Coutinho,
15  2005). By a dynamical downscaling of the global ECMWF-EPS, we rely on the sampling of the IC/LBCs uncertainties provided
by the system, but we simulate more precisely the small-scale interactions so as to enhance the high-resolution quantitative
precipitation forecasts (Branković et al., 2008). Lateral boundary fields are updated every 3-h.

Physical parameterizations are identical across all WRF ensemble members and involve: the WRF single-moment 6-class mi-
crophysical scheme with graupel (WSM6; Hong and Lim (2006)); the 1.5-order Mellor-Yamada-Janjić boundary layer scheme
20  (MYJ; Janjić (1994)); the Dudhia short-wave scheme (Dudhia, 1989); the RRTM long-wave scheme (Mlawer et al., 1997); the
unified NOAH land surface model (Tewari et al., 2004); and the eta similarity surface layer model (Janjić, 1994). Atmospheric
simulations span over two consecutive 48-h periods for the Llobregat event: 02-04 November and 03-05 November 2011 00
UTC, respectively. For the Guadalentín episode, NWP simulations span over two consecutive 48-h periods: 27-29 Septem-
ber and 28-30 September 2012 00 UTC, respectively. Next, hourly QPFs outputs are used to force the hydrological models.
25  Note that the model set-up matches the WRF operational configuration employed at the University of the Balearic Islands
($http : //meteo.uib.es/wrf$).

#### 3.1.2   The AROME-WMED model

AROME-WMED is a research convection-permitting model dedicated to HYMEX (Fourrié et al., 2015). It is based on AROME-
France that is the operationnal non-hydrostatic model of Météo-France (Seity et al., 2011). The models run at a 2.5-km hori-
30  zontal resolution. It has 60 vertical levels that range from 10 m above ground to 1 hPa. The deep convection is thus explicitly
resolved. Its microphysical parameterization is a one-moment 5-class scheme (Pinty and Jabouille, 1998; Caniaux et al., 1994)
with rain, snow, graupel, cloud liquid water and cloud ice. In the boundary layer, the vertical turbulent transport follows by

two schemes: the parameterization of Cuxart et al. (2000) for prognostic turbulent kinetic energy for an eddy diffusivity part or the mass flux scheme of Pergaud et al. (2009) for dry thermal and shallow convection. The surface scheme is SURFEX (Masson et al., 2013) and the surface boundary layer is SLB (Masson and Seity, 2009). AROME-WMED covers a domain that encompasses most of the Western Mediteranean Sea (34N-11W,48N-20E; see Fourrié et al. (2015) to vizualize it). Its lateral

5  boundary conditions come from forecasts of the French operational global model ARPEGE (Courtier et al., 1991). During HYMEX SOP1, a daily 48-h AROME-WMED forecast was run in real time, starting at 00 UTC every day from September to December 2012. A 3D-var data assimilation scheme is used to produce the initial atmospheric state. Compared to AROME-France analysis, more satellite and surface observations and some experimental measurements such as low-layer ballons data and additional radiosoundings were included in the data assimilation process. The overall quality of the AROME-WMED

10  model proved to be as good as the AROME-FRANCE one, especially QPFs. All details can be found in Fourrié et al. (2015).

Vincendon et al. (2011) used the object-oriented verification method called SAL (Wernli et al., 2008) to evaluate AROME-France QPF, in terms of location and amplitude errors. The references used were the quantitative precipitation estimates (QPE) provided by radar data (see Tabary, 2007;Tabary et al., 2007). No systematic biases in the magnitude of the rainfall were found

15  and location errors remained lower than 50 $km$ in the 80% of the cases. So, the high-resolution process-based model trajectory can be used to feed an hydrological model subject to the consideration of the uncertainty that affects QPF.

A classical way to take the uncertainty of atmospheric forecasts into account is to use a meteorological EPS. A prototype of a meteorological ensemble at convective-scale based on AROME model has been developed (Vié et al., 2011; Nuissier et al., 2012; Bouttier et al., 2012). This system is called AROME-EPS. Unfortunatelly, the AROME-EPS domain does not encompass

20  the Murcia area. So, it has not been used in this work and another approach has been adopted.

Vincendon et al. (2011) developed the so-called "Perturbation method" so as to create a set of QPF scenarios on the basis of the AROME-France deterministic forecast that contains valuable information. Primarly, an object-oriented climatology of forecasts errors was established besides the SAL evaluation cited in the former section. First rainy objects were defined as corresponding to connected grid cells with more than 2 $mm$ an hour. The same is performed for convective objects, which

25  were connected grid cells with more than 100 $mm$ an hour. Some probability density functions (PDFs) of errors in term of amounts of rain and location of those objects were computed. For the present study, this climatology has been extended using a more extended domain (including Northern Spain) and including more rainy events. The study sample gathered rainy events from 2008 to 2010. Then a perturbation tuned from those PDFs was introduced into the deterministic forecast. The perturbation method follows the following steps:

30  - Shifting of the rainy objects in accordance with the PDF of the errors in location of the AROME deterministic forecasts.
- Change of the rainfall intensity inside the rainy objects in accordance with the PDF of the errors in amplitude of the rainy objects.
- Change of the convective objects within each rainy object that are set more or less peaked/flat in accordance with the PDF of the amplitude errors of the convective objects.

35  Next, this method was applied to the AROME-WMED model forecasts. This is possible owing to the very close statistics

Natural Hazards
and Earth System
Sciences
Discussions

obtained with AROME-France and AROME-WMED in QPF evaluations. This method tuned on AROME-France thus can be applied to AROME-WMED forecasts on other regions and periods, given that the rainy events considered have common characteristics with some in the sample.

This system is not a meteorological EPS but it allows to obtain several members for AROME forecasts. Consequently, the
5   system will be called AROME ensemble in the following in order to simplify the writing and understanding. It is worth to mention that the AROME ensemble has finally an actual horizontal resolution that is higher than the WRF ensemble, due to the different approaches used to generate the members of those ensembles. In this work, 48-h AROME-WMED forecasts started on 02 and 03 November 2011 at 00 UTC for the Llobregat study case, and 27 and 28 September 2012 at 00 UTC for the Guadalentín study case.

10   ## 3.2   Hydrological models

### 3.2.1   The HEC-HMS model

HEC-HMS has been used as semi-distributed, conceptual- and event-based configuration for both river basins (USACE-HEC, 2000). A single hyetograph is used to drive the hydrological model for each subbasin. First, the hourly rainfall spatial accumulations are obtained from the 40 rain-gauges available. Spatial distributions are obtained by the kriging method. The
15   semi-variogram has been matched after applying a linear model. This technique is founded in minimizing the error variance and is prescribed for rain-gauge stations with an uneven spatial distribution (Bhagarva and Danard, 1994; Seo, 1998). Next, we have computed the time series of the hourly rainfall amounts for each individual subwatershed as the area-average of the spatially gridded precipitation within the corresponding subbasin. Note that the same procedure is used to drive the hydrological model with the QPFs, but we have used NWP model grid points instead of the observed precipitation. Spatially gridded
20   precipitation is 1-$km$ resolution.

Runoff is computed according to the curve number methodology (CN; USDA, 1986). CNs depends non-linearly on a wide range of facets: accumulated precipitation, lithology, land uses and soil's antecedent moisture condition (AMC; Chow and W. (1988)). CNs were obtained after experimental field campaigns with AMC II for both river basins. The dimensionless unit hydrograph (UH) provided by the American Soil Conservation Service has been applied so as to transform the effective rainfall
25   into overland flow for every sub-basin. This conceptual scheme links the runoff maximum with the time-to-peak by accounting for the sub-basin area and a conversion constant (USACE-HEC, 2000). The Muskingum method has been implemented so as to propagate the flood hydrograph (Chow and W., 1988). Therefore, the model set-up employs spatially uniform conceptual model parameters for each sub-basin and for the dynamical formulation as well.

The dams have been modelled by using the following information provided by the CHS and ACA hydraulic divisions: the
30   storage capacity, maximum outflow and elevation, and initial water level. The reservoirs have been simulated by using their elevation-storage-outflow relationships (USACE-HEC, 2000). Finally, we have implemented a diversion element to account for the diversion of the flood volumes towards the Mediterranean Sea in the Guadalentín river. Note that data on diverted flows have been provided by the CHS.

Model calibration focused on peak discharges, timings and runoff volumes, which are strongly dependent on infiltration and routing processes. In the semi-arid Mediterranean Spain, sparse vegetation together with torrential convective precipitations, that easily exceed the high initial soil infiltration capacity, favour fast Hortonian flows and rapid flow velocities in the river streams (Belmonte and Beltrán, 2001). Consequently, we calibrated the following model parameters: SCS-CNs, times of

5 concentrations and flood wave celerities. Note that SCS-CNs encompass effectively the initial soil moisture conditions. The calibration procedure was carried out by combining a manual and an automatic approach. The automatic procedure uses as objective function the root mean square error weighted according to the peak; and as search algorithm, the univariate-gradient method (USACE-HEC, 2000). A complete description of the HEC-HMS set-ups and the calibration and validation tasks for both catchments can be found in (Amengual et al., 2007, 2009, 2015).

10 All the hydrologic simulations comprise a 96-h period. For the Llobregat flooding, HEC-HMS has been run from 02 to 06 November 2011 00 UTC. For the Guadalentín flash flood, model simulations span from 27 September to 01 October 2012 00 UTC. These simulation periods encompass the whole flash-flood episodes. Note that the hydrological model applies a linear interpolation to convert hourly rainfall amounts to the model time-step.

### 3.2.2 The ISBA-TOP model

15 The hydrological model ISBA-TOP (Bouilloud et al., 2010) is dedicated to Mediterranean catchments simulations. ISBA-TOP fully couples the land surface model ISBA (Interaction Surface Biosphere Atmosphere, Noilhan and Planton, 1989) and a version of TOPMODEL (Beven and Kirkby, 1979) that has been adapted to the Mediterranean areas (Pellarin et al., 2002). This coupling consists of introducing a lateral distribution of soil water following the TOPMODEL concept into ISBA. ISBA deals with the soil-atmosphere exchanges : water and energy budgets are managed over a rectangular domain described by 1

20 $km^2$ grid cells. The hydrological processes are simulated over soil vertical columns. The catchments are described by small grid cells (called pixels) according to a DTM. The ISBA soil moisture over a grid cell allows determining the water-storage deficit on the corresponding catchment pixels as well as the hill slope recharge. TOPMODEL equations allow computing the lateral water transfers within the catchment using the topographical information and rainfall spatial distribution. The pixels water-storage deficit are thus updated and are transformed back into ISBA soil moisture. Pixels with null deficits define the

25 saturated contributive areas.

ISBA computes soil water flows from those new saturated areas and soil moisture fields. Runoff over saturated areas (Dunne, 1978) occurs when water excess concerns the contributive areas that are diagnosed by the TOPMODEL approach. Hortonian runoff that occurs when rainfall intensity exceeds the infiltration capacity (Horton, 1933) is estimated on the the non-saturated fraction of the grid. ISBA also computes gravitational drainage at the bottom of the deepest soil layer. The drainage flow

30 computed for an ISBA grid cell is distributed over all the corresponding catchment pixels, whereas total runoff is assumed to occur on the saturated catchment pixels only. The DTM informs about the geomorphology of the catchment, including the distance between each hill slope pixel and the river. The water flows are routed up to the river so as to compute total discharges at each river pixel through a geomorphological method. Artinyan et al. (2016) link the river discharge variable and the water velocity in the river.

A full description of the coupling principle is available in Bouilloud et al. (2010). This ISBA-TOP original version has recently been improved so as to obtain satisfactory results without a calibration of the parameters. They are defined through pedotransfert functions (Vincendon et al., 2016). Reservoirs and diversion elements have been considered on the same basis as described in section 3.2.1.

5    In this work, ISBA-TOP has been run with an hourly time step. Hourly rainfall amounts collected by rain gauges and spatially distributed (through a linear horizontal interpolation method) are used to drive ISBA-TOP. Other parameters (2m-air temperature, 2m-air humidity,10m-wind speed and direction,...) are set constant. AROME-WMED analyses provide ISBA-TOP initial soil water contents and temperatures. ISBA-TOP simulations start systematically 2 days before the beginning of the studied period, so on 01/11/2011 at 00 UTC and on 26/09/2012 at 00 UTC to ensure soil moisture balance at the beginning
10   of the rainfall event.

### 3.3   Designed HEPS

Several HEPS have been built by coupling HEC-HMS and ISBA-TOP hydrological models with both WRF and AROME-WMED SREPSs. Recall that the SREPSs have been run for two 48-h different periods for each case study: on 02/11/2011 00 UTC and 03/11/2011 00 UTC for the Llobregat flood, and on 27/09/2012 00 UTC and 28/09/2012 00 UTC for the Guadalentín
15   flooding. Each SREPS has 50 members plus the unperturbed simulation. The resulting HEPSs have been labelled as: AROME-ISBA, AROME-HMS, WRF-ISBA and WRF-HMS, respectively, and have been run for the aforementioned periods. In brief, we have four experimental set-ups encompassing the 03/11/2011 and 28/09/2012 floods: RUN02 (starting on 02/11/2011 00 UTC), RUN03 (starting on 03/11/2011 00 UTC), RUN27 (starting on 27/09/2012 00 UTC) and RUN28 (starting on 28/09/2012 00 UTC). For each experimental set-up we have two different 48-h EPSs and four subsequent HEPS. Several experiments have
20   been designed for evaluation purposes. They are listed in Table 1 together with their starting time and experimental name. Summarizing the different experiments, we have distinct types of data that have been used to drive the hydrological models:

- A reference simulation consists of using rainfall collected by raingauges and spatially interpolated by kriging. These rainfall data will be called QPE, Quantitative Precipitation Estimates. The discharge simulations obtained with both hydrological models driven by measured rainfall are annotated $REF2011$ for the November 2011 case and $REF2012$
25   for the HYMEX IOP8 case (see Table 1).

- The hydrological models driven by deterministic QPF (Quantitative Precipitation Forecasts) allow to obtain deterministic QDF (Quantitative Discharge Forecasts). AROME produces deterministic QPF originally (as described in section 3.1.2 ) while, for the WRF model, the unperturbed member of ECMWF-EPS dowscaled by WRF is used as control and will be denoted as "deterministic" as well. For each event, two simulations of deterministic forecasts have been produced at
30   two distinct starting times.

- Ensemble rainfall scenarios provided by WRF and AROME SREPSs allow to obtain ensembles of QDFs. The same experiments than those deterministic are produced (two starting times for each study case).

Finally, more than 2400 discharge time series have been obtained.

### 3.4 Evaluation method

The hourly discharges ensemble forecats skills have been assessed computing objective scores on the whole data sample, using all the outlets and both cases in order to increase the statistical significance. See A for their definition. Classically, Root Mean Square Error ($RMSE$) has been computed to provide a comparison with observed discharge data and $\sigma$ to inform about

5  the spread of the HEPSs. For an informative ensemble, $RMSE$ has weak values and $\sigma$ has the same order of magnitude as $RMSE$. So ratios $\frac{\sigma}{RMSE}$ lower than one indicate a lack of spread. Brier Scores ($BS$) have been computed also to quantify the ability of the ensembles to forecast threshold exceedance. The threshold exeedance for the HEPS is compared generally with a threshold exceedance in observations, which are considered as a reference. In the present study, raingauge driven discharges simulations are used as the reference. Computing $BS$ for several thresholds allows to point out a potential threshold impact.

10  An extention of $BS$ for all the considered thresholds is the Ranked Probability Score ($RPS$). In the present study, the $RPS$ of the ensembles are compared against the reference -that is the simulations with the derteministic versions of the models- by computing the Ranked Probability Skill Scores ($RPSS$). $RPSS$ assesses the benefit on the studied ensembles compared with the deterministic version of the same model. $RPSS$ score greater than zero means better skill for the ensemble than for the reference.

## 4  Results

### 4.1  Hydrological models

The different experiments have been assessed at the following outlets for each river basin, where the stream mesurements are quality-controlled by the ACA and CHS hydraulic divisions:

– Despí, Castellbell and Sadurní for the Llobregat River and

20  – Lorca and Paretón outlets for the Guadalentín River.

Figure 3 presents the cumulated Nash-Sutcliffe efficiency frequencies at these five stream-gauges for the simulated hourly discharges within $REF2011$ and $REF2012$ experiments with HEC-HMS model (in blue) and ISBA-TOP (in orange). These empirical cumulative distributions are built following the method of Le Lay and Saulnier (2007). The Nash-Sutcliffe efficiencies computed for each simulation are ranked from the smallest to the largest. The probability of a value being less than the $i_{th}$ smallest Nash-Sutcliffe effyciency (given by $(i - 0.5)/n$, with $n$ total number of data) gives the cumulative frequency. On the graph, the more the distribution is shifted to the right, the better is the skill of the model. This representation allows to characterise the overall accuracy of a model from a regional point of view. Better results are obtained with HEC-HMS model with more frequent high Nash-Sutcliffe efficiencies. This is confirmed by the representation of the observed versus

30  simulated discharges (see Fig. 4). The points are quite well organised along the bisectrix with the HEC-HMS model. Note that HEC-HMS was previously calibrated and validated by using observed streamflow and precipitation data for both river basins

(Amengual et al., 2007, 2009, 2015), while ISBA-TOP has been run without any previous calibration task as described by Vincendon et al. (2016), but by using an spatial estimation of the initial soil moisture. Note that the initially dry soils and high infiltration capacities of these semi-arid catchments enhance the nonlinear response of runoff to intense precipitation and large rainfall amounts. In addition, heterogeneities also arise in the hydraulics of these basins' response to flash floods. Without a
5  previous task of calibration and verification, it is difficult to cope with these high non-linearities (Amengual et al., 2017).

An illustration of the raingauge driven simulated hourly discharge time series is given by Figures 5 and 6, that show some examples at Castellbell and Paretón of the deterministic as well as ensemble forecasts. For the November 2011 case, the flow peaks obtained with ISBA-TOP driven by QPE are underestimated compared to the observations at Despí (not shown) and Castellbell (see Fig. 5a) and overestimated at Sadurní (not shown). Two flow peaks are obtained with both hydrological models,
10  a moderate time lag is obtained with HEC-HMS; In addition, this model simulates a peak of the good order of magnitude at Castellbell (Fig. 5b), but an overestimation is obtained at Sadurní (not shown). That is, ISBA-TOP presents more inaccuracies when modelling the surface flow of the first wave of intense rains, probably as consequence of estimating the initial conditions of the soil drier than the actual. Another reason could be that the ISBA-TOP model first computed the runoff over saturated areas, so the saturation excess process contributes more to runoff generation than the Hortonian process. However, both flash-
15  floods were mainly dominated by a rapid surface flow production as consequence of the very intense precipitations.

The hourly discharges produced by the QPE driven HEC-HMS simulations are quite close to the observed discharges for the September 2012 case (Fig. 6b), slightly underestimated with ISBA-TOP (Fig. 6a). Naturally, the distinct physical schemes of both hydrological models lead to distinct hydrological responses yielding useful information from both configurations. The calibrated HEC-HMS model simulates more properly the peak flow amplitude but the timing seems better reproduced by the
20  uncalibrated ISBA-TOP coupled system as it simulates more accurately the dynamic processes of this flooding. This tends to show that a multi-model approach can be really informative. That is, the SCS-CN method is more suitable for simulating flow volumes before fast response Hortonian floods, but an spatially-distributed modelling of the dynamic of the overland flow is paramount for capturing properly the flood timing. A reason could be that once the rain has moistened the soil, the saturation excess runoff generation process plays a stronger role in streamflow dynamics and ISBA-TOP succeed better in simulating it.

25  ## 4.2 Skill of the designed HEPS

### 4.2.1 QPF evaluation

Catchment-averaged precipitation amounts have been computed for 24-h periods for each experiment of Table 1 and for the distinct rainfall sources : QPE based on raingauge measurements, deterministic QPFs from the AROME and WRF models and ensemble forecasts. Figures 7 and 8 show the results for the cases of November 2011 and September 2012, respectively.
30  The Llobregat catchment-averaged precipitation amounts from 03/11/2012 at 00 UTC to 04/11/2012 at 00 UTC is over-/underestimated by AROME/WRF deterministic QPFs for both $RUN02$ and $RUN03$ (see Fig. 7). On the contrary, all the ensemble strategies succeed in encompassing the raingauges value either in the interquartile range (for the AROME ensemble)

or in the "highest scenarios" for the WRF ensemble as figure 7 shows. Note also the generalized higher ensemble spreads for AROME than for WRF.

The deterministic QPFs from both models and both starting times underestimate the 24h-accumulated rainfall averages on the Guadalentín catchment (see Fig. 8). The ensemnle QPF scenarios do not succeed in correcting this tendency, except for the

5   $RUN28$ corresponding to the AROME ensemble. These facts illustrate the arduous task of simulating extreme precipitations properly over medium-sized basins in terms of location, timing and rainfall amounts. Recall that small-scale convective systems are highly sensitive to different NWP model configurations, physical schemes and IC/LBCs.

As far as rainfall the spatial distribution is concerned, AROME deterministic QPF overestimates the highest rainfall for

10   $RUN02$ and $RUN03$ experiments and locates them too north-west compared to the observation. The location is better for WRF deterministic QPF but the rainfall amounts are underestimated (not shown). The behaviour of the ensembles is illustrated by figure 9. AROME ensemble still leads to overestimate the spatial extension of the highest precipitation whereas WRF ensemble still mislocate them. So the errors of the deterministic NWP models are still present in the ensembles. AROME $RUN27$ and WRF $RUN27$ deterministic as well as ensemble runs do not succeed in simulating neither good rainfall amounts

15   nor a good spatial distribution (see Fig. 10). WRF $RUN28$ forecasts a good location for the heaviest rains but the amounts are far too weak (except for the highest percentiles of the ensemble). The contrary is obtained with AROME $RUN28$ whose QPF reaches the good totals but with a mislocation. AROME $RUN28$ has a clear tendency to extend more westward the rainfall. That is, WRF fails to simulate realistically the extreme rainfall values produced by the convective-scale systems anchored by the complex orography, while AROME fails to locate them realistically. It is worth noting that the WRF ensemble does not

20   account for those uncertainties associated to the physical parameterization diversity. Some previous studies have shown the benefits of using multiple physics ensemble strategies to further reduce biases when forecasting HPEs in the Mediterranean Spain (Tapiador et al., 2012; Amengual et al., 2017). Moreover the higher actual resolution of the AROME ensemble can induce a slightly better rainfall intensity forecast.

### 4.2.2   QDF evaluation

25   The raingauge driven discharges remain uncertain as shown in section 4.1 due to both hydrological modelling and QPE uncertainties. Therefore, the rain-gauge driven discharge simulations are considered as the references so as to assess the HEPS performances. That is, the $REF2011$ and $REF2012$ experiments. In this way, the hydrological modeling uncertainties are avoided when evaluating the QDFs. As aforementioned, figures 5 and 6 show some examples of the hourly discharge time series. Those examples illustrate that the HEPS approach improves the forecast compared to the deterministic experiments for

30   both extreme floods. Indeed, the inter-quartile envelopes better encompass the raingauge driven discharges (blue lines) than the deterministic forecasts (green lines). This conclusion is the same than what was concluded for rainfall. A focus on the $RUN03$ experiment is interesting. The observed rainfall is well approximated by the AROME ensemble 25%-percentile or by the WRF ensemble 75%-percentile (see Fig. 7). But as far as the discharges are concerned (see figure 5), the peak discharge simulation is

close to the 75%-percentile for the AROME ensemble driven HEPS and is out of the interquartile range for the WRF ensemble driven HEPS. This fact highlights the highly nonlinear nature of the rainfall-runoff transformation in semi-arid basins.

The cumulative distribution functions (CDFs) of the flow peaks at Castellbell and Paretón are plotted for the previously comented examples (Figs. 11 and 12, Ferraris et al., 2002). This additional information further complements the benefits of

5  accounting for HEPSs when coping with flash floods. Peak discharge exceedance probabilities $[P(Q \geq q)]$ of the observed flows quantify the likelihood of forecasting these extremes. $[P(Q \geq Qobs)]$ are of 0.36, and 0.20 for the AROME-ISBA HEPSs at Castellbell and Paretón, respectively. Similarly, peak discharge exceedance probabilities are of 0.18 and 0.16 for the WRF-HEC HEPSs (Table 2 ). As reference, AEMET declares an alert when extreme weather has a probability of occurrence greater than or equal to 0.20. In addition, almost all HEPSs issue unequivocal probabilities of surpassing the different discharge

10  return periods ($Qp_T$), spaning from 0.34 to 0.98 depending on the hydrometric section (Table 2 ; for further information about the values of the $Qp_T s$, see Amengual et al., 2009, 2015).

Even if the observed peak discharge exceedance probabilities range from low to moderate for both HEPSs, the AROME-ISBA HEPSs would have triggered emergency procedures before both episodes according to the AEMET criterion. Furthermore, the distinct hydrological ensemble strategies would prove useful for conveying proper information to civil protection and

15  emergency decision-makers before both floods as $[P(Q \geq q)s]$ are all well above 0.2. Note that the different discharge return periods quantify the risk of facing hazardous floodings.

Table 3 points out positive $RPSS$ values for all the ensembles that confirm a benefit of driving the hydrological models with the ensembles rather than with the deterministic QPFs. Nevertheless, the $RPSS$ is lower using WRF meteorological ensemble. The $RMSE$ values are not significantly different for these extreme floodings (less than 10%). On the contrary, the $\sigma$ scores

20  are very far one from the others. The spread is higher using AROME than using WRF and higher with HEC-HMS model than with ISBA-TOP. However, the ratio $\frac{\sigma}{RMSE}$ is always lowerthan 1. This tends to show a lack of spread except when HEC-HMS is driven with AROME ensemble. The lesser spread of the WRF than the AROME ensemble can be attributed to the fact that the synoptic and large mesoscale dynamical and thermodynamical environment is sufficiently accurate in the ECMWF unperturbed member. Therefore, less additional information is conveyed by the perturbed IC/LBC ensemble members. On the

25  other hand, the AROME ensemble is founded on a climatological model error database, accounting not just for inaccuracies in the IC/LBCs, but also in the model formulation. In this case, it would be advisable to also account for inaccuracies in the WRF model physical schemes when using a perturbed IC/LBC ensemble in order to further span the EPS spread. The different spreads in the HEC-HMS and ISBA-TOP models can be attributed to the distinct model infiltration schemes, when initiating runoff as a based-threshold process.

30  ### 4.2.3  Impact of the forecasting lead time

In the following, the study cases are considered separately. The Brier scores have been computed for the hourly QDFs rendered by the meteorological models started at different lead times. Figures 13) and 14 depict the results for the November 2011 and September 2012 episodes, respectively. These forecast probability levels are: $Qq < 0.05$; $0.05 \leq Qq < 0.15$; $0.15 \leq Qq < 0.25$; ...; $0.85 \leq Qq < 0.95$; and $Qq \geq 0.95$. All HEPS are verified for the following hourly discharge thresholds: 0, 1, 2, 4, 8,

Natural Hazards
and Earth System
Sciences
Discussions

16, 32, 128, 256 and 512 $m^3.s^{-1}$. To increase the statistical robustness, a boot strap with 1000 repetitions has been applied (Diaconis and Efron, 1983). It appears that the forecasting lead time of the atmospheric EPSs has a strong impact on the flash-flood forecasts performances. Figure 13 shows that, for the November 2011 case, except for the HEPS where HEC-HMS is driven by AROME, the lower Brier scores are obtain with $RUN03$ rather than $RUN02$. For the September 2012 case, $RUN28$ leads to better results that $RUN27$ for the four HEPS (see Fig. 14). This tends to show that better results are obtained using the latest meteorological forecasts. This can be expected for the WRF ensemble since it is built by perturbing the initial conditions. So, the closer the lead times to the episode are, the more accurate is the representation of the initial state of the atmospheric conditions that result in the flash-floods. With AROME, the results are more varied. In fact, this AROME ensemble comes from a perturbation of the forecasted rainfall, so it depends more on the deterministic scenario. The AROME deterministic simulations also rely on the initial conditions. So, in theory, the later the simulations start, the more accurate are the IC/LBCs to those producing the flash-floods. But closer lead times to the episode date does not guarantees to have better simulations than starting before, especially when dealing with extremes. This has to be confirmed on more a wider climatology, but it would point out to a possible improvement of the forecasting and warning schemes before hazardous floods in order to better establish the confidence levels from an operational perspective.

### 4.2.4  Impact of the number of ensemble members

The computational cost of a HEPS with 50 members is very high from an operational point of view. So, another issue of the maximum interest is to examine the ensemble forecasting skill in terms of the ensemble size. That is, are so large-sized ensembles really necessary or smaller size ensembles have similar forecasting skill? To investigate this issue, a sub-sample of ensemble members is selected randomly among the whole population for each ensemble experiment. How the performance is affected in terms of probabilistic QDFs is assessed by computing the aforementioned statistical scores (i.e. $RPSS$, $RMSE$, $\sigma$, $\frac{\sigma}{RMSE}$) for the different sub-sets of HEPSs. The size of the additional ensembles range from 10 to 40 members. Fig. 15 shows the scores depending on the number of members for the September 2012 case study. Reducing the number of members leads to a deterioration of the scores but this relationship is not linear and depends on the considered HEPS. It seems that the best compromise in terms of RPSS and spread (i. e. $\frac{\sigma}{RMSE}$ for instance) is obtained around 20 or 25 members. These outcomes do not change by just considering the November 2011 episode or by considering both episodes together (figs. not shown).

### 5  Conclusions

In line with the major scientific goal of the HyMeX program of improving flood early warning procedures and mitigation measures, several distinct HEPSs have been built to forecast two flash-floods events that have occurred over two semi-arid Spanish Mediterranean watersheds. These HEPSs make use of two different meteorological ensembles at convective scale and two hydrological modelling systems. First, the performances of the hydrological models driven by observed rainfall data have been assessed first. Next, the skill of the HEPSs is studied for the two flash-flood events. The aim was two-folded: to analyze

Natural Hazards
and Earth System
Sciences
Discussions

the forecasting skill of several distinct HEPSs as well as to investigate the impact of the forecast lead time and of the number of ensemble members. The main conclusions are :

– The calibrated SCS-CN method better simulates the peak discharges but a spatially-distributed (even uncalibrated) model better reproduces the flood dynamics.

5 – The use of an ensemble rather than a deterministic approach clearly improves the forecasts of both extreme rainfall episodes over the watersheds and discharges at the different hydrometric sections. This is clear whatever the case, catchment and HEPS.

– The AROME ensemble shows a higher spread than the WRF ensemble. Recall that high ensemble spread can avoid underdispersion. This result would indicate that accounting for the physical scheme uncertainty could enhance the ensemble
10 spread.

– The results of the HEPS intercomparison vary depending on the case and catchment. But it is interesting to note that the conclusions made on the catchment-averaged rainfall forecasts can be different from those made on the discharge forecasts owing to the strong non linearities in the rainfall-runoff transformation. This fact clearly shows the added value of assessing ensembles through an hydrological point of view.

15 – As far as lead time is concerned, a better skill is obtained for ensembles with shorter forecasting lead times, at least in the sample considered in our study.

– Regarding the ensemble size, the ensembles with 50 members obtain the best objective scores, but considering more computationally-affordable ensemble sizes (around 20 or 25 members) does not deteriorate the ensemble performance significantly. This result could help to the optimal design of real-time hydrometeorological forecasting chains.

20 Although two case studies do not allow to reach general conclusions about the predictability of this kind of hydrometeorological hazards or about the optimal hydrometeorological forecasting strategy in an operational framework, it points out important aspects to take into account in future studies. Note that the 28 September 2012 episode is a prototype of long-lasting mesoscale convective systems that are responsible for the most hazardous flash floods over the Western Mediterranean. Finally, this study aims at assessing HEPSs that combine multi-model and ensemble approaches. This kind of hydrometeorological approach will
25 develop in the future. Integrative studies with integrated chains considering meteorology, hydrology but also hydraulics and impacts are promising approaches for a future transference to an operational context and breakthrough improvements in risk management strategies.

## Appendix A: Scores for objective evaluation

The probabilistic scores used in this work to compare the skill of the ensembles are described in this section. The root mean square error is computed as:

$$RMSE = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(m_i - o_i)^2} \tag{A-1}$$

5  where $T$ is the total number of time steps , $o_i$ the reference value at time step $i$ and $m_i$ the mean of the forecast members. $RMSE = 0$ for an ensemble which ensemble mean perfectly matches the observation.

The spread is defined as:

$$\sigma = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_{k,i} - m_i)^2} \tag{A-2}$$

with $N$ the total number of reference discharge data, $n$ the number of forecast members and $x_{k,i}$ the value of member $k$ at time
10  step $i$. If $\sigma$ and $RMSE$ are of the same order of magnitude, the ensemble prediction system brings valuable information.

The Brier Score (Brier, 1950) measures the root mean square distance between the probability of a binary event (that can be a threshold exceedance) and its occurrence. It is computed as :

$$BS = \frac{1}{N}\sum_{t=1}^{N}(p_t - o_t)^2 \tag{A-3}$$

with $p_t$ the probability at time step $t$ of a threshold exceedance, and $o_t = 1$ if the observation exceeds this threshold at time
15  step $t$ and $o_t = 0$ otherwise. This score is 0 for a perfect forecast, and the worst forecast gives a score of 1.

The ranked probability score is a Brier Score (Wilks, 1995) but for multiple thresholds.

$$RPS = \frac{1}{K-1}\sum_{k=1}^{K}[Y_k - O_k]^2 \tag{A-4}$$

where $K$ is the number of considered classes (or thresolds), $k$ the index of a given threshold, $Y_k$ the cumulated distribution of forecasts at a given threshold idexed $k$ and $O_k$ the cumulated distribution of observation at the threshold idexed $k$. The Ranked
20  probability skill score ($RPSS$) links the RPS of a forecasting system to the RPS of a reference:

$$RPSS = 1 - \frac{RPS}{RPS_{ref}} \tag{A-5}$$

$RPSS$ is positive when the evaluated forecast system is beneficial compared to the reference.

*Competing interests.* No competing interests are present

Natural Hazards
and Earth System
Sciences
Discussions

Open Access

EGU

# References

Addor, N., S., J., F., F., and M., Z.: An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios., Hydrology and Earth System Sciences, 15, 2327–2347, 2011.

Amengual, A., Romero, R., Gómez, M., and Martín, A., a. A. S.: A hydrometeorological modeling study of a flash-flood event over Catalonia,
5      Spain., Journal of Hydrometeorology, 8, 292–303, 2007.

Amengual, A., Romero, R., and Alonso, S.: Hydrometeorological ensemble simulations of flood events over a small-size basin of Majorca Island, Spain, Quarterly Journal of the Royal Meteorological Society, 134, 1221–1242, 2008.

Amengual, A., Romero, R., Vich, M., and Alonso, S.: Inclusion of potential vorticity uncertainties into a hydrometeorological forecasting chain: application to a medium size basin of Mediterranean Spain., Hydrology and Earth System Sciences, 13, 793–811, 2009.

10    Amengual, A., Homar, V., and Jaume, O.: Potential of a probabilistic hydrometeorological forecasting approach for the 28 September 2012 extreme flash flood in Murcia, Spain, Atmospheric Research, 166, 10–23, 2015.

Amengual, A., Carrio, D., Ravazzani, G., and Homar, V.: A Comparison of Ensemble Strategies for Flash Flood Forecasting: The 12 October 2007 Case Study in Valencia, Spain., Journal of Hydrometeorology, 18, 1143–1166, 2017.

Artinyan, E., Vincendon, B., Krumova, K., Nedkov, N., Tsarev, P., Balabanova, S., and Koshinchanov, G.: Flood forecasting and alert system
15    for Arda River basin, Journal of Hydrology, 541, 2016.

Belmonte, A. M. C. and Beltrán, F. S.: Flood events in Mediterranean ephemeral streams (ramblas) in Valencia region, Spain., Catena, 45, 229–249, 2001.

Beven, K. and Kirkby, M.: A physically based, variable contributing area model of basin hydrology, Hydrological Sciences Bulletin, 24, 43–69, 1979.

20    Bhagarva, M. and Danard, M.: Application of optimum interpolation to the analysis of precipitation in complex terrain., Journal of Applied Meteorology, 33, 508–518, 1994.

Bouilloud, L., Chancibault, K., Vincendon, B., Ducrocq, V., Habets, F., Saulnier, G.-M., Anquetin, S., Martin, E., and Noilhan, J.: Coupling the ISBA land surface model and the TOPMODEL hydrological model for Mediterranean flash-floods forecasting : description, calibration and validation, Journal of Hydrometeorology, 11, 315–332, 2010.

25    Bouttier, F., Vié, B., Nuissier, O., and Raynaud, L.: Impact of stochastic physics in a convection-permitting ensemble., Monthly Weather Review, 140, 3706–3721, 2012.

Branković, C., Matjacić, B., and S., I.-S.: Downscaling of ECMWF Ensemble Forecasts for Cases of Severe Weather: Ensemble Statistics and Cluster Analysis , Monthly Weather Review, 136, 3323–3342, 2008.

Brier, G. W.: Verification of forecasts expressed in term of probability, Monthly Weather review, 78, 1–3, 1950.

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

Buizza, R.: Weather Prediction: Ensemble Prediction. Published in the Encyclopaedia of Atmospheric Sciences, Academic Press, London, 2003.

Caniaux, G., Redelsperger, J.-L., and J.-P., L.: A Numerical Study of the Stratiform Region of a Fast-Moving Squall Line. Part I: General
35    Description and Water and Heat Budgets , Journal of the Atmospheric Sciences, 51, 2046–2074, 1994.

Chow, V.T.and Maidment, D. R. and W., M. L.: Applied hydrology , McGraw-Hill International Editions, Civil engineering series, 13, 572 pp, 1988.

Cloke, H., Pappenberger, F., van Andel, S., Schaake, J., Thielen, J., and Ramos, M.: Special Issue on Hydrological Ensemble Prediction Systems(HEPS), Hydrological Processes, 27, 1–4, 2013.

Courtier, P., C., F., Geleyn, J., F., R., and M., R.: The Arpege project at Météo-France. , ECMWF 1991 Seminar Proceedings: Numerical Methods in Atmospherics Models, ECMWF, Reading, United Kingdom, pp. 193–231, 1991.

5  Cuxart, J., Bougeault, P., , and Redelsperger, J.-L.: A turbulence scheme allowing for mesoscale and large-eddy simulations., Quarterly Journal of the Royal Meteorological Society, 126, 1–30, 2000.

Diaconis, P. and Efron, B.: Computer-intensive methods in statistics., Scientific American, 248, 116–130, 1983.

Done, J., Davis, C. A., and Weisman, M.: The next generation of NWP: Explicit forecasts of convection using the Weather Research and Forecasting (WRF) model., Atmospheric Science Letters, 5(6), 110–117, 2004.

10  Doocy, S., Daniels, A., Murray, S., and Kirsch, T.: The Human Impact of Floods: a Historical Review of Events 1980-2009 and Systematic Literature Review, PLOS Currents Disasters, 5, doi: 10.1371/currents.dis.f4deb457904 936b07c09daa98ee8171a, 2013.

Drobinski, P., Ducrocq, V., Alpert, P., Anagnostou, E., Béranger, K., Borga, M., Braud, I., Chanzy, A., Davolio, S., and Delrieu, G. e. a.: Hymex: a 10-year multidisciplinary program on the mediterranean water cycle, Bulletin of the American Meteorological Society, 95, 1063–1082, 2014.

15  Ducrocq, V., Braud, I., Davolio, S., Ferretti, R., Flamant, C., Jansa, A., Kalthoff, N., Richard, E., Taupier-Letage, I., Ayral, P.-A., Belamari, S., Berne, A., Borga, M., Boudevillain, B., Bock, O., Boichard, J.-L., Bouin, M.-N., Bousquet, O., Bouvier, C., Chiggiato, J., Cimini, D., Corsmeier, U., Coppola, L., Cocquerez, P., Defer, E., Delanoë, J., Di Girolamo, P., Doerenbecher, A., Drobinski, P., Dufournet, Y., Fourrié, N., Gourley, J., Labatut, L., Lambert, D., Le Coz, J., Marzano, F., Molinié, G., Montani, A., Nord, G., Nuret, M., Ramage, K., Rison, B., Roussot, O., Said, F., Schwarzenboeck, A., Testor, P., Van-Baelen, J., Vincendon, B., Aran, M., , and Tamayo, J.: HyMeX-SOP1, the field
20  campaign dedicated to heavy precipitation and flash-flooding in Northwestern Mediterranean, Bulletin of the American Meteorological Society, 95, 2014.

Dudhia, J.: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale twodimensional model., Journal of the Atmospheric Sciences, 46, 3077–3107, 1989.

Dunne, T.: Field studies of hillslope flow processes., In Kirkby M. J. (ed.) Hillslope hydrology, pp. 227–293, 1978.

25  Ferraris, L., Rudari, R., and Siccardi, F.: The uncertainty in the prediction of flash-floods in the northern Mediterranean environement , Journal of Hydrometeorology, 3, 714–726, 2002.

Fiori, E., Comellas, A., Molini, L., Rebora, N., Siccardi, F., Gochis, D., Tanelli, S., and Parodi, A.: Analysis and hindcast simulations of an extreme rainfall event in the Mediterranean area: the Genoa 2011 case., Atmospheric Research, 138, 13–29, 2014.

Fourrié, N., Bresson, E., Nuret, M., Jany, C., Brousseau, P., Doerenbecher, A., Kreitz, M., Nuissier, O., Sevault, E., Bénichou, H., Amodei, M.,
30  and Pouponneau, F.: AROME-WMED, a real-time mesoscale model designed for the HyMeX special observation periods, Geoscientific Model Development, 8, 1919–1941, 2015.

Grimit, E. P. and Mass, C.: Measuring the ensemble spread-error relationship with a probabilistic approach: Stochastic ensemble results., Monthly weather review, 135, 203–221, 2007.

Hally, A., Richard, E., Fresnay, S., and Lambert, D.: Ensemble simulations with perturbed physical parametrizations: Pre-HyMeX case
35  studies, Quarterly Journal of the Royal Meteorological Society, 140, 1900–1916, 2013.

Hearman, A. and Hinz, C.: Sensitivity of point scale surface runoff predictions to rainfall resolution, Hydrology and Earth System Sciences, 11, 965–982, 2007.

Natural Hazards
and Earth System
Sciences

Discussions

Open Access

EGU

Hong, S.-Y. and Lim, J.-O.: The WRF single-moment 6-class microphysics scheme (WSM6). , Journal of the Korean Meteorological Society, 42, 129–151, 2006.

Horton, R.: The role of infiltration in the hydrological cycle, Transactions - American Geophysical Union, 14, 446–460, 1933.

Hoskins, B. and Coutinho, M.: Moist singular vectors and the predictability of some high impact european cyclones. , Quarterly Journal of
5      Royal Meteorological Society, 131, 581–601, 2005.

Janjić, Z. I.: The step-mountain eta coordinate model: further developments of the convection, viscous sublayer and turbulence closure schemes, Monthly Weather Review, 122, 927–945, 1994.

Le Lay, M. and Saulnier, G.: Exploring the signature of spatial variabilities in flash flood events: Case of the 8-9 september 2002 Cévennes-vivarais catastrophic event, Geophysical Research Letters, 34, 2007.

10    Leoncini, G., Plant, R. S., Gray, S. L., and Clark, P. A.: Ensemble forecasts of a flood-producing storm: comparison of the influence of model-state perturbations and parameter modifications., Quarterly Journal of the Royal Meteorological Society, 139, 198–211, 2013.

Leutbecher, M.: On ensemble prediction using singular vectors started from forecasts., Monthly Weather Review, 133, 3038–3046, 2005.

Llasat, M. C., Caumont, O., Flores, I., Garrote, L., Gilabert, J., Llasat-Botija, M., Marcos, R., Nuissier, O., Richard, E., and Rigo, T.: The 6 November 2011 flood event in Catalonia: Analusis using the DRIHM infrastructure, Proceedings of the Mediterranean Meeting on
15    "Monitoring, modelling and early warning of extreme events triggered by heavy rainfalls". MED-FRIEND project. University of Calabria, Cosenza (Italy), 2014.

Masson, V. and Seity, Y.: Including Atmospheric Layers in Vegetation and Urban Offline Surface Schemes, Journal of Applied Meteorology and Climatology, 48, 1377–1397, https://doi.org/10.1175/2009JAMC1866.1, http://dx.doi.org/10.1175/2009JAMC1866.1, 2009.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P.,
20    Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, Geoscientific Model Development, 6, 929–960, https://doi.org/10.5194/gmd-6-929-2013, http://www.geosci-model-dev.net/6/929/2013/, 2013.

25    Mlawer, E. J., Taubman, S., Brown, P., Iacono, M., and S.A., C.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, Journal of Geophysical Research, 102, 663–682, 1997.

Noilhan, J. and Planton, S.: A simple parametrization of land surface processes for meteorological models., Monthly Weather Review, 117, 536–549, 1989.

Nuissier, O., Joly, B., Vié, B., and Ducrocq, V.: Uncertainty of lateral boundary conditions in a convection-permitting ensemble: a strategy
30    of selection for Mediterranean heavy precipitation events, Natural Hazards and Earth System Science, 12, 2993–3011, 2012.

Palmer, T. N., Gelaro, R., Barkmeijer, J., and Buizza, R.: Singular vectors, metrics, and adaptive observations., Journal of the Atmospheric Sciences, 55, 633–653, 1998.

Pellarin, T., Delrieu, G., Saulnier, G., Andrieu, H., Vignal, B., and Creutin, J.: Hydrologic Visibility of Weather Radar Systems Operating in Mountainous Regions: Case Study for the Ardeche Catchment (France), Journal of Hydrometeorology, 3, 539–555, 2002.

35    Pergaud, J., Masson, V., Malardel, S., and Couvreux, F.: A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction., Boundary Layer Meteorology, 132, 83–106, 2009.

Pinty, J. and Jabouille, P.: A mixed-phase cloud parameterization for use in mesoscale non-hydrostatic model: simulations of a squall line and of orographic precipitations. Proc., Conf. of Cloud Physics, Everett, WA, USA, Amer. Meteor. soc., pp. 217–220, 1998.

Natural Hazards
and Earth System
Sciences
Discussions

Randrianasolo, A., R. M. T. G. A. V. M. E.: Comparing the scores of hydrological ensemble forecasts issued by two different hydrological models, Atmospheric Science Letters, 11(2), 100–107, 2010.

Ravazzani, G., Amengual, A., Ceppi, A., Homar, V., Romero, R., Lombardi, G., and Mancini, M.: Potentialities of ensemble strategies for flood forecasting over the Milano urban area., Journal of Hydrology, 539, 237–253, 2016.

5   Rebora, N., Molini, L., Casella, E., Comellas, A., Fiori, E., F, P., Siccardi, F., Silvestro, F., Tanelli, S., and Parodi, A.: Extreme rainfall in the Mediterranean: What can we learn from observations?, Journal of Hydrometeorology, 14, 906–922, 2013.

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

Romero, R., Doswell, C., and Ramis, C.: Mesoscale numerical study of two cases of long-lived quasi- stationary convective systems over
10   Eastern Spain. , Monthly Weather Review, 128, 3731–3752, 2000.

Schaake, J., Hamill, T., Buizza, R., and M., C.: HEPEX: the Hydrometeorological Ensemble Prediction EXperiment, Journal of HydroMeteorology, 88, 1541–1547, 2007.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bernard, P., Bouttier, F., Lac, C., and Masson, V.: The AROME-France convective scale operational model, Monthly Weather Review, 139, 976–991, 2011.

15   Seo, D.: Real-time estimation of rainfall fields using radar rainfall and rain gage data., Journal of Hydrology, 208, 37–52, 1998.

Silvestro, F., Gabellani, S., Giannoni, F., Parodi, A.and Rebora, N., Rudari, R., and Siccardi, F.: A hydrological analysis of the 4 November 2011 event in Genoa., Natural Hazards and Earth System Science, 12, 2743–2752, 2012.

Skamarock, W.C., e. a.: A Description of the Advanced Research WRF Version 3., NCAR Tech. Note NCAR/TN-4751STR, p. 125 p, 2008.

Tabary, P.: The new French operational radar rainfall product: Part I, Methodology., Weather and forecasting, 22, 393–408, 2007.

20   Tabary, P., Desplats, J., Dokhac, K., Eideliman, F., Guéguen, C., and Heinrich, J.: The new French operational radar rainfall product: Part II, Validation., Weather and forecasting, 22, 409–427, 2007.

Tapiador, F., Tao, W., Shi, J., Angelis, M., Marcos, C., Rodríguez, A., and Hou, A.: A comparison of perturbed initial conditions and multiphysics ensembles in a severe weather episode in Spain., Journal of Applied Meteorology and Climatology, 51, 489–504, 2012.

Tewari, M., F., Chen, W., Wang, J., Dudhia, M. A., LeMone, K., Mitchell, M., Ek, G., Gayno, J., Wegiel, K., and H., C. R.: Implementation
25   and verification of the unified NOAH land surface model in the WRF model., 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, 2004.

Thielen, J., Schaake, J., Hartman, R., and Buizza, R.: Aims, challenges and progress of the Hydrological Ensemble Prediction EXperiment (HEPEX) following the thrid HEPEX workshop head in Stresa 27 to 29 June 2007, Atmospheric Science Letters, 9, 29–35, 2008.

Thirel, G., Rousset-Regimbeau, F., Martin, E., and Habets, F.: On the impact of short range meteorological forecasts for ensemble streamflow
30   predictions, Journal of Hydrometeorology, 9, 1301–1317, 2008.

USACE-HEC, .: Hydrologic Modeling System HEC-HMS: Technical reference manual, p. 157 pp, 2000.

Verbunt, M., Walser, A., Gurtz, J., Montani, A., and Schär, C.: Probabilistic flood forecasting with a limited-area ensemble prediction system: selected case studies., Journal of Hydrometeorology, 8, 897–909, 2007.

Vié, B., Nuissier, O., and Ducrocq, V.: Cloud-resolving ensemble simulations of Mediterranean heavy precipitating events: uncertainty on
35   initial conditions and lateral boundary conditions, Monthly Weather Review, 139, 403–423, 2011.

Vincendon, B., Ducrocq, V., Saulnier, G., Bouilloud, L., Chancibault, K., Habets, F., and Noilhan, J.: Benefit of coupling the ISBA land surface model with a TOPMODEL hydrological model dedicated to Mediterranean flash floods , Journal of Hydrology, 394, 256–266, 2010.

Vincendon, B., Ducrocq, V., Nuissier, O., and Vié, B.: Perturbation of convection-permitting NWP forecast for flash-flood ensemble fore-casting, Natural Hazards and Earth System Science, 11, 1529–1544, 2011.

Vincendon, B., Edouard, S., Dewaele, H., Ducrocq, V., Lespinas, F., Delrieu, G., and Anquetin, S.: Modeling flash floods in southern France for road management purposes, Journal of Hydrology, in press, available online, –, 2016.

5  Weisman, M. L., S. W. C. and Klemp, J. B.: The resolution dependence of explicitly modeled convective systems, Monthly Weather Review, 125, 527–548, 1997.

Wernli, H., Paulat, M., Hagen, M., and C., F.: SAL-A novel quality measure for the verification of quantitative precipitation forecasts., Monthly Weather Review, 136, 4470–4487, 2008.

Wilks, D.: Statistical methods in the Atmospheric sciences : An introduction, Academic Press, p. 467pp., 1995.

10  Zheng, Y., Alapaty, K., Herwehe, J. A., Del Genio, A. D., and Niyogi, D.: Improving High-Resolution Weather Forecasts Using the Weather Research and Forecasting (WRF) Model with an Updated Kain-Fritsch Scheme., Monthly Weather Review, 144, 833–860, 2016.
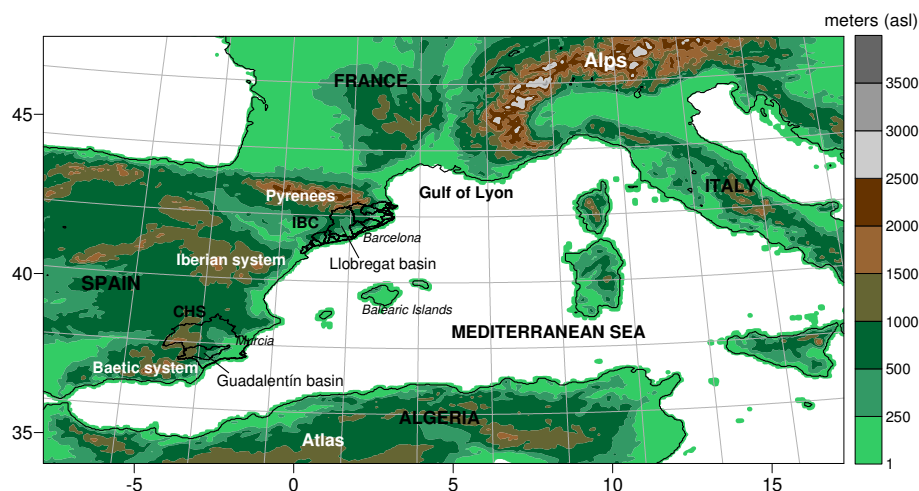
**Figure 1.** Configuration of the computational domain used for the WRF numerical simulations. Main geographical features mentioned in the text are shown. The thick continuous lines show the regions where the Llobregat and Guadalentín River basins are located in the Internal Basins of Catalonia (IBC) and the Confederación Hydrográfica del Segura (CHS), respectively. Coordinates are indicated in latitude and longitude.

Natural Hazards
and Earth System
Sciences

Discussions

Open Access
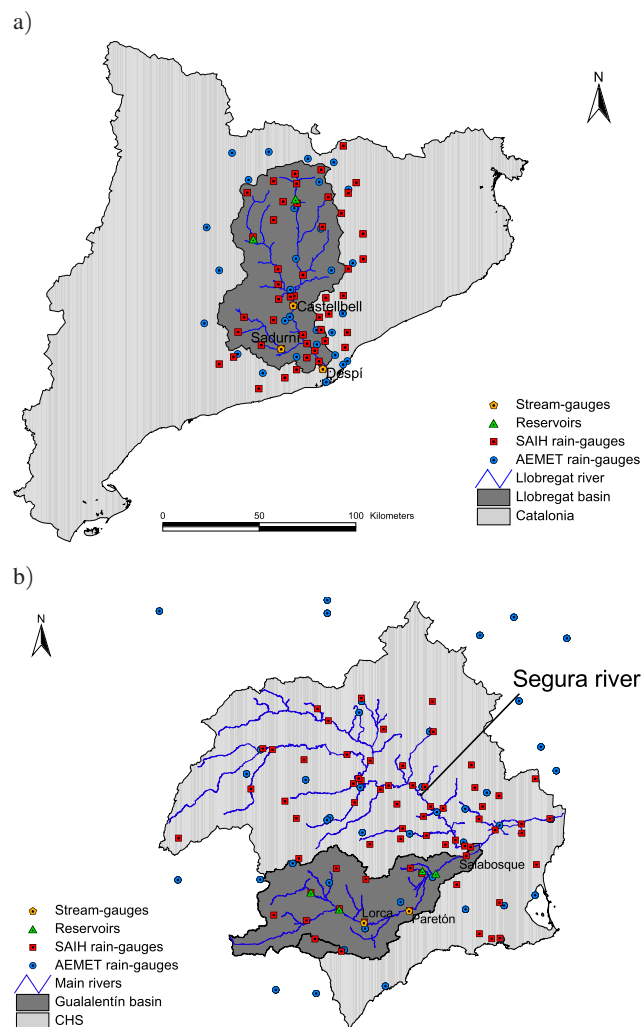


**Figure 2.** (a) Spatial distribution of the rain-gauges for the Llobregat catchment. It includes a total of 81 automatic rainfall stations. The Llobregat River basin is highlighted in shaded dark gray. (b) Spatial distribution of the rain-gauges for the CHS. It includes a total of 108 automatic rainfall stations distributed over an area of 18 208 $km^2$. The Guadalentín river basin is highlighted in shaded dark gray.
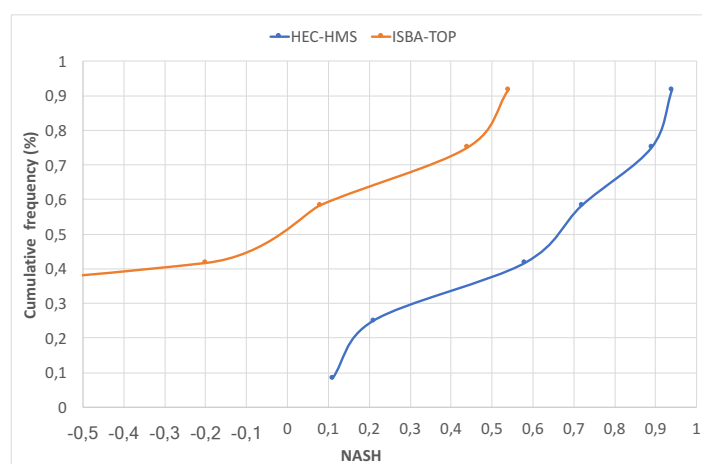
**Figure 3.** Cumulated Nash efficiency frequencies for hourly discharges simulated with HEC-HMS (blue line) and ISBA-TOP (orange line) within $REF2011$ and $REF2012$ experiments.
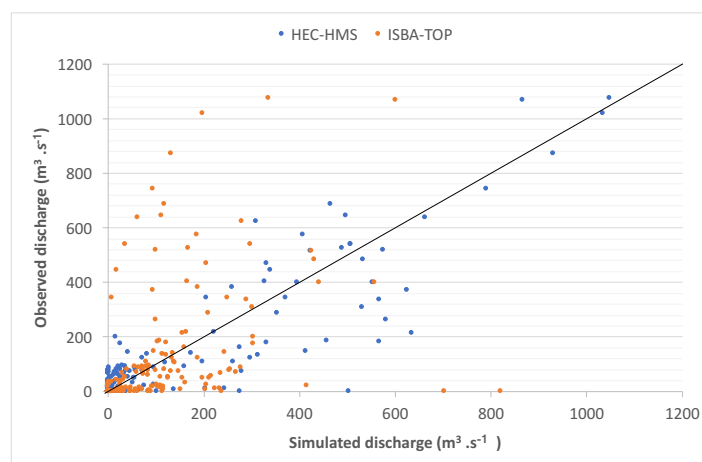
**Figure 4.** Hourly observed discharges versus simulated discharges ($m^3.s^{-1}$) for all watersheds simulated with HEC-HMS (blue points) and ISBA-TOP(orange points) within $REF$2011 and $REF$2012 experiments.
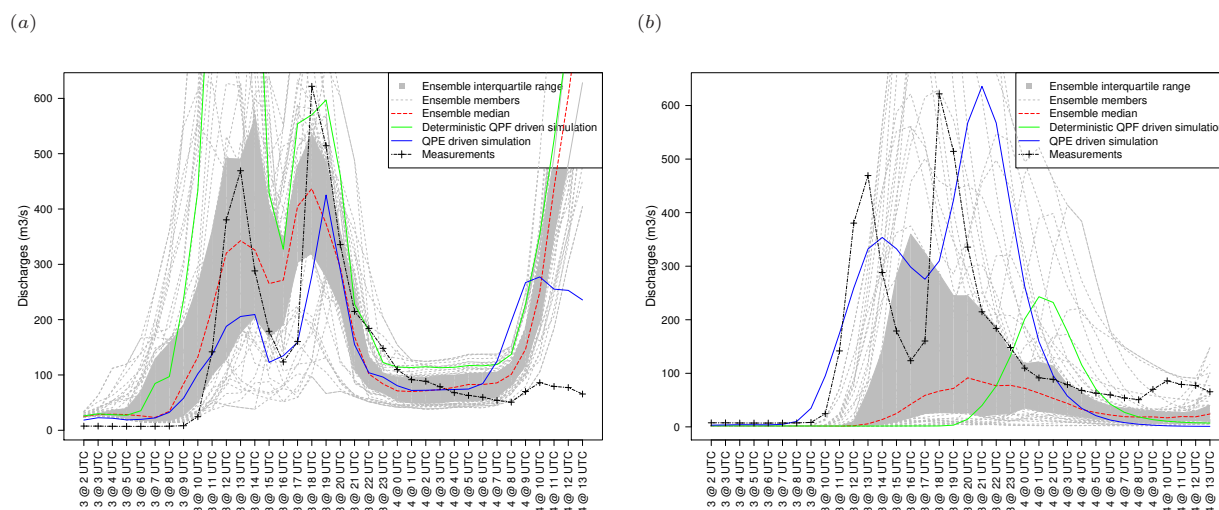
**Figure 5.** Simulated and observed hourly discharges of the Llobregat River at Castellbell for (a) the AROME $RUN03$ experiments driving ISBA TOP, and (b) WRF $RUN03$ experiments driving HEC-HMS. Black crosses denote the measurements from 03/11/2011 00 UTC to 04/11/2011 12 UTC. The blue lines the QPE driven discharge simulations. The green lines are the deterministic driven discharge forecasts. The grey dashed lines denote the individual HEPS members. The shaded areas represent the HEPS interquartile range. The red lines stand for the HEPS median.

Natural Hazards
and Earth System
Sciences
Discussions
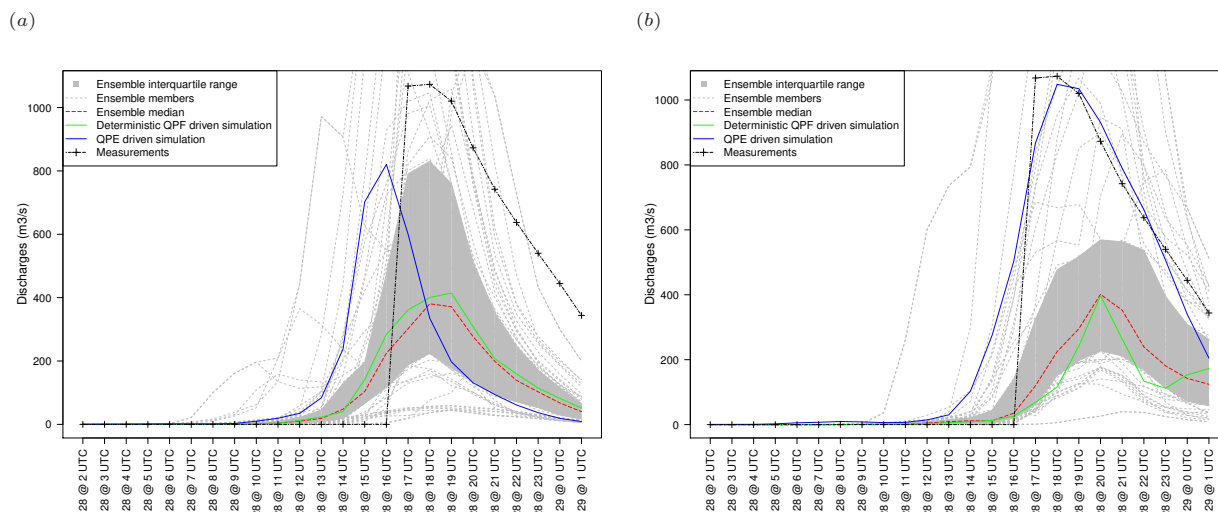


**Figure 6.** As Fig. 5 but for the Guadalentín River at Paretón and (a) the AROME RUN28 experiments driving ISBA-TOP, and (b) the WRF RUN28 experiments driving HEC-HMS from 28/09/2012 00 UTC to 29/09/2012 00 UTC.

Natural Hazards
and Earth System
Sciences
Discussions
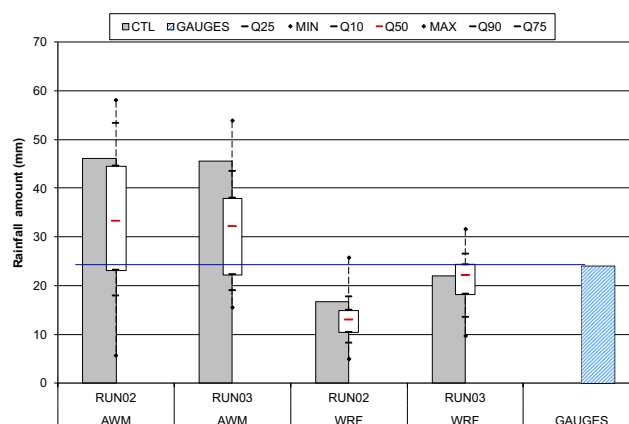


**Figure 7.** 24h-accumulated rainfall averaged on the Llobregat catchment starting on 03/11/2011 00 UTC from raingauges data (purple column), deterministic QPF (shaded columns) and ensemble forecasts (boxplots) from AROME and WRF 48-h runs started on 02/11/2011 at 00 UTC ($RUN02$) or 03/11/2011 at 00 UTC ($RUN03$).
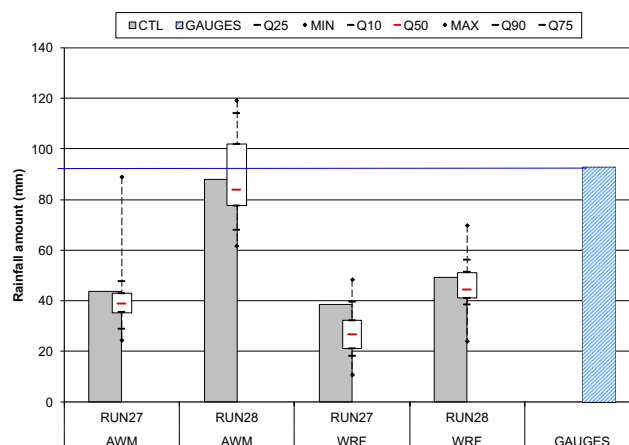
**Figure 8.** 24h-accumulated rainfall averaged on the Guadalentín catchment starting on 28/09/2012 00 UTC from raingauges data (purple column), deterministic and ensemble forecasts from AROME and WRF 48-h runs started on 27/09/2012 00 UTC ($RUN27$) or 28/09/2012 00 UTC ($RUN28$).
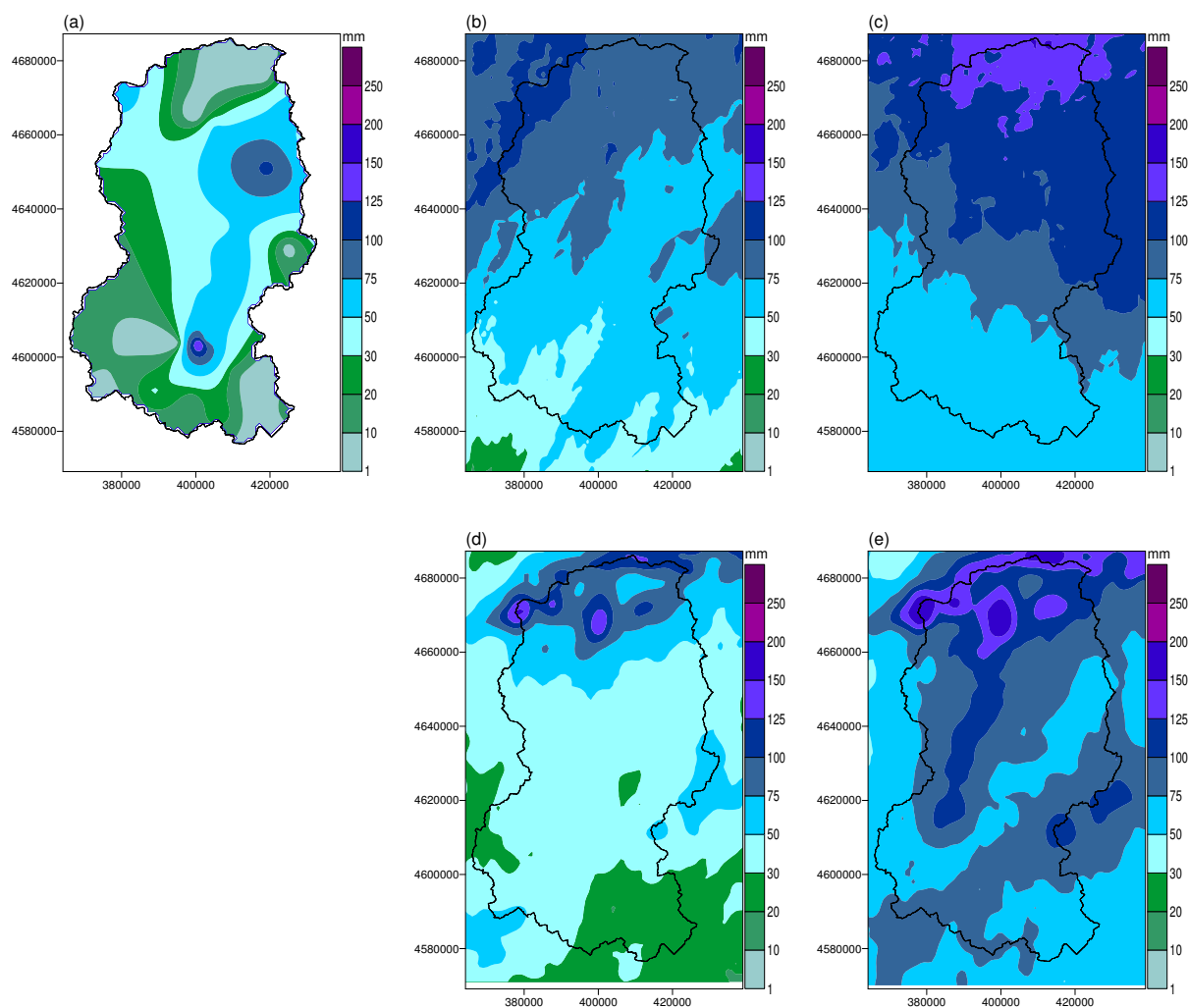
**Figure 9.** 24-h accumulated precipitation (03/11/2011 00 UTC - 04/11/2011 00 UTC) from raingauges (a), $90^{th}$-percentile of QPF from AROME ensemble RUN02 (b), RUN03 (c) and WRF ensemble RUN02 (d) and RUN03 (e) over the Llobregat catchment. Coordinates are indicated in UTM 31 N projection.
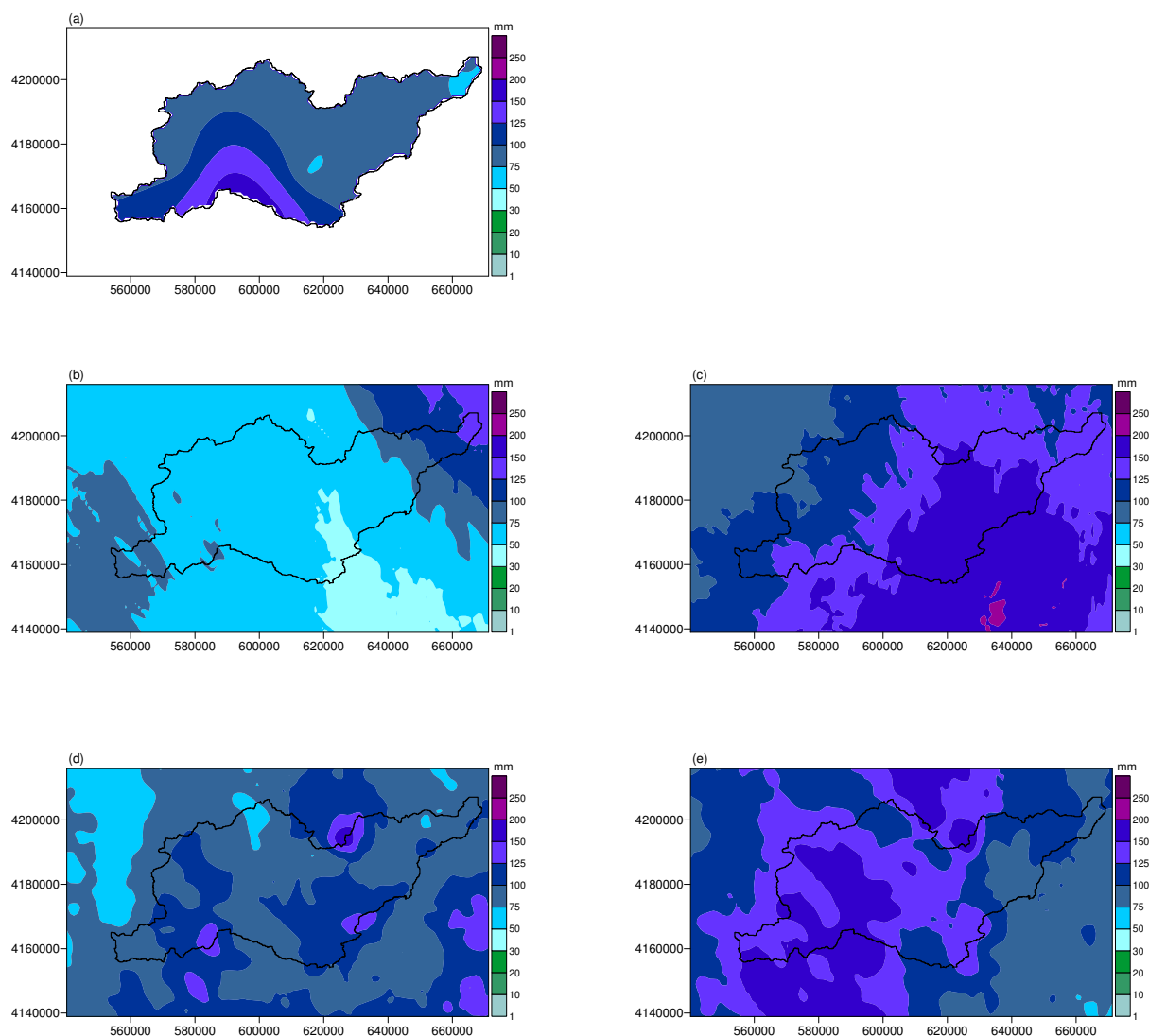
**Figure 10.** 24-h accumulated precipitation (28/09/2012 00 UTC - 29/09/2012 00 UTC) from raingauges (a), $90^{th}$-percentile of QPF from AROME ensemble RUN27 (b), RUN28 (c) and WRF ensemble RUN27 (d) and RUN28 (e) over the Guadalentín catchment.Coordinates are indicated in UTM 30 N projection.
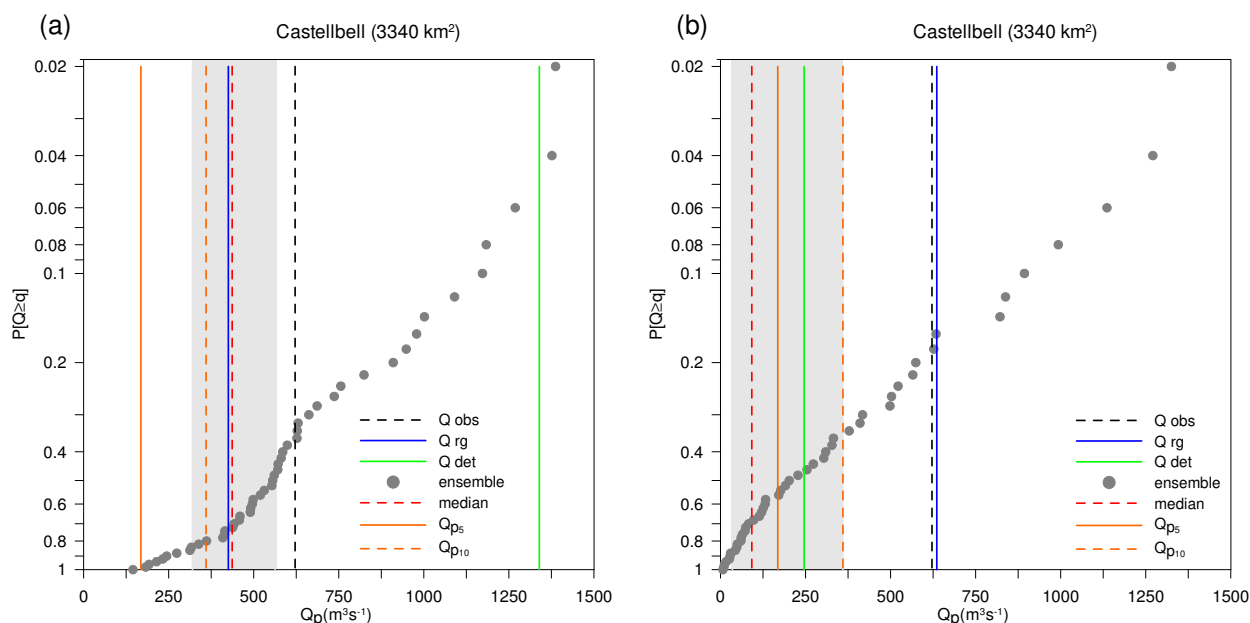
**Figure 11.** Peak discharge exceedance probabilities for the 02 November 2011 hydrometeorological episode at Castellbell for (a) AROME driven ISBA-TOP and (b) WRF driven HEC-HMS runoff experiments. The vertical dashed black line denotes the observed maximum flow. The solid blue and green lines correspond to the rain-gauge and NWP-deterministic driven maximum discharges, respectively. The dash red line represents the ensemble median peak discharge. The light gray shaded area depicts the ensemble spread between quantiles q0.25 and q0.75 of the members. Variables $Qp_5$ and $Qp_{10}$ denote peak discharge exceedance probabilities of the 5-, 10-year return periods, respectively.
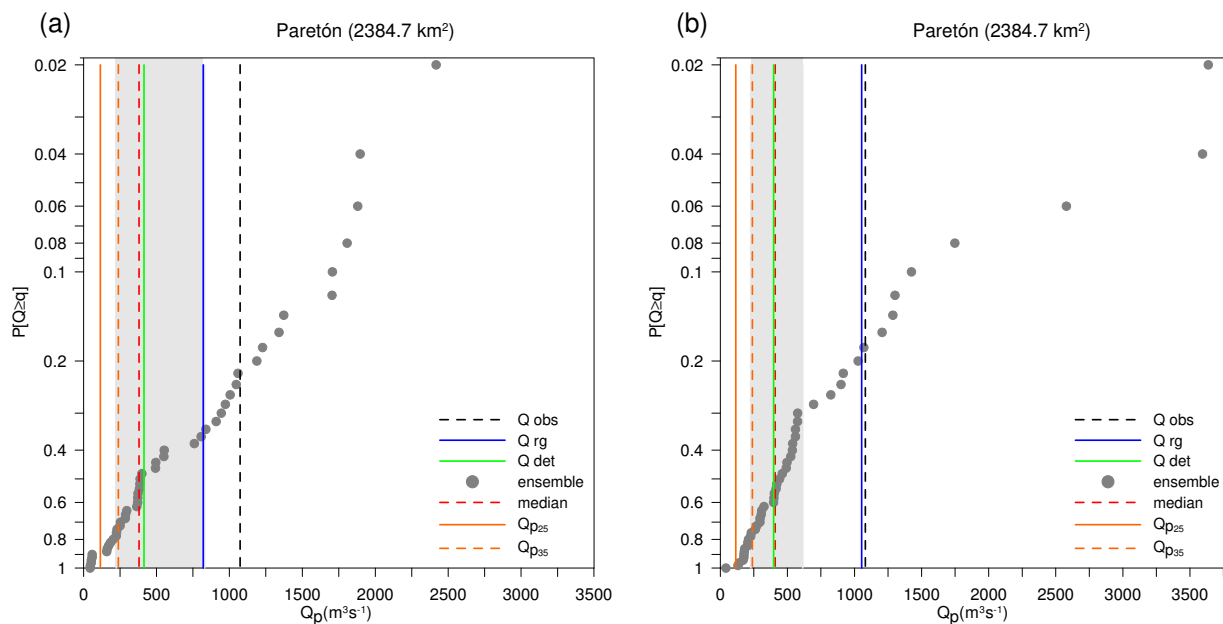
**Figure 12.** As figure 11 but for the 28 September 2012 hydrometeorological episode at Paretón. Variables $Qp_{25}$ and $Qp_{35}$ denote peak discharge exceedance probabilities of the 25- and 35-year return periods, respectively.
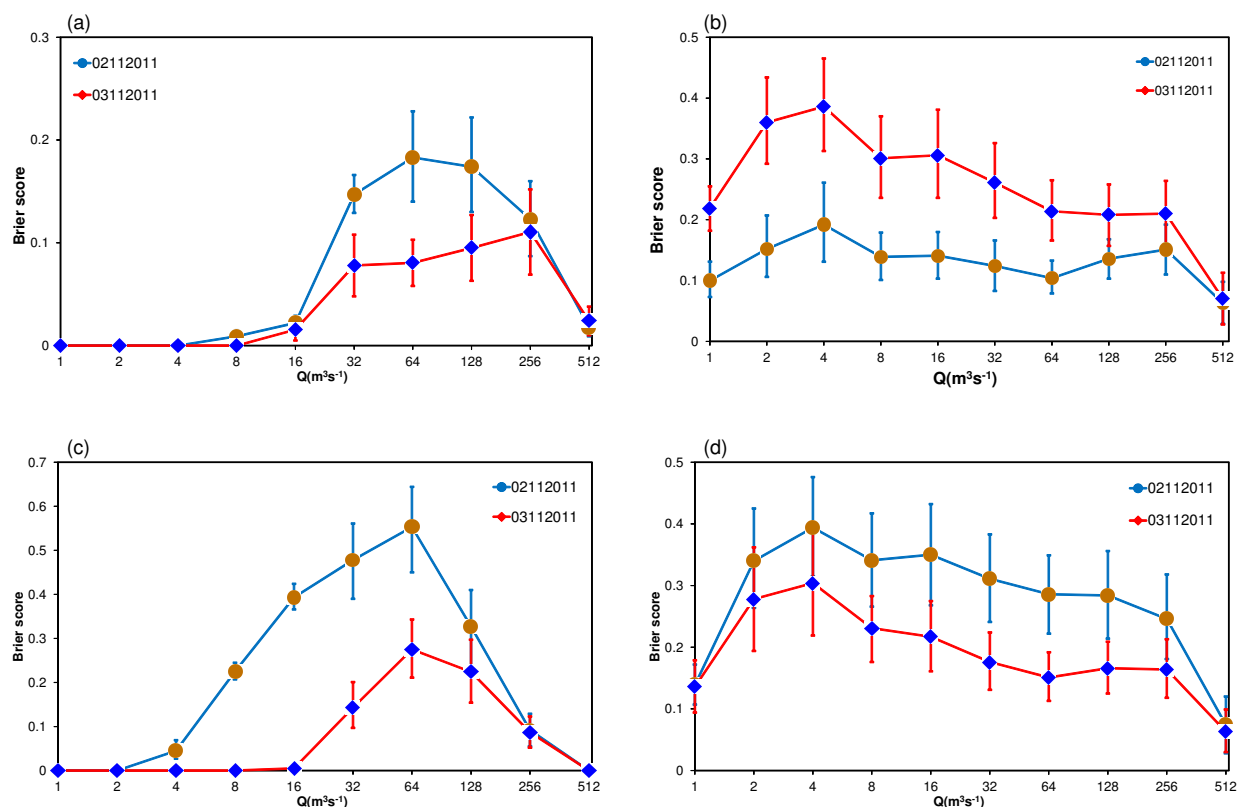
**Figure 13.** Brier Scores of discharges forecasts by AROME-ISBA ensemble (a), AROME-HMS ensemble (b), WRF-ISBA ensemble (c) and WRF-HMS ensemble (d) for the November 2011 study case . The reference is raingauge driven discharges simulation with ISBA-TOP for a and c and with WRF for b and d. The blue line is obtained for $RUN02$ experiments, the red line is obtained for $RUN03$ experiments.
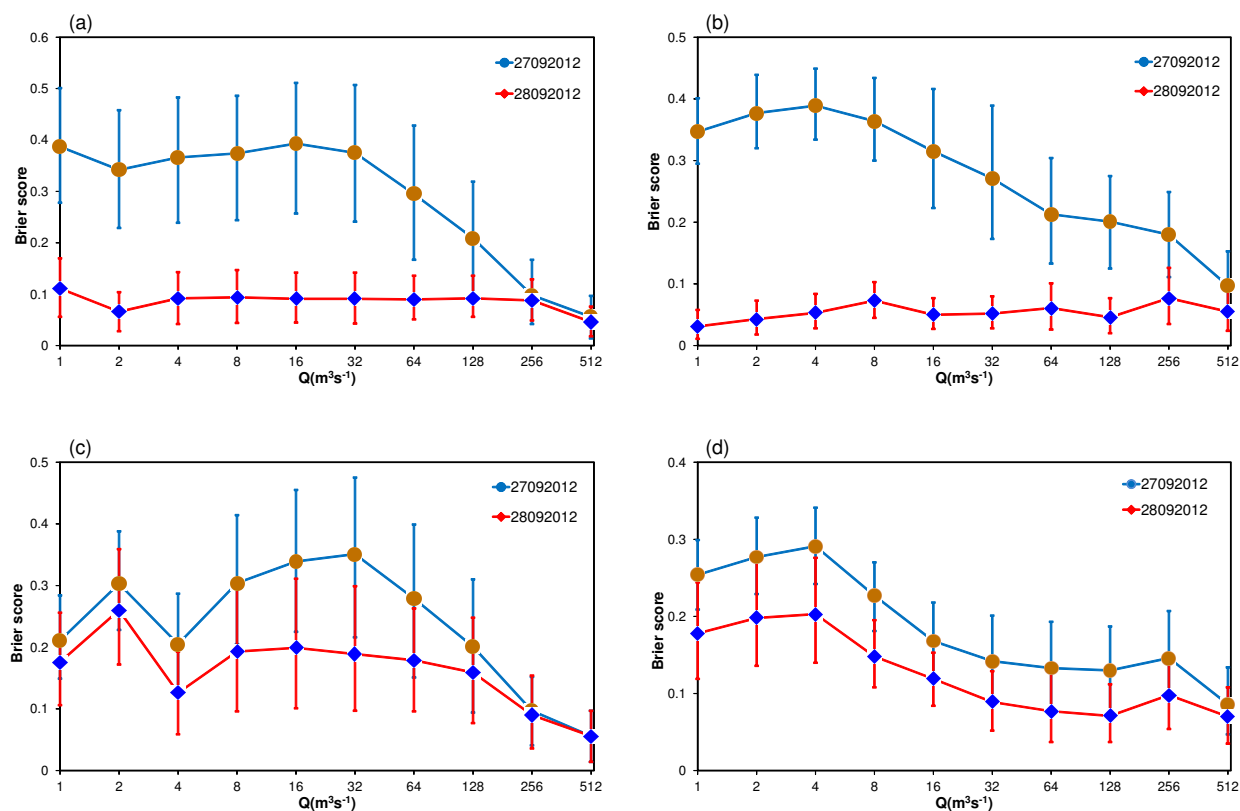
**Figure 14.** As Fig. 13 but for the September 2012 study case. The blue line is obtained for $RUN27$ experiments, the red line is obtained for $RUN28$ experiments.
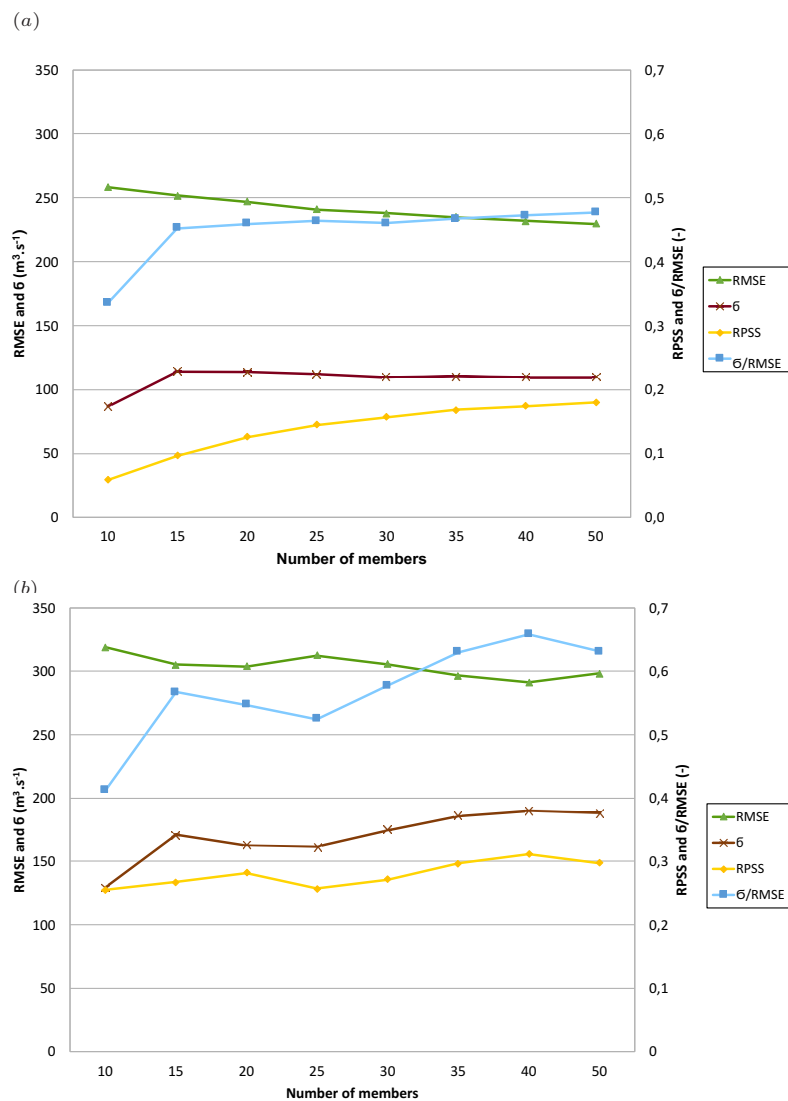
**Figure 15.** Probabilistic Scores of discharges forecasts by AROME-ISBA (a) and WRF-HMS (b) for the September 2012 study case. The reference is discharges simulations obtained with each hydrological model driven by rainfall observations.

Natural Hazards
and Earth System
Sciences
Discussions

**Table 1.** Summary of all the numerical experiments performed for the 03 November 2011 and 28 September 2012 flash floods.

| Name of the experiment | Starting time | Meteorological data | Hydrological models | Catchments where discharge is simulated | Number of simulated discharge time-series |
|---|---|---|---|---|---|
| **November 2011 event** | | | | | |
| REF2011 | | QPE | HEC-HMS | Lorca/Paretón/Salabosque | 3 |
| | | " | ISBA-TOP | " | 3 |
| RUN02 | 02/11/2011 at 00 UTC | AROME deterministic QPF | HEC-HMS | " | 3 |
| | " | " | ISBA-TOP | " | 3 |
| | " | AROME ensemble QPF | HEC-HMS | " | 150 |
| | " | " | ISBA-TOP | " | 150 |
| | " | WRF deterministic QPF | HEC-HMS | " | 3 |
| | " | " | ISBA-TOP | " | 3 |
| | " | WRF ensemble QPF | HEC-HMS | " | 150 |
| | " | " | ISBA-TOP | " | 150 |
| RUN03 | 03/11/2011 at 00 UTC | AROME deterministic QPF | HEC-HMS | " | 3 |
| | " | " | ISBA-TOP | " | 3 |
| | " | AROME ensemble QPF | HEC-HMS | " | 150 |
| | " | " | ISBA-TOP | " | 150 |
| | " | WRF deterministic QPF | HEC-HMS | " | 3 |
| | " | " | ISBA-TOP | " | 3 |
| | " | WRF ensemble QPF | HEC-HMS | " | 150 |
| | " | " | ISBA-TOP | " | 150 |
| **September 2012 event** | | | | | |
| REF2012 | | QPE | HEC-HMS | Castelbell/Sadurní/Despí | 3 |
| | | " | ISBA-TOP | " | 3 |
| RUN27 | 27/09/2012 at 00 UTC | AROME deterministic QPF | HEC-HMS | " | 3 |
| | " | " | ISBA-TOP | " | 3 |
| | " | AROME ensemble QPF | HEC-HMS | " | 150 |
| | " | " | ISBA-TOP | " | 150 |
| | " | WRF deterministic QPF | HEC-HMS | " | 3 |
| | " | " | ISBA-TOP | " | 3 |
| | " | WRF ensemble QPF | HEC-HMS | " | 150 |
| | " | " | ISBA-TOP | " | 150 |
| RUN28 | 28/09/2012 at 00 UTC | AROME deterministic QPF | HEC-HMS | " | 3 |
| | " | " | ISBA-TOP | " | 3 |
| | " | AROME ensemble QPF | HEC-HMS | " | 150 |
| | " | " | ISBA-TOP | " | 150 |
| | " | WRF deterministic QPF | HEC-HMS | " | 3 |
| | " | " | ISBA-TOP | " | 3 |
| | " | WRF ensemble QPF | HEC-HMS | " | 150 |
| | " | " | ISBA-TOP | " | 150 |

Natural Hazards
and Earth System
Sciences
Open Access
EGU
Discussions

**Table 2.** Peak discharge exceedance probabilities $[P(Q \geq q)]$ for the AROME-ISBA and WRF-HEC HEPSs for the 03 November 2011 and 28 September 2012 episodes and the indicated stream gauges. Variables $Qp_5$, $Qp_{10}$, $Qp_{25}$ and $Qp_{35}$ denote peak discharge exceedance probabilities of the 5-, 10-, 25- and 35-year return periods, respectively.

| Episode | 03/11/2011 | 28/09/2012 |
|---|---|---|
| Streamgauge | Castellbell | Paretón |
| Experiment | $[P(Q \geq Qp_{obs} = 621.8m^3.s^{-1})]$ | $[P(Q \geq Qp_{obs} = 1073.3m^3.s^{-1})]$ |
| AROME-ISBA | 0.36 | 0.20 |
| WRF-HEC | 0.18 | 0.16 |
| Experiment | $[P(Q \geq Qp_5 = 168.5m^3.s^{-1})]$ | $[P(Q \geq Qp_{25} = 115.0m^3.s^{-1})]$ |
| AROME-ISBA | 0.98 | 0.88 |
| WRF-HEC | 0.56 | 0.98 |
| Experiment | $[P(Q \geq Qp_{10} = 359.9m^3.s^{-1})]$ | $[P(Q \geq Qp_{35} = 238.5m^3.s^{-1})]$ |
| AROME-ISBA | 0.80 | 0.72 |
| WRF-HEC | 0.34 | 0.74 |

Natural Hazards
and Earth System
Sciences

Open Access

Discussions

EGU

**Table 3.** Scores for hourly discharges ($m^3.s^{-1}$) with the meteorological ensembles AROME-EPS and WRF-EPS and the hydrological models ISBA-TOP and HEC-HMS. All catchments and cases are considered.

| Meteorological model | AROME | WRF | AROME | WRF |
|---|---|---|---|---|
| Hydrological model | ISBA-TOP | ISBA-TOP | HEC-HMS | HEC-HMS |
| $RPSS\,(-)$ | 0.41 | 0.02 | 0.46 | 0.13 |
| $RMSE\,(m^3.s^{-1})$ | 222.1 | 233.3 | 229.9 | 219.6 |
| $\sigma\,(m^3.s^{-1})$ | 73.4 | 23.9 | 220.5 | 50.2 |
| $\frac{\sigma}{RMSE}(-)$ | 0.33 | 0.10 | 0.95 | 0.23 |