Potential of stochastic methods for improving convection-permitting ensemble forecasts of extreme events over the western Mediterranean

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Potential of stochastic methods for improving convection-permitting ensemble forecasts of extreme events over the Western Mediterranean

Alejandro Hermoso [a,*], Victor Homar [a], Robert S. Plant [b]

[a] *Meteorology Group, Physics Department, Universitat de les Illes Balears, Palma, Mallorca, Spain*
[b] *Department of Meteorology, University of Reading, Reading, United Kingdom*

A B S T R A C T

The design of convection-permitting ensemble prediction systems capable of producing accurate forecasts of disruptive events is an extraordinarily challenging effort. The difficulties associated with the detection of extreme events found at these scales motivates the research of methodologies that efficiently sample relevant uncertainties. This study investigates the potential of multiple techniques to account for model uncertainty. The performance of various stochastic schemes is assessed for an exceptional heavy precipitation episode which occurred in eastern Spain. In particular, the stochastic strategies are compared to a multiphysics approach in terms of both spread and skill. The analyzed techniques include stochastic parameterization perturbation tendency and perturbations to influential parameters within the microphysics scheme. The introduction of stochastic perturbations to the microphysics processes results in a larger ensemble spread throughout the entire simulation. Conversely, modifications to microphysics parameters generate small-scale perturbations that rapidly grow over areas with high convective instability, in contrast to the other methods, which produce more widespread perturbations. A conclusion of specific interest for the western Mediterranean, where deep moist convection and local orography play an important role, is that stochastic methods are shown to outperform a multiphysics-based ensemble for this case, indicating the potential positive impact of stochastic parameterizations for the forecast of extreme events in the region.

1. Introduction

Since the pioneering works of Lorenz in the 1960s, it is extensively recognized that numerical weather forecasts are uncertain, and that small differences amplify with time, eventually rendering a forecast useless (Lorenz, 1963, 1969). Although the theoretical frameworks of Liouville and Fokker-Planck equations (Hasselmann, 1976; Ehrendorfer, 1994) provide a basis to describe the time evolution of uncertain systems, their practical application in operational weather forecasting is unachievable and thus, currently restricted to low complexity systems (Hermoso et al., 2020a). In this context, ensemble prediction systems (EPS) emerge as a feasible alternative tool to account for the inherent uncertainties in numerical weather prediction, specifically initial condition and model errors.

Regarding initial conditions, different methodologies are used for the global and synoptic scales with the aim of identifying fast growing modes and optimizing the sampling of plausible states under the ever-present limitation of computational resources. Popular approaches include the singular vectors technique (Molteni et al., 1996), used at the European Centre for Medium Range Weather Forecasts (ECMWF), and bred vectors (Toth and Kalnay, 1993), which have been used at the National Center for Environmental Prediction (NCEP). Alternatively, ensemble data assimilation is an ensemble generation strategy designed to estimate the flow-dependent uncertainty of the forecast, accounting for and propagating uncertain information originating from observations and background fields. Multiple operational centers exploit this approach to initialize global scale ensembles (Whitaker et al., 2008; Houtekamer et al., 2009; Kazumori, 2014, among others). Although error sampling methodologies for large scales are mature (e.g., Buizza et al., 2005; Buizza and Leutbecher, 2015), the translation of these techniques to smaller scales and limited area configurations is not straightforward. Indeed, not only initial, but also boundary condition uncertainty must be sampled owing to the use of limited area models. A prevailing strategy to generate initial and boundary condition perturbations for regional ensembles directed at representing large scale uncertainty consists of the dynamical downscaling of a coarser-resolution

---

global ensemble (e.g., Stensrud et al., 1999; Marsigli et al., 2014; Kühnlein et al., 2014). Furthermore, the adaptation to regional scales of techniques that dynamically sample growing modes, mainly breeding, has been attempted (Thanh et al., 2016; Hermoso et al., 2020b). Additionally, the use of 3D variational perturbations to generate small-scale perturbations blended with large-scale perturbations from a global ensemble has also been explored (Keresturi et al., 2019). In any case, ensemble data assimilation systems, based on conventional observations (i.e., surface data, radiosondes, buoys, among others), and also remote sensing information, such as radar information (e.g., Wheatley et al., 2014; Wang and Wang, 2017; Degelia et al., 2018; Carrió et al., 2019) or satellite derived products (e.g., Romine et al., 2013; Velden et al., 2017; Zhang et al., 2018; Okamoto et al., 2019), constitute the usual initialization method for meso- and convective-scale EPS. The initialization and evolution of surface variables are also an important error source. Indeed, some operational ensembles introduce perturbations to sea surface temperature and/or soil variables to sample uncertainties associated with these fields (e.g., Tennant and Beare, 2014; Bouttier et al., 2016).

Convection-permitting EPS have been shown to improve over coarser resolution predictions (e.g., Duc et al., 2013; Schellander-Gorgas et al., 2017; Klasa et al., 2018; Gowan et al., 2018). However, convective ensemble forecasts are usually underdispersive, that is, ensemble spread is lower than forecast errors (e.g, Clark et al., 2011; Vié et al., 2011; Schwartz et al., 2014; Romine et al., 2014; Chen et al., 2018). The effective prediction of deep convection is complex due to the nonlinearities that dominate in small-scale processes (Hohenegger and Schär, 2007; Melhauser and Zhang, 2012; Sun and Zhang, 2016). This issue severely restricts the predictability horizon of socially relevant aspects of high-impact phenomena, such as location, timing, and intensity of convection, to extremely short lead times of the order of a few hours (e.g., Cintineo and Stensrud, 2013; Surcel et al., 2015; Zhang et al., 2015, 2016). This limited predictability is in turn highly case-dependent (Frogner et al., 2019), and it is strongly conditioned by the ensemble generation strategy as well as local and case-dependent characteristics. For instance, Bachmann et al. (2019, 2020) identify an increase in predictability related to the presence of orography and the use of radar data assimilation.

The difficulties associated with convective-scale forecasting are particularly severe in the Mediterranean region. The area is frequently affected by heavy precipitation events linked to deep convective activity (Ramis et al., 1998; Homar et al., 2002; Nuissier et al., 2008; Fiori et al., 2014; Faccini et al., 2016; Lorenzo-Lacruz et al., 2019, among others), especially during late summer and autumn (Llasat et al., 2010). The relatively warm sea during this time of the year acts as a heat and moisture source which, in combination with upper-level cold disturbances, generates convective instability that favors the triggering of deep moist convection. Furthermore, the Mediterranean Sea is surrounded by prominent mountain ranges (e.g., Atlas, Pyrenees, Alps) and upland areas located near the coast, which play an important role in the development of convective activity (e.g., Ramis et al., 1998; Ducrocq et al., 2008). Indeed, complex topography poses a supplementary challenge to precipitation forecasting as has been shown in multiple studies over different regions (e.g., Goger et al., 2016; Li et al., 2016; Yáñez Morroni et al., 2018). Moreover, the presence of many small-to-medium size catchments in the area with hydrological response times of a few hours, combined with the high population density of the Mediterranean coastal areas make this region highly vulnerable to flash flooding. Hydrometeorological forecasting chains can extend the period to take protective action beyond the short response times of small catchments, thus providing civil protection agencies with valuable information in order to mitigate the devastating effects of flash floods. Accurate high-resolution quantitative precipitation forecasts (QPFs) are required to produce useful flash flood forecasts. The characteristics of the watersheds in the area demand very precise forecasts of intensity and location of rainfall. However, the chaotic nature of convective precipitation critically hampers the accuracy of QPFs at the scale of interest.

Although initial and boundary condition perturbations can generate substantial diversity, and constitute the core of many studies aimed at improving short-range probabilistic forecasts (Peralta et al., 2012; Johnson and Wang, 2016; Keresturi et al., 2019; Hermoso et al., 2020b, among others), it is widely acknowledged that model error must also be considered to comprehensively account for uncertainties at meso- and convective-scales. This important source of uncertainty is related to the representation of specific physical processes, mainly those not explicitly resolved by the model. Different strategies have been proposed to sample uncertainty in model formulation. One of them consists of the combination of multiple models from distinct operational centers. The rationale behind this method is that the combination of the most advanced and accurate models developed at multiple centers can adequately reflect uncertainties present in model formulation. Although positive results have been obtained with this methodology, even at regional scale (Hou et al., 2001; García-Moya et al., 2011; Beck et al., 2016), this strategy requires substantial human resources to simultaneously maintain various models. Considering that the major source of model error is the representation of subgrid processes, a more popular elementary approach is to build a multiphysics ensemble. The technique is based on using multiple combinations of physical parameterization schemes, across the ensemble. This methodology shows good performance in terms of ensemble diversity and skill (e.g., Hacker et al., 2011b; Berner et al., 2011; Wang et al., 2011; Amengual et al., 2017). However, this approach presents some downsides: (i) each ensemble member has a different climatology and error characteristics, and (ii) its inclusion in operational tasks is costly as it requires the development and maintenance of various parameterization schemes, which in turn need to be physically consistent with each other. A variant of this method is the multiparameter approach (e.g., Hacker et al., 2011a; Yussouf and Stensrud, 2012; Duda et al., 2017) in which only specific parameters are modified among different ensemble members. Even though this approach allows a more specific representation of uncertainty and suppresses the requirement of dealing with distinct schemes, it generally displays a poorer performance compared to the classical multiphysics method (e.g., Berner et al., 2015).

An additional standpoint, which is receiving increasing attention recently, consists of inserting stochastic perturbations as a way to represent subgrid model error and introduce spread (Berner et al., 2017). The most popular techniques implemented in operational applications are the Stochastic Perturbed Physics Tendency (SPPT; Buizza et al., 1999) and the Stochastic Kinetic Energy Backscatter (Shutts, 2005). In the latter, stochastic perturbations are applied to potential temperature and streamfunction tendencies in order to account for the dissipated kinetic energy that interacts with the flow at the resolved scales, while in the former, the physical parameterizations tendencies in the state variables equations are perturbed. This is achieved by multiplying the tendency by a random perturbation which is spatially and temporally correlated. The characteristics of the random pattern (spatial and temporal decorrelation and variance) can vary depending on the application. For instance, in the original development of Buizza et al. (1999), random numbers were constant within boxes of certain side and over determined time steps and with diverse patterns for each state variable. The version currently operationally implemented at ECMWF uses a prescribed space and time correlation, includes a tapering to 0 for the lowest and top model levels and applies the same pattern to all state variables (Palmer et al., 2009). A similar version has been used in research studies with the Weather Research and Forecasting (WRF) model, although with no tapering (Berner et al., 2015). Even though the SPPT scheme has shown good performance (e.g., Berner et al., 2011), it presents some physically questionable aspects. The perturbation is applied to the full physics tendency, which generates several troublesome results: (i) the effect of the perturbation is zero when the net physical tendency is zero, although contributions - and thus, errors - from specific parameterizations may be high, (ii) error characteristics
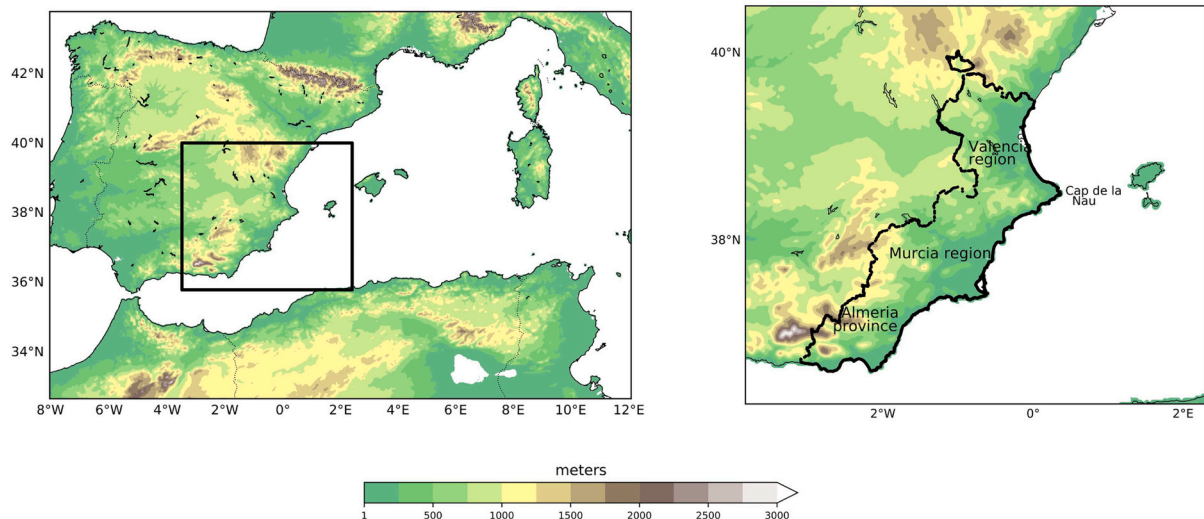
Fig. 1. Computational domain used in WRF simulations (left panel). The study area is enclosed by a thick black line. A zoom of the study area is shown in the right panel, indicating the geographical locations mentioned in the text.

are assumed to be equal among the different parameterizations, and (iii) the perturbation is not applied when and where uncertainty is originated (i.e., within the parameterization formulation), but at a later stage. Furthermore, this stochastic procedure to sample model error can generate numerical instabilities, mainly at the boundary layer (Lock et al., 2019).

Distinct suggestions have been made to solve some of the aforementioned issues of SPPT. The random pattern can be applied to specific parameters of each parameterization scheme, thus targeting the perturbation at the process level. This technique is called Stochastically Perturbed Parameters (SPP; Jankov et al., 2017). Although some positive results are reported (improvement in surface variables forecast at ECMWF, Ollinaho et al., 2017), SPP generally produces an insufficient spread (Christensen et al., 2015; Jankov et al., 2017), so that SPPT usually outperforms SPP. However, recent progresses on SPP have led to a similar performance of this technique compared to SPPT (Lang et al., 2021). Furthermore, the random parameters method designed at the UK Met Office (Bowler et al., 2008; McCabe et al., 2016) is based on randomly perturbing selected parameters, so that they vary over the forecast lead time, but are kept fixed for the whole domain at a specific lead time. McCabe et al. (2016) describes a positive impact in spread and skill in surface variables when this procedure is applied. An additional proposed modification of SPPT introduces independent perturbations for each parameterization scheme to solve the problematic of cancelling tendencies. Christensen et al. (2017) applied this approach to the Integrated Forecasting System of ECMWF, finding an increase in spread in areas with strong convective activity. Wastl et al. (2019a) investigate the implementation of this idea for a regional EPS and Wastl et al. (2019b) combine this methodology with SPP, obtaining a significant improvement of spread, especially when convective activity is intense.

The optimal strategy to represent relevant uncertainties in convection-permitting ensembles remains unclear. The challenges associated with convective-scale forecasting and the need to alleviate the consequences derived from extreme weather events motivate research focused on all methods available to sample uncertainties. In this context, the aim of this study is contributing to answer this essential question regarding the optimal ensemble generation strategy for the convective-scale by focusing on model error. The potential of various stochastic schemes to generate skilful forecasts for a Mediterranean high-impact episode is tested and compared to the multiphysics technique. This approach is a popular methodology to account for model error, which is applied in numerous operational ensembles worldwide (e.g., Descamps et al., 2015; Du et al., 2015), and is also widely used in

the hydrometeorological forecasting context (Lagasio et al., 2017; Tian et al., 2019; Roux et al., 2020). In the analysis, special emphasis is put on certain local specificities, which are of the utmost importance due to the topographic and hydrological particularities of the area. The remainder of the paper is structured as follows: Section 2 describes the most important characteristics of the case study selected, in Section 3 the model and ensemble configurations used in the different experiments are detailed, results are presented and discussed in Section 4, and conclusions are provided in Section 5.

2. Case study

A heavy precipitation episode took place on 12 and 13 September 2019 over eastern Spain. Total accumulated precipitation reached 500 mm at some sites, and also intensities in excess of 30 mm in 10 min were registered at some rain gauges. The intense rainfall and the subsequent widespread flash flooding that developed across the regions of *València*, *Murcia* and *Almería* (Fig. 1) produced devastating effects including 7 fatalities and economic costs in excess of 425 milion EUR (Consorcio de Compensación de Seguros (CCS), 2019).

The synoptic situation on the days prior to the onset of the episode displays the classical pattern that precedes heavy precipitation over eastern Spain (Romero et al., 1999): a deep trough over the north Atlantic at upper-levels, which moved southward towards the Iberian Peninsula during 10 and 11 September. This trough evolved to a cut-off, which remained roughly stationary during 12 and 13 September over north Africa. At low levels, the combination of a cyclonic system over north Africa and an anticyclone over the north Atlantic resulted in an easterly warm and moist flow which impinged the eastern coast of Spain between 11 and 13 September. As a result, the lower troposphere was convectively destabilized, the continuous moisture supply resulted in both high precipitable water levels and sustained water vapor flux convergence along the south-eastern coast of Mediterranean Spain.

Three differentiated phases of the episode arise from the analysis of observational data. (see Hermoso et al., 2021 for a detailed observational and physical analysis of this case study):

- Phase 1 (12 September 00–06 UTC): This phase is defined by the presence of an extended area of high equivalent potential temperature (EPT) and convective instability over the southwestern Mediterranean. A persistent convective band formed over the *València* region near *Cap de la Nau* promoted by the local orography of the

area. Total rainfall over 200 mm was registered in some locations during this phase.

- Phase 2 (12 September 06–18 UTC): The area of strongest instability extended southward and individual convective cells were triggered over the sea. These cells organized in a linear precipitation structure, which produced a record of 180 mm in two hours during the most active period of this phase.

- Phase 3 (12 September 18–13 September 12 UTC): It was the longest and most intense phase. The principal feature was the formation of a mesoscale convective system (MCS) that affected *Murcia* during the last hours of 12 and first hours of 13 September. The local topography anchored persistent precipitation systems, with maximum 10-min intensities exceeding 32 mm, at some rain gauges in *Murcia*. Additionally, western *Almería* was affected by convective bands with accumulations exceeding 200 mm in 6 h. From 13 September 05 UTC, the MCS moved northward affecting the *València* region.

The particularities of this case, specifically its long duration, the combination of different convection initiation mechanisms, with marine and orographic triggering, and the characteristics of the catchments embedded in the study area, with extensions in the range between 100 and 1000 km^2, make it an excellent testbed for convection-permitting ensemble strategies. The consideration of a single case study does not allow statistically significant conclusions to be drawn, but the results may point to promising directions in the prediction of extreme events. The statistical analysis is beyond the scope of this study and is left as future work.

3. Methodology

The performance of multiple stochastic techniques to predict the impacts of the 12–13 September 2019 exceptional heavy precipitation episode is investigated by using the WRF-ARW version 3.9.1 (Skamarock et al., 2008). A domain centered over the western Mediterranean is defined (Fig. 1). It has a horizontal resolution of 2.5 km and 50 vertical levels up to 50 hPa. The time step is set to 12 s. Given the long duration of the severe activity, the period is split in two independent 30-h simulations to benefit from an analysis update of the global model before the onset of phase 3. Indeed, attempts to cover the entire episode with a single simulation resulted in low skill during phase 3. Note that no data assimilation cycle is applied in these experiments. Therefore, the first simulation covers the period 11 September 18 UTC - 13 September 00 UTC, while the second one spans from 12 September 18 UTC to 14 September 00 UTC. Only the last 24 h of each simulation are evaluated, whereas the first six hours are left as a spin-up period.

Initial and boundary conditions are extracted from the dynamical downscaling of the ECMWF-EPS, which has a horizontal spectral resolution of O640 ($\sim$ 18 km at midlatitudes) and includes 91 vertical levels up to 0.01 hPa. Initial condition perturbations in ECMWF-EPS are built by combining ensemble data assimilation and singular vectors (Isaksen et al., 2010; Bonavita et al., 2017). These perturbations are added to the deterministic analysis, while model error is sampled by means of SPPT. A multi-scale pattern is used with a combination of three different variances, and temporal and spatial correlations which are specified for each individual pattern. Furthermore, a tapering function is introduced to reduce perturbation amplitude at the boundary layer and near the stratosphere (Leutbecher et al., 2017). From the 50 members of the ECMWF-EPS, the 10 initial and boundary conditions that display the greatest variability over the defined numerical domain are selected by using a k-means clustering. All ECMWF-EPS members are classified in 10 clusters based on their geopotential height, temperature and relative humidity at 500, 700 and 850 hPa. The 10 closest members to cluster centroid are considered. From these 10 initial and boundary conditions, multiple ensembles of 50 members are constructed with distinct techniques to account for model uncertainty (i.e., multiphysics and stochastic methods), which are thoroughly described below. Since the
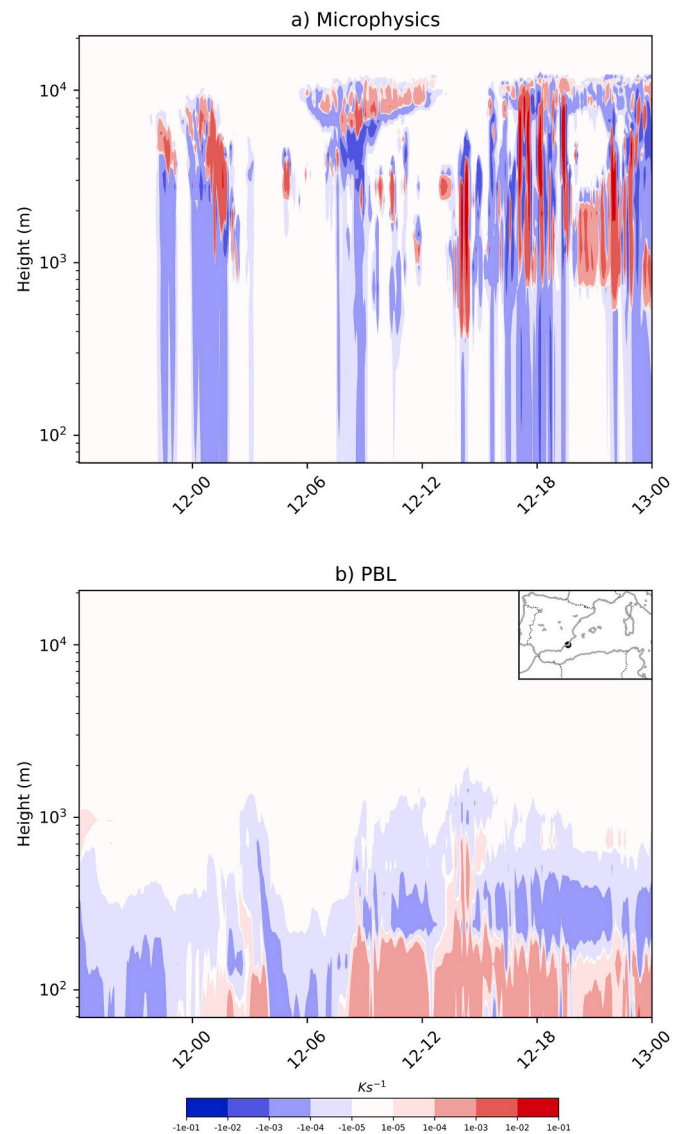

Fig. 2. Temperature tendencies from a) microphysics and b) PBL parameterizations for a MPS ensemble member with baseline configuration during the first simulation (11 September 18 UTC - 13 September 00 UTC). The point where the tendencies are computed is represented on the inner map of panel b).

objective of this study is investigating the impact of model error sampling, no additional initial condition perturbations are introduced, and thus differences among experiments are attributable to the model error methodology. However, a certain degree of variability in initial and boundary conditions is kept, so that the configuration presents some similarities to operational settings. It is noteworthy that these ensembles are prone to be underdispersive owing to the lack of small-scale initial perturbations. This could magnify the benefit of the most dispersive model perturbation strategy, when it may only be counteracting the lack of small-scale initial perturbations.

The baseline physical configuration includes the RRTMG radiation scheme (Iacono et al., 2008), the RUC land-surface model (Smirnova et al., 2016), the NSSL 2-moment microphysics (Mansell et al., 2010) and the Mellor-Yamada and Nakanishi Niino level 2.5 planetary boundary layer (PBL; Nakanisi and Niino, 2006). Convection is explicitly resolved with this set-up, and thus convective parameterization is switched off. For the multiphysics experiment this suite is one of the combinations applied.

Table 1
Summary of the microphysics and PBL schemes used in the MPS experiment.

Microphysics	PBL
NSSL 2-mom	**MYNN**
WSM6	MYNN
Thompson	MYNN
NSSL 2-mom	MYJ
WSM6	MYJ

Baseline configuration options are highlighted in bold.

3.1. Ensemble experiments

3.1.1. Multiphysics (MPS)

Considering the importance of convective processes involved in this episode, cumulus parameterizations would be a leading contributor to model uncertainty in coarse resolutions (e.g., Jankov et al., 2005). However, convection is mostly explicitly resolved in kilometric-scale simulations. In this context, microphysics and PBL are the parameterized mechanisms that most significantly affect deep moist convection. Microphysics accounts for phase transitions between water species, and thus this scheme can produce considerable effects on storm development. Indeed, recent studies emphasize the role of microphysics processes at convective scales in cases of orographic precipitation (Morales et al., 2019) or supercell development (Qiao et al., 2018). As a matter of fact, microphysics perturbations can yield impacts of the same order as initial condition modifications in some situations (Posselt et al., 2019). The PBL parameterization alters temperature and moisture in the lower troposphere, which is of the utmost importance for convective development (Coniglio et al., 2013). Moreover, turbulent processes in the boundary layer exert an important influence on convection triggering, especially under weak synoptic forcing (Kober and Craig, 2016; Keil et al., 2019).

As an illustrative example, microphysics and PBL temperature tendencies at a model grid point are depicted in Fig. 2. Microphysics tendencies are more intense in areas and times with active convective development (mainly during the unfolding of phase 3 at the end of this simulation), and exhibit lower spatial and temporal correlation than PBL tendencies. Moreover, as expected the effect of the PBL tendencies is only significant at the lowest 1000 m.

The 50 MPS ensemble members are generated by using 5 different physics suites combined with the 10 different initial conditions. Considering the high influence of microphysics processes, three schemes are included for this parameterization: NSSL 2-moment, WRF Single-Moment 6-class (WSM6; Hong and Lim, 2006) and Thompson (Thompson et al., 2008). Conversely, only two PBL options are selected: Mellor-Yamada and Nakanishi Niino (MYNN) level 2.5 and Mellor-Yamada-Janjic (MYJ; Janjic, 1994). The choice of these parameterizations is based on the compatibility among schemes and their suitability for kilometer-resolution simulations. Among the options that do not incorporate any of the schemes of the baseline configuration, only the combination of WSM6 and MYJ is adopted, as these options constitute the standard suite of the operational configuration used by the research Meteorology Group at the University of the Balearic Islands (http ://meteo.uib.es/wrf). A summary of the parameterizations considered is provided in Table 1.

3.1.2. Stochastically perturbed physics tendencies (SPPT)

The SPPT methodology consists of stochastically perturbing the physics tendencies originating from subgrid processes. It was originally introduced by Buizza et al. (1999), and has been used in many operational and research applications. The idea is that mechanisms occurring on unresolved scales should be sampled from a distribution rather than represented by the equilibrium mean (Berner et al., 2015). The version used in this study is the one implemented in WRF v3.9.1. Physics
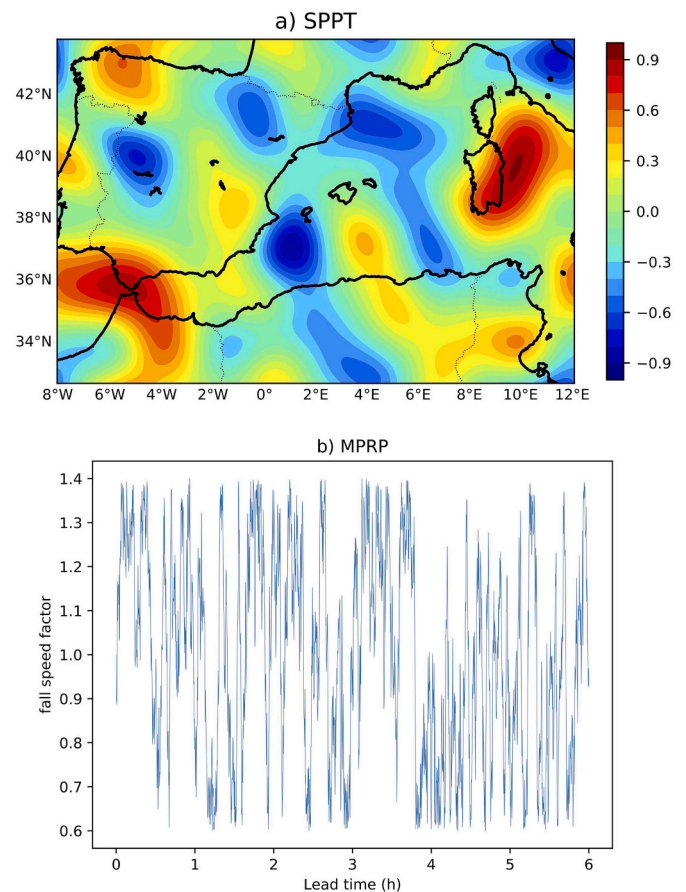

Fig. 3. a) Illustrative example of stochastic patterns for SPPT for a specific ensemble member and time step and b) time evolution of one microphysics perturbed parameter (fallfac) for the same ensemble member.

tendencies of potential temperature perturbation (θ), zonal and meridional wind components (u, v), and specific humidity (q) at each grid point are multiplied by a random number:

$$X'_{phys} = (1 + r(x, y, t))X_{phys}, \tag{1}$$

where X_{phys} stands for the tendency of variable $X \in T, u, v, q$ and $r(x, y, t)$ is the random pattern, which is a space and time dependent function. The perturbation is constant with height. Prescribed spatial and temporal decorrelation parameters are used to ensure the smoothness of $r(x, y, t)$. The random field $r(x, y, t)$ is defined as a 2D inverse Fourier transform:

$$r(x, y, t) = \sum_{k=-K/2}^{K/2} \sum_{l=-L/2}^{L/2} r_{k,l}(t) exp\left(2\pi\mathbf{i}\left(\frac{kx}{X} + \frac{ly}{Y}\right)\right), \tag{2}$$

where the total $(K+1)(L+1)$ wavenumber components in zonal (x) and meridional (y) directions in physical space are denoted by k and l. Spectral coefficients $r_{k,l}$ are computed through a first-order autoregressive process:

$$r_{k,l}(t + \Delta t) = \left(1 - \frac{\Delta t}{\tau}\right) r_{k,l}(t) + g_{k,l} \varepsilon_{k,l}(t), \tag{3}$$

which is controlled by the decorrelation time τ. $\varepsilon_{k,l}(t)$ is a Gaussian white noise process with 0 mean and variance equal to 1. Amplitudes $g_{k,l}$ correspond to a two-dimensional Gaussian in spectral space:

$$g_{k,l} = \left\{ \frac{\sigma_{k,l}^2 \left[1 - \left(1 - \frac{\Delta t}{\tau}\right)^2\right]}{2\sum_k\sum_l exp\left(-8\pi\kappa\rho_{k,l}^2\right)} \right\}^{1/2} exp\left(-4\pi\kappa\rho_{k,l}^2\right), \quad (4)$$

where κ is the decorrelation length, $\sigma_{k,l}^2$ the spectral variances and $\rho_{k,l} = \sqrt{k^2/X^2 + l^2/Y^2}$. The definition of $g_{k,l}$ is made such that the variance at any grid point σ^2, which is an additional user-specified parameter, corresponds to the total variance in spectral space.

Hence, the characteristics of the random pattern are established by the parameters τ, κ and σ. Being aware of the temporal and spatial scales involved in the convective systems developed during the episode, τ is set to 1 h and κ to 100 km (Table 3, Fig. 3). These values are similar to those used in other convective-scale studies (e.g., Romine et al., 2014). The variance is fixed to 0.25, which is the maximum value that does not generate numerical instabilities, and the amplitude of $r(x,y,t)$ is restricted to the interval $[-2\sigma,2\sigma]$. In some implementations of SPPT a tapering of the stochastic perturbations near surface and model top levels is introduced (Palmer et al., 2009; Bouttier et al., 2012) in order to prevent numerical instabilities. The WRF SPPT scheme does not include this tapering, but no numerical instabilities are found. As a result, a greater impact on surface variables can be achieved (e.g., Romine et al., 2014). An important drawback of this version is the exclusion of microphysics tendency perturbations. The effect of this parameterization in WRF-ARW is added after the state is updated with the other physical tendencies in order to preserve physical balances. Because of this different treatment of the microphysics processes, its inclusion in SPPT could result in a double counting of these perturbations (Jankov et al., 2019).

3.1.3. Microphysics random parameters (MPRP and SPPT_MPRP)

The relevance of microphysics process for moist convection dynamics (e.g., latent heat exchanges during phase changes of microphysics species), together with the lack of precise information at the subgrid scales, motivates the development of a stochastic scheme that copes with the uncertainties associated with these processes.

The proposed method to perturb microphysics follows the update of McCabe et al. (2016) of the random parameter scheme originally developed by Bowler et al. (2008). A set of parameters are perturbed so as to vary along forecast lead time, but are fixed for the whole domain. The time evolution of the random pattern p is governed by a first-order regressive process, similarly to SPPT, although in this case, only the temporal variation is considered:

$$p(t + \Delta t) = \left(1 - \frac{\Delta t}{\tau}\right)p(t) + \varepsilon(t) \quad (5)$$

The values of the parameters are limited within a predefined range, based on physical properties. In order to avoid the introduction of biases in the case that the default value does not correspond to the center of the parameter range, the updated value of the parameters is adjusted such that $p = 0$ matches the default value, while $p = -1$ coincides with the minimum value, and $p = 1$, with the maximum. A boundary condition is applied at the limits (i.e., $p = \pm 1$) to avoid peaks in the distribution of p at the bounds. The modified pattern $p*$ is obtained by reflecting p into the $[-1,1]$ range:

$$p^*(t) = \begin{cases} -2 - p(t) & if \quad p(t) < -1, \\ 2 - p(t) & if \quad p(t) > 1, \\ p(t) & otherwise. \end{cases}$$

The random process $\varepsilon(t)$ is formulated in terms of the autocorrelation $\phi = 1 - \Delta t/\tau$ (McCabe et al., 2016):

$$\varepsilon(t) = 2\chi(t)\sqrt{1 - \phi^2}, \quad (6)$$

where $\chi(t)$ is a random number obtained from a uniform distribution in

Table 2
Summary of the ensemble configurations tested.

Experiment	Description
MPS	Microphysics and PBL multiphysics
SPPT	Stochastic perturbations to total physics tendencies (except microphysics)
MPRP	Stochastic perturbations to microphysics parameters
SPPT_MPRP	Combination of SPPT and MPRP experiments

Table 3
Coefficients used in the stochastic schemes.

Scheme	κ (km)	τ (h)	σ^2
SPPT	100	1	0.25
MPRP	–	0.5	–

the interval $[-1,1]$. Given the mean duration of the influence of microphysics in convective processes (Fig. 2), the decorrelation parameter is fixed at 30 min (Table 3, Fig. 3).

Some parameters of the NSSL 2-mom scheme are perturbed in order to account for uncertainties in important microphysics processes for convective-scale forecasting. The parameters selected are described below and a summary is presented in Table 4:

- Cloud condensation nuclei concentration (CCN): The perturbation of this quantity is motivated by the specificities of the Mediterranean region. Its geographical location facilitates the intrusion of air masses of different origin and aerosol content. Thus, incursions of air with high particle content originated in the Saharan desert, continental air masses from central Europe or clean maritime air stemming form the Atlantic are common (e.g., Prospero, 1996; Clarke et al., 1997). This variability in CCN can have a substantial influence on precipitation processes (e.g., Hudson and Mishra, 2007; Barthlott and Hoose, 2018; Keil et al., 2019), although the impacts are variable and largely depend on the particular environmental conditions (White et al., 2017).
- Graupel and hail fall speed coefficient (graupel/hail fallfac): This factor modifies the fall speed of ice particles. This alteration can have an influence on mixing ratios and melting at low levels, affecting downdrafts and cold pool intensity (Falk et al., 2019). Furthermore, including perturbation to hydrometeors fall speed led to improvements in convective scale precipitation forecasts (Wang et al., 2020).
- Saturation percentage for initial cloud formation (ssmxinit): This parameter controls the number of activated CCN that can generate cloud droplets (Phillips et al., 2007). The number of activated CCN is an important property for convective clouds, as it controls the rate of droplet growth with cloud depth and the conversion into precipitation-sized particles (Freud et al., 2011).

Based on this design, two different ensembles are produced: (i) MPRP, in which only the microphysics scheme is stochastically perturbed following the above methodology, and (ii) SPPT_MPRP, wherein these perturbations to microphysics are combined with the sampling of uncertainties from other parameterizations by means of SPPT. A summary of all configurations tested is provided in Table 2. The former experiment is aimed at examining the isolated influence of microphysics perturbations to effectively introduce diversity, while the latter is

Table 4
Microphysics parameters perturbed in the MPRP scheme.

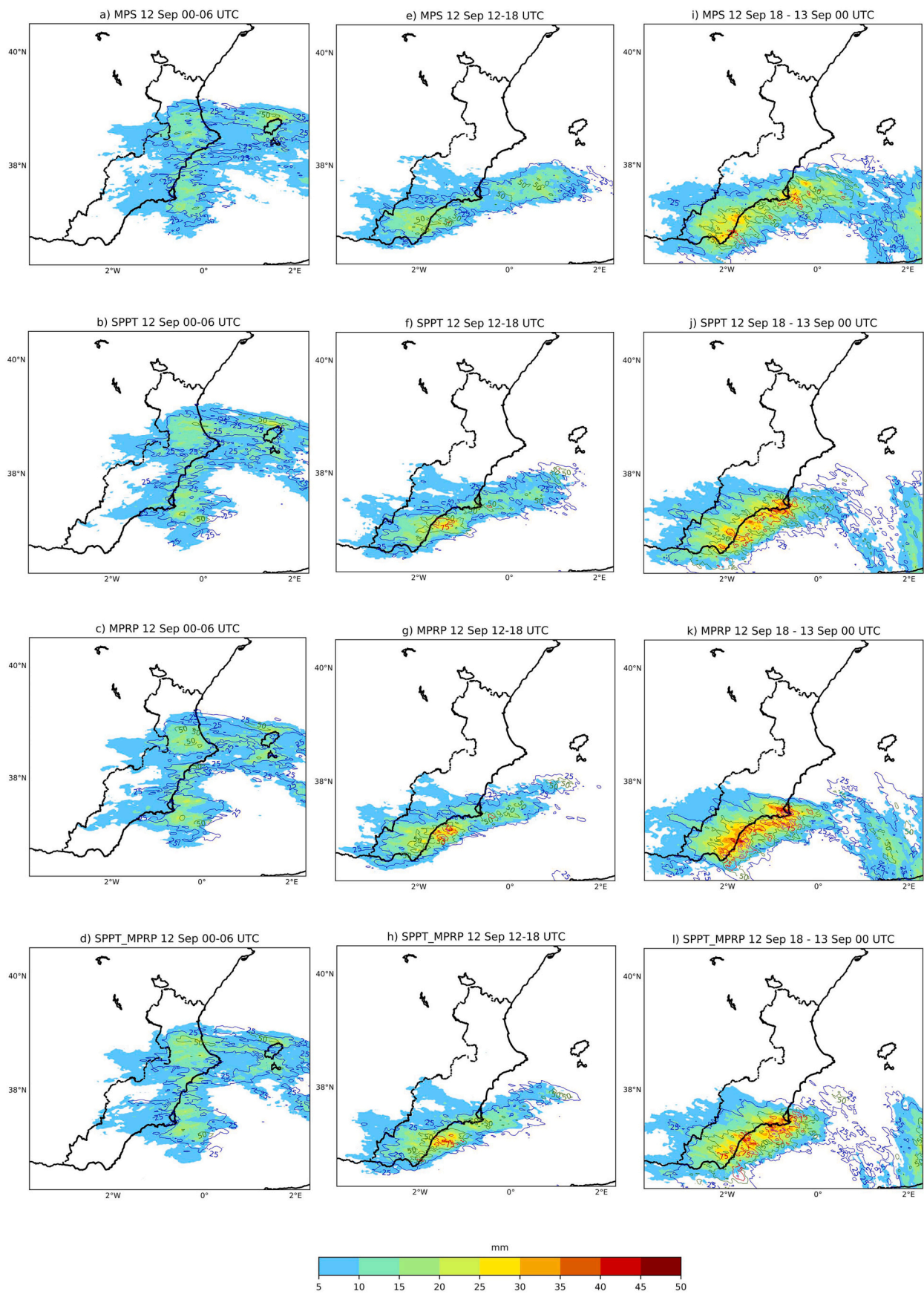| Parameter	Min	Default	Max
CCN	0.4E+9	0.5 E+9	1E+9
graupel/hail fallfac	0.7	1	1.3
ssmxinit	0.2	0.4	0.6

Fig. 4. 95th percentile of maximum hourly accumulated precipitation in 6-h periods (contours) and interquartile range (shaded) for the 12 September simulations (phases 1, 2 and beginning of phase 3): 00–06 (a-d), 12–18 (e-h) and 18–13 September 00 UTC (i-l).
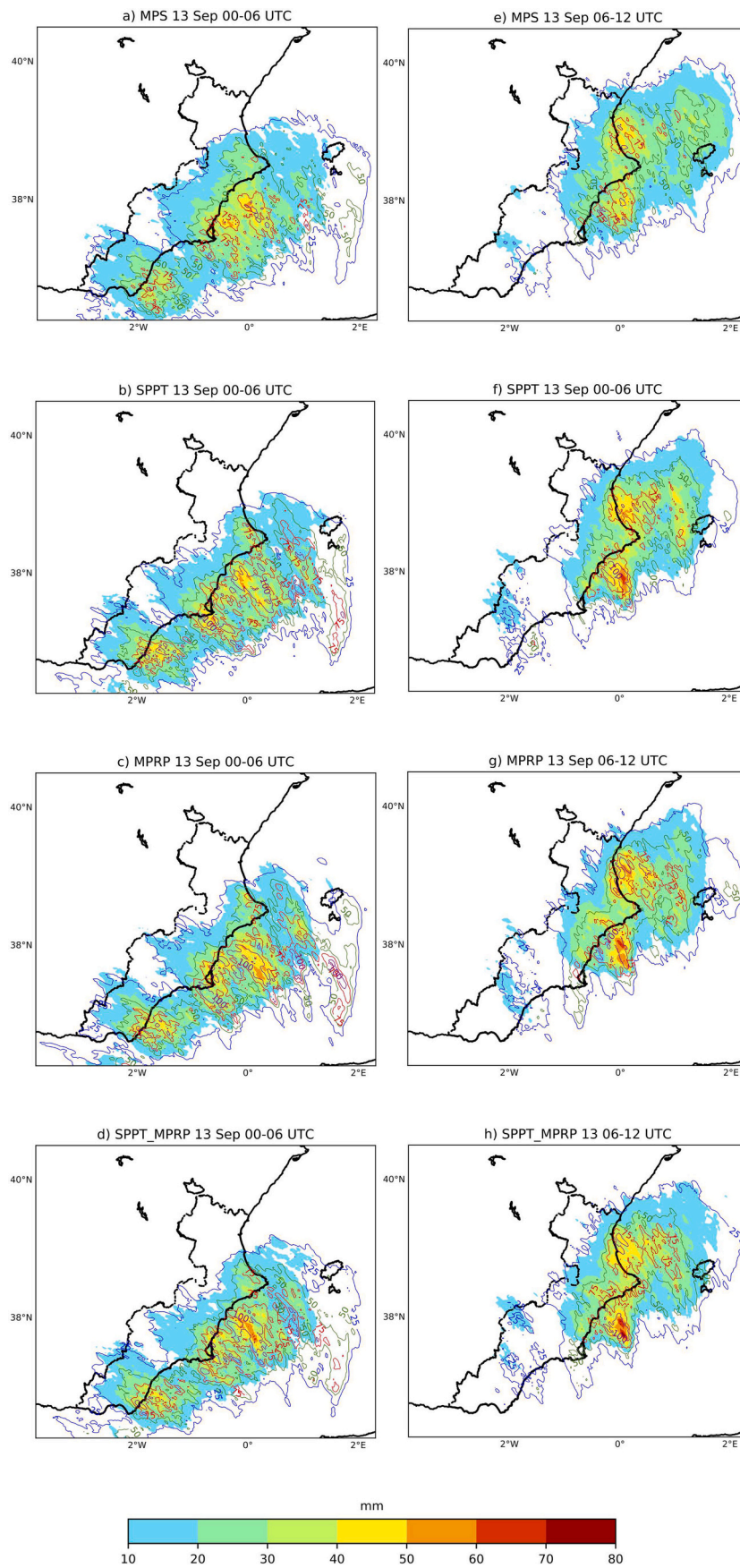
Fig. 5. As in Fig. 4 for the 13 September simulations (phase 3): 00–06 (a-d), and 12–18 (e-h).

designed to ascertain the mixed impact of perturbations from multiple error sources. It is noteworthy in the highly nonlinear regime of the convective-scale, that the sum of the individual effects of each perturbation may not produce an additive result.

3.2. Verification scores

Different probabilistic verification scores are used to assess the ability of the stochastic ensembles proposed to adequately sample the probability density function of scenarios consistent with the uncertainties. In this sense, we focus on socially relevant aspects of the episode, and thus the verification is directed to rainfall. In particular, 3-h accumulated precipitation is considered and three distinct thresholds are set (20, 35 and 50 mm). These values are selected as a compromise between using extreme accumulations and getting statistically significant results. Radar estimated precipitation fields are used for the verification. This information is derived from the reflectivity volume-scanning of the three radars located in the study area (*València, Murcia* and *Almería*), and calibrated by means of automatic rain gauge data. Further details on this procedure are explained on Hermoso et al. (2021). Radar data is used instead of station data due to the high observational density of the radar dataset over the study area. Observation errors are not considered in the verification.

The model grid length used is 2.5 km, but the effective resolution (i. e., the resolution in which structures are predictable) is lower. The skill of convection-permitting precipitation forecasts at grid scale is generally low (e.g., Ebert, 2008). The neighborhood maximum ensemble probability (Schwartz and Sobash, 2017) is applied to mitigate this problem. In this approach, a grid point has a value of 1 (i.e., occurrence of the event) if the binary event occurs at any grid point contained within a predefined distance. This method is more appropriate for severe weather than averaging a binary value across the neighborhood given the low frequency of extreme events. A square of 20 km side centered on each grid point is used. This spatial scale is of the same order of the extension of the watersheds contained in the study area, and thus is suitable to assess the skill of the different methods considered.

The Brier score compares the forecast probability of a binary event extracted from the ensemble to the occurrence of the event for each available observation (Jolliffe and Stephenson, 2003). The score is negatively oriented, with values ranging from 0 (perfect forecast) to 1. However, note that with the application of the neighboring approach, a perfect score can be obtained with spatial shifts up to 20 km in the forecasted precipitation field. An advantage of this score is that it can be decomposed in three independent components that represent different properties of the ensemble: reliability, resolution and uncertainty.

In order to present a better comparison between the different methodologies, the Brier score is transformed to a skill score, taking the MPS ensemble as reference:

$$BSS = 1 - \frac{BS}{BS_{ref}}. \tag{7}$$

Null BSS indicates an identical skill, while positive (negative) values imply an improvement (degradation) with respect to the reference.

Additionally, Receiver Operative Characteristics (ROC; Mason, 1982; Mason and Graham, 2002) are computed, as this curve is widely used to evaluate the skill of probabilistic forecasts. It compares the hit and false alarm rates for distinct probability thresholds, and the area under the ROC measures the ability of the ensemble forecast to discriminate between events and non-events. A value of 1 is obtained for a perfect probabilistic forecast, while 0.5 corresponds to no skill above the random forecast.

The statistical significance of the verification scores computed across the study is assessed by means of a bootstrapping technique. 95% confidence intervals are generated from 1000 resamples with replacements. However, given the limited volume of data considered, this significance test can only be used to identify relatively important differences among experiments for the weather situation analyzed in this study.

4. Results

4.1. Ensemble characteristics

Before analyzing the performance of the different ensemble generation strategies proposed, their main features in terms of extremes and diversity in accumulated precipitation are inspected. 95th percentiles and interquartile ranges are used to quantify extremes and spread, respectively. Over the course of phase 1 (Fig. 4a–d), ensembles containing perturbations to microphysics yield the largest spread and higher rainfall amounts than MPS and SPPT, with 6-h precipitation in excess of 50 mm in the 95th percentile. However, a northward displacement error with respect to the observed position of the thin convective band is identified, and reaches near 50 km for the SPPT_MPRP.

During the unfolding of phase 2 (Fig. 4e–h), all ensembles render similar diversity and precipitation quantities over land, and only minor differences among experiments are detected over sea, where ensembles based on microphysics perturbations display the largest spread. At the end of phase 2, ensemble spread highlights the area of convective development over the sea, and MPS features more widespread variability than stochastic-based ensembles. At this stage, the differences between SPPT and the other stochastic ensembles are lower, with similar spread and maximum values located off the coast of *Almería*.

This behavior continues at the beginning of phase 3, while significant spread extends along the littoral of *Murcia* and *Almería* (Fig. 4i–l), where intense convective activity occurred. Diversity is larger for stochastic techniques, especially those acting on microphysics perturbations. The extension of the area with notable ensemble spread adjacent to the coastline reflects changes in precipitation location sampled by the ensembles, and the maximum variability is found transversal to the coastline in accordance to the observed evolution of storms. Indeed, the spread provides an uncertainty quantification of rainfall location and intensity.

During the first hours of 13 September, when the MCS was in its mature stage, clear differences between MPS and ensembles including stochastic perturbations are detected (Fig. 5). Both ensemble diversity and maximum 6-h precipitation accumulations are smaller for MPS. In addition, areas of larger spread lay offshore *València* and *Almería*, while stochastic methods provide larger diversity over land at southeast *Murcia* and *Almería*. This latter structure is consistent with the areas where deep moist convection developed, namely near *Murcia* owing to the presence of the MCS and *Almería*, which was affected by intense convective bands. Comparable results are reproduced during the northward evolution of the MCS at the end of the episode. The largest spread is found over the sea north and south of *Cap de la Nau*, and is higher for the stochastic ensembles.

The larger diversity generated by the stochastic schemes is consistent with the results of other studies (e.g., Romine et al., 2014; McCabe et al., 2016; Jankov et al., 2017; Gasperoni et al., 2020). Indeed, the areas with higher diversity reveal intensity and location variability among ensemble members and are consistent with the main features of the heavy precipitation episode. This result is particularly relevant for probabilistic hydrometeorological forecasts. An accurate representation of the uncertainty associated with rainfall intensity and location is crucial, due to the highly nonlinear effects in the hydrological response related to small variations in these quantities. Moreover, the stochastic procedure applied to the microphysics processes may be suitable to improve model error sampling for these kind of episodes characterized by prominent maritime activity and deep moist convection. The relevance of microphysics in heavy precipitation events involving deep moist convection is well established (e.g., Keil et al., 2019; Lupo et al., 2020), owing to multiple microphysical processes related to condensation and the presence of multiple frozen hydrometeors.
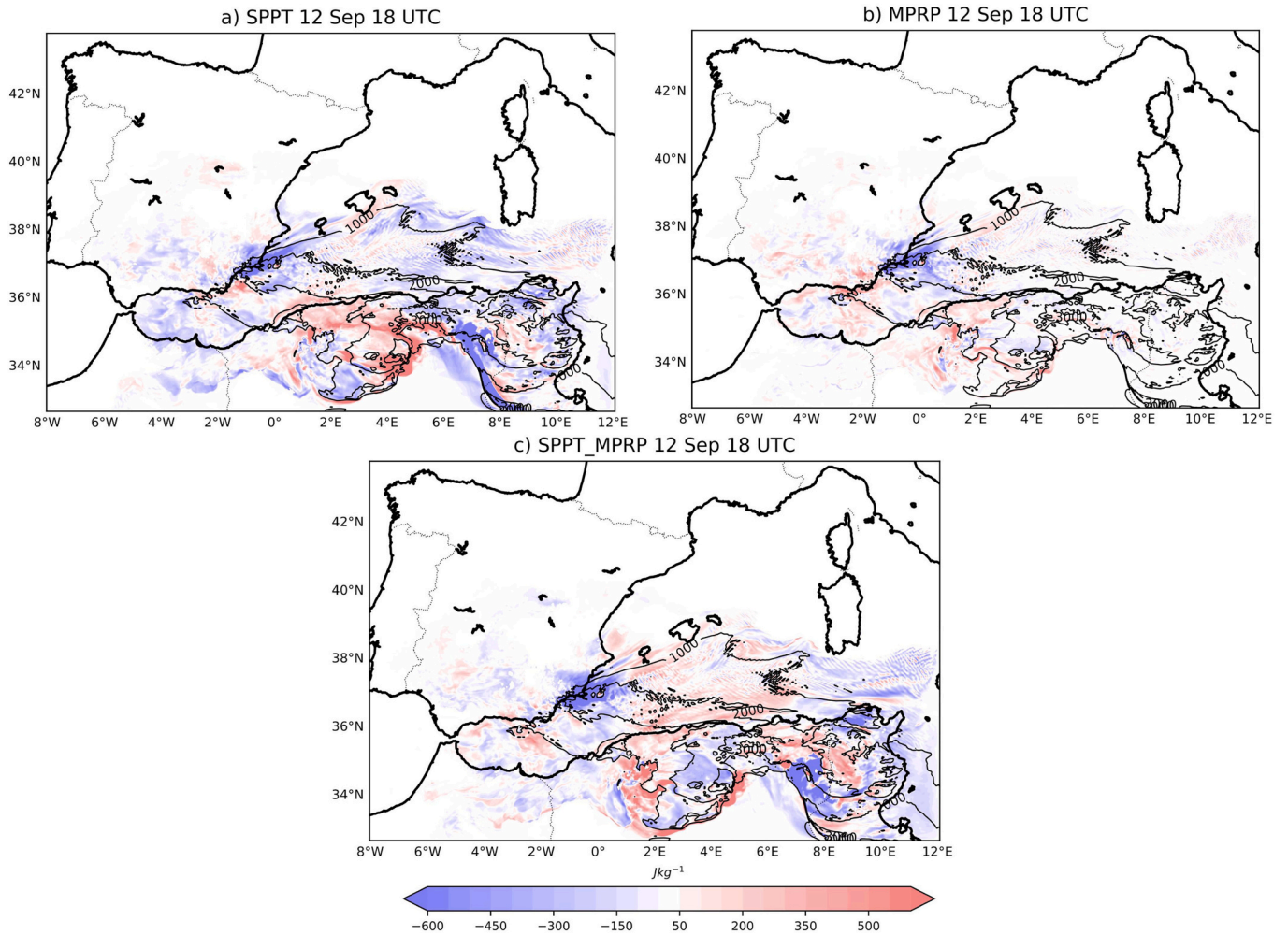
Fig. 6. CAPE difference (shaded) between a stochastically perturbed member and the corresponding unperturbed member (black contour) on 12 September at 18 UTC for the three experiments containing stochastic perturbations.

4.2. Environment characteristics

The efficiency of the proposed techniques to effectively introduce modifications to the thermodynamical environment is investigated. For the sake of simplicity, and as an illustrative example, an individual member of each stochastic ensemble is compared to the corresponding MPS member (unperturbed). Note that the differences between these simulations are attributable to the stochastic perturbations, because initial and boundary conditions, as well as physical parameterizations remain unchanged. Convective available potential energy (CAPE) is analyzed as a measure of instability. The value considered is the most unstable CAPE.

Stochastic modifications to microphysics parameters initially generate small-scale perturbations to CAPE in areas with convective instability, in which microphysics is active. However, these perturbations only grow in intensity and acquire higher spatial correlation in areas where deep moist convection is developing, mainly over the southwestern Mediterranean (Figs. 6 and 7). Conversely, SPPT introduces larger scale perturbations (more spatially correlated), regulated by the decorrelation length parameter, and these also intensify around areas of strong convective activity. Nevertheless, the spatial scale of perturbations outside the areas of most intense convective growth is larger than for the microphysics perturbations. Besides design differences, the variations between the two methods can be attributed to PBL processes, which have larger spatial correlation scales than microphysics. The combination of stochastic schemes introduces

perturbations of different spatial scales, which eventually evolve similarly to MPRP and SPPT perturbations in the neighborhood of active convective development (Figs. 6 and 7). This result is consistent with studies that investigate upscale perturbation growth from convective-scales, which is much enhanced in areas with deep moist convection (Selz and Craig, 2015; Zhang, 2019).

The spatial characteristics of the perturbation can be quantified using two parameters - $\log \rho$ and ω^2 - representing the amplitude and the variance of the perturbation, respectively. These quantities were introduced by Primo et al. (2008), and mathematically correspond to the geometrical mean of the perturbation ($\log \rho$) and the variance (ω^2):

$$log\rho = log\left(\prod_{i=1}^{N} |\delta x_i|^{\frac{1}{N}}\right), \tag{8}$$

$$\omega^2 = \frac{1}{N}\sum_{i=1}^{N}(log\delta x_i - log\rho)^2, \tag{9}$$

where N is the number of grid points with non-zero perturbation $|\delta x_i|$. For stochastic methods, δx_i is computed as the difference between the stochastically perturbed ensemble member and the corresponding member of MPS with the same initial conditions and physical parameterizations. For MPS, δx_i is computed as the anomaly with respect to the ensemble mean, as there is no unperturbed reference for this experiment. δx_i is computed considering all model levels below 5000 m and the
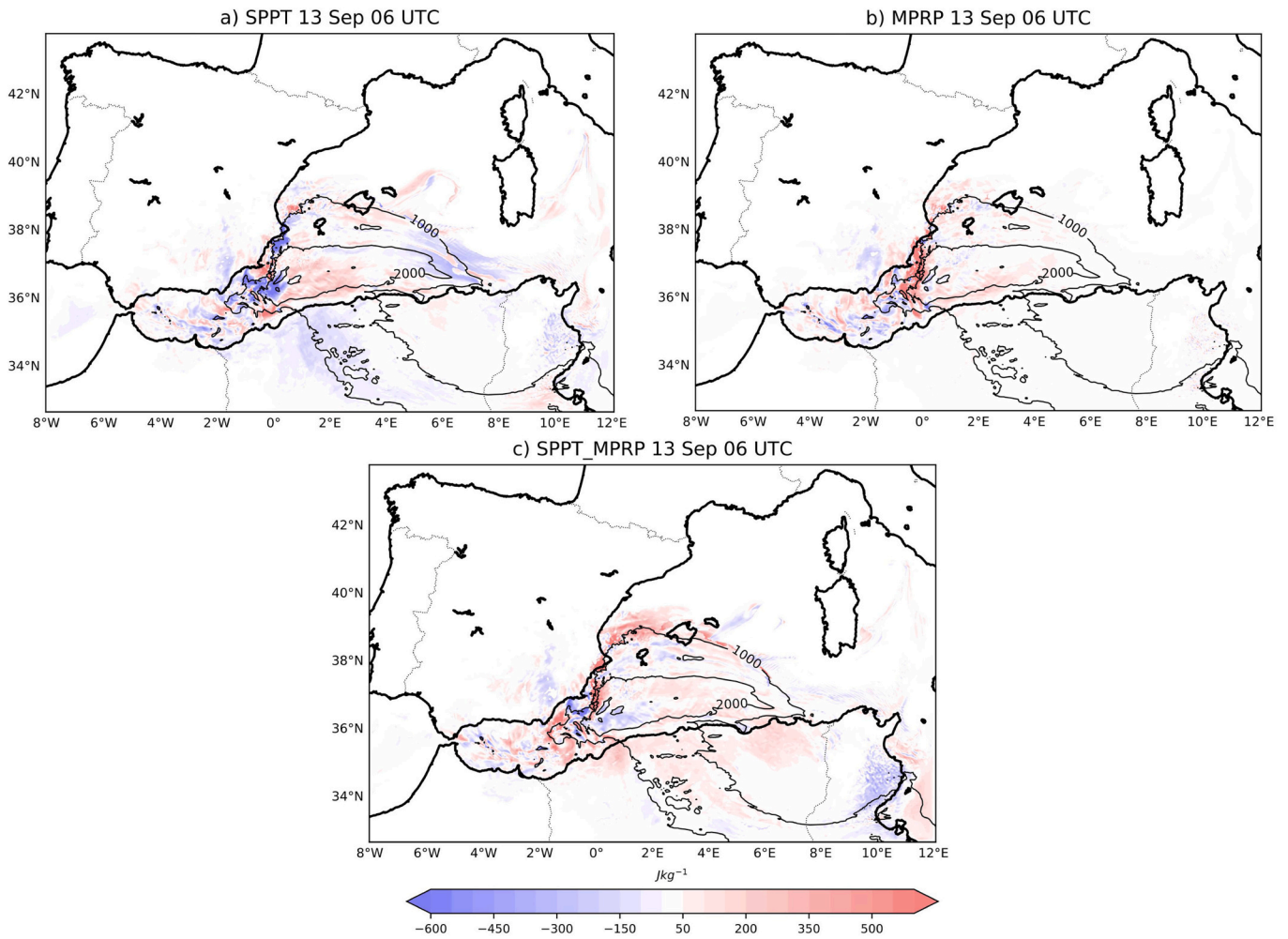
Fig. 7. As in Fig. 6 on 13 September at 06 UTC.

four state variables (T, u, v and q). Since there are more model levels near the surface, boundary layer perturbations are more represented in δx_i. Potential temperature and specific humidity perturbation are rescaled to ensure dimensionality consistence:

$$
T_{fact} = \sqrt{\frac{C_p}{T_{ref}}},
$$
$$
q_{fact} = L\sqrt{\frac{w_q}{C_p T_{ref}}},
$$

(10)

where C_p is the air specific heat capacity, L the latent heat of condensation, $T_{ref} = 300$ K, and factor $w_q = 0.1$ is introduced to level the magnitude of humidity perturbations with the rest.

Although ω^2 is defined as the variance of perturbation amplitude, it can be interpreted as a measure of localization. Admittedly, ω^2 does not provide a universal perturbation scale characterization, but the value of this quantity (i.e., perturbation amplitude dispersion) for a particular dynamical system - defined by the model and the domain numerical configuration - is related to its scale, and to the spatial correlations generated by the dynamics of the system (Primo et al., 2008). Low (high) ω^2 imply large (small) scale perturbations with respect to domain size.

The evolution of perturbations shows a rapid amplitude growth for MPRP perturbations, which are highly localized during the first hours of simulation (i.e., large ω^2). The two ensembles including SPPT display virtually identical development, starting from larger scale perturbations, conditioned to the spatial decorrelation scale parameter, and growing at

a smaller rate than MPRP (Fig. 8). The saturation level is higher for SPPT, as perturbation amplitudes are intense over larger areas compared to MPRP, which translates into higher log ρ and lower ω^2. This behavior is consistent with the conventional scale-dependent evolution of perturbations (e.g., Toth and Kalnay, 1993). Nonetheless, saturation variations between perturbations with different initial spatial correlations are slight and of minor importance over areas with deep moist convection. Therefore, diversity in log ρ originates in areas distant from convective activity. When SPPT and MPRP perturbations are combined, the former dominates the scale and amplitude of the full perturbation, so that negligible differences are obtained for log ρ and ω^2. MPS perturbations display the largest amplitude and spatial correlation, although the procedure to compute these perturbations is fundamentally different from the other experiments.

The modification of convective instability caused by stochastic perturbations has an influence on precipitation. As an illustrative example, changes produced in the above analyzed members with respect to the unperturbed MPS simulation are inspected and qualitatively compared to the 6-h accumulated rainfall errors (Figs. 9 and 10a). These fields are obtained by using radar estimated precipitation. Minor variations are identified during phases 1 and 2. However, they are substantial at the onset of phase 3, during the last hours of 12 September. The unperturbed member produces precipitation in areas of southern *Murcia* and *Almería*, with notable location errors with respect to observations. Stochastic perturbations introduce modifications in the forecasted precipitation field, with scales compatible with these localization errors (Fig. 9b–d). In this example, the methodologies investigated drive the most intense
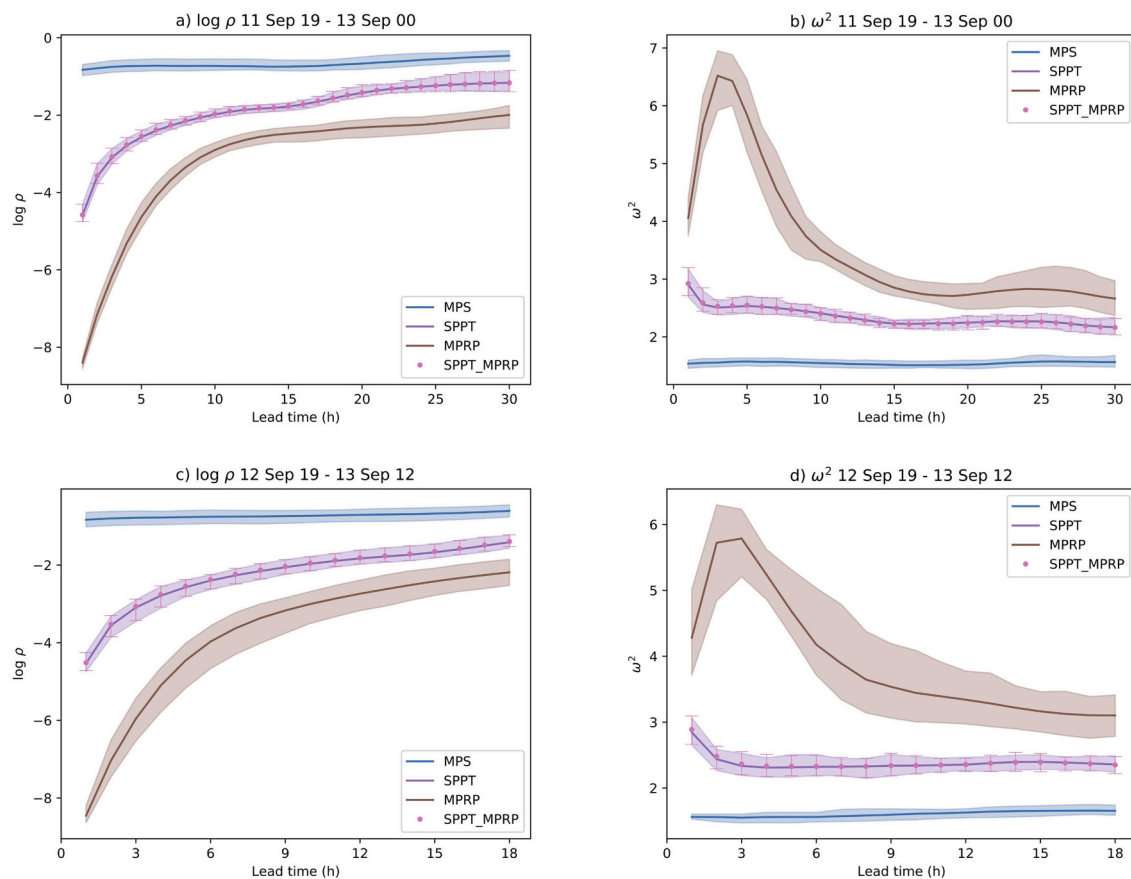
Fig. 8. Time evolution of log ρ (a, c) and ω^2 (b, d) for the first (a, b) and second (c, d) simulations for the five experiments. Solid lines denote the ensemble mean of these quantities, while the shaded area encompasses minimum and maximum log ρ and ω^2. Since the evolutions for SPPT and SPPT_MPRP virtually overlap, SPPT_MPRP values are represented by vertical bars.

rainfall areas southward and eastward, closer to observations. During the first hours of 13 September, SPPT and MPRP members produce a more accurate forecast around the *València* coastline and over southeast *Murcia* (Fig. 10b,c). In this case, the combination of different stochastic perturbations causes a degradation of the precipitation forecast in some areas of *València*, where excessive rainfall is already predicted by the unperturbed member, and an additional increment is introduced (Fig. 10d).

Therefore, stochastic perturbations tend to grow in areas with deep moist convection and introduce modifications to the thermodynamic environment. These changes produce alterations in precipitation of the same scale as the model errors, and thus this methodology has a great potential to sample precipitation uncertainties, including extremes. In addition to the useful information distilled from this analysis, a precise probabilistic verification must be conducted to comprehensively assess the performance of each ensemble strategy investigated.

4.3. Verification

4.3.1. Probabilistic scores

Given the promising diversity of outcomes generated by the stochastic strategies and the effective perturbation growth characteristics, a quantitive assessment is carried out by means of the probabilistic verification scores introduced in Section 3.2. Since they are only computed for an individual heavy precipitation episode, which implies a certain level of correlation between verification locations, the analysis of these scores is not generalizable to other cases in a strict sense.

During phases 1 and 2, multiphysics exhibits the lowest Brier score for all considered thresholds, and thus negative BSS are obtained for the

stochastic ensembles tested (Fig. 11a–c). The area affected by the linear precipitation structure is better captured by the multiphysics-based ensemble (Fig. 12). However, differences are only important at the 20 mm threshold. The first phase is strongly influenced by an elongated area of high EPT over the western Mediterranean. Uncertainties related to initial conditions play a primary role at this stage, and variability introduced by stochastic techniques is modest. Concerning Brier score components - reliability and resolution - modest variations among experiments are found, especially for the largest threshold.

These low contrasts between experiments are propagated to the second phase. This period is primarily governed by the southward evolution of the high EPT intrusion. At the beginning of this phase, low forecast probability is found over large areas of the occurrence of the binary event over the southern *València* region (Fig. 12a,b), resulting in low reliability. During the subsequent evolution, maximum values of probability are moderate (around 0.5–0.6). Thus, ensemble ability to distinguish between the occurrence and non-occurrence of the event is low, leading to low forecast resolution (Fig. 12c,d).

Additionally, ROC areas show better discrimination skill for stochastically perturbed ensembles for the 50 mm threshold during phases 1 and 2, especially for ensembles with microphysics perturbations, owing to higher hit rates and probability maxima. This fact is related to the larger rainfall extremes produced by these ensembles. Despite this increase in high values, areal accumulated precipitation is slightly lower than for the unperturbed simulations. By contrast, the skill in terms of discrimination is better for MPS at the low thresholds, due to a lower proportion of false alarms (Fig. 13a–c). Stochastic ensembles slightly increase the probability of the event in areas of non-occurrence and reduce the probability in areas of occurrence. Singularly, the false alarm
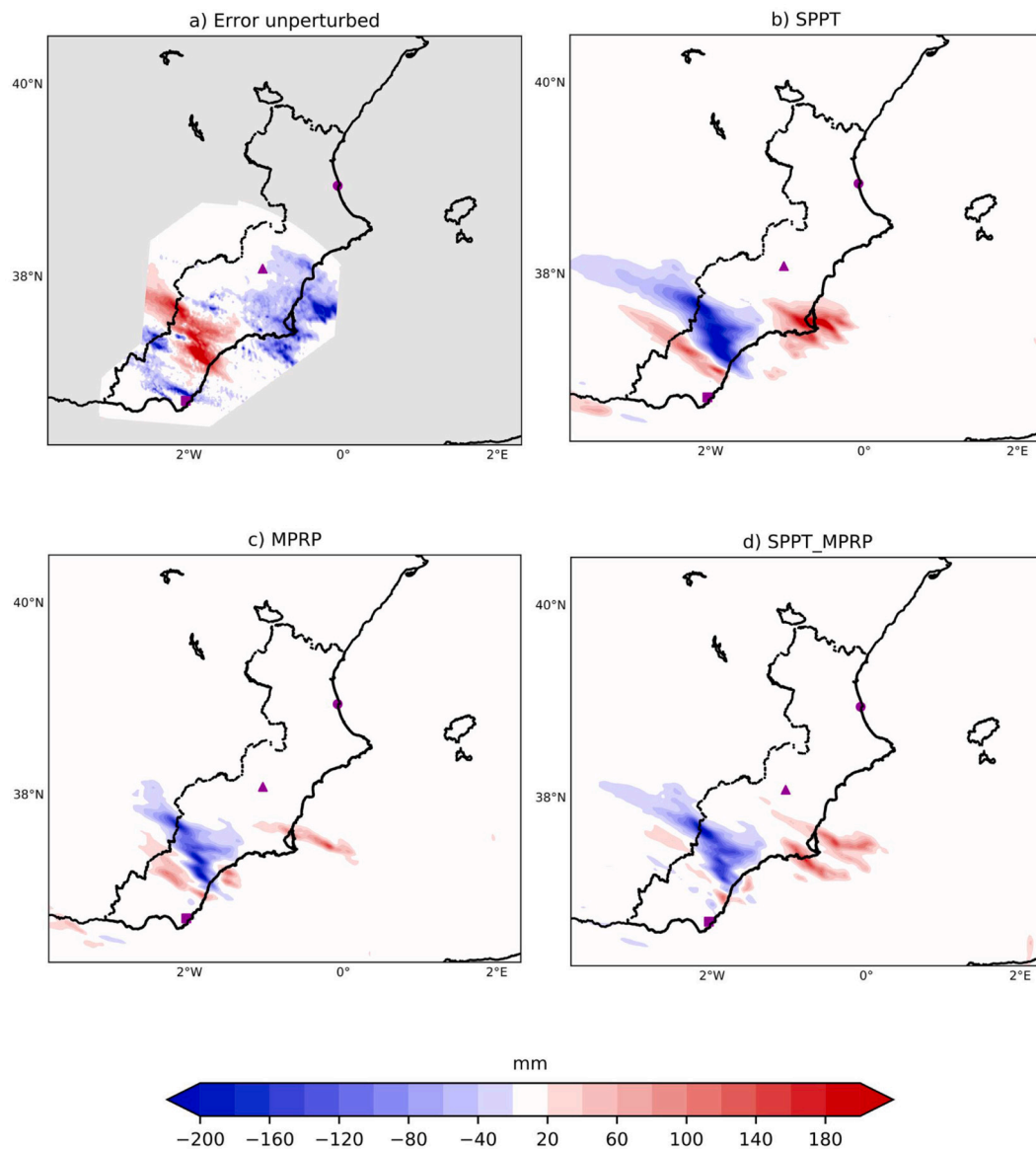
Fig. 9. a) Difference between 6-h accumulated precipitation (12 September 18–13 September 00 UTC) for an unperturbed member and radar estimated rainfall, and b)-d) modifications to the unperturbed member produced by the distinct stochastic techniques. Locations of *València*, *Murcia* and *Almería* radars are represented by a purple dot, triangle, and square, respectively. The area outside radar coverage is grayed out in panel a).

increment effect for MPRP is lower than for the other stochastic ensembles during phase 1. Regarding the first part of phase 2, hit rate and forecasted probabilities are low, while during the subsequent hours both hit and false alarm rates increase, with fewer false alarms for MPS. The stochastic modifications introduced are not sufficient to capture the formation of the convective line that distinguish this stage.

By contrast, during phase 3 differences between experiments are higher (Fig. 11). This period is characterized by the formation of a quasi-stationary MCS over an area with complex topography, which produced persistent and intense precipitation over *Murcia*. At this stage, model error is more influential than in the previous phases and significant differences between multiphysics and stochastic techniques are attained. During the last part of the first simulation, stochastic methods including microphysics perturbations yield the best performance, which is consistent with the intense effect of microphysics tendencies during this phase (Fig. 2). The area of highest probability of occurrence is more concentrated for ensembles with stochastic parameterizations (Fig. 14). Reliability is improved by a better correspondence between forecasted probabilities and observations. Moreover, larger probabilities are

forecasted, especially by stochastic ensembles (Fig. 14b), which leads to an increased resolution.

This improvement of stochastic approaches with respect to multiphysics extends to the first part of the second simulation, although in this case no relevant differences are obtained between the various stochastic strategies. At this time, variations are mostly attributable to resolution modifications. Indeed, forecasted probabilities > 0.9 are produced by MPRP over an elongated area of approximately 100 km affected by the MCS (Fig. 14d). Conversely, during the northward evolution of the MCS, skill is slightly better for multiphysics, mainly due to a better reliability. The introduction of stochastic perturbations produces a clear enhancement of ensemble performance in terms of discrimination between the occurrence and non-occurrence of the selected binary events for the period characterized by the quasi-stationary MCS (Fig. 13c–e), as MPS produces more false alarms compared to stochastic methods at this stage. In the course of the northward evolution of this system, differences among ensembles reduce.

The combination of multiple stochastic perturbations does not translate into a clear improvement of probabilistic forecast skill for this
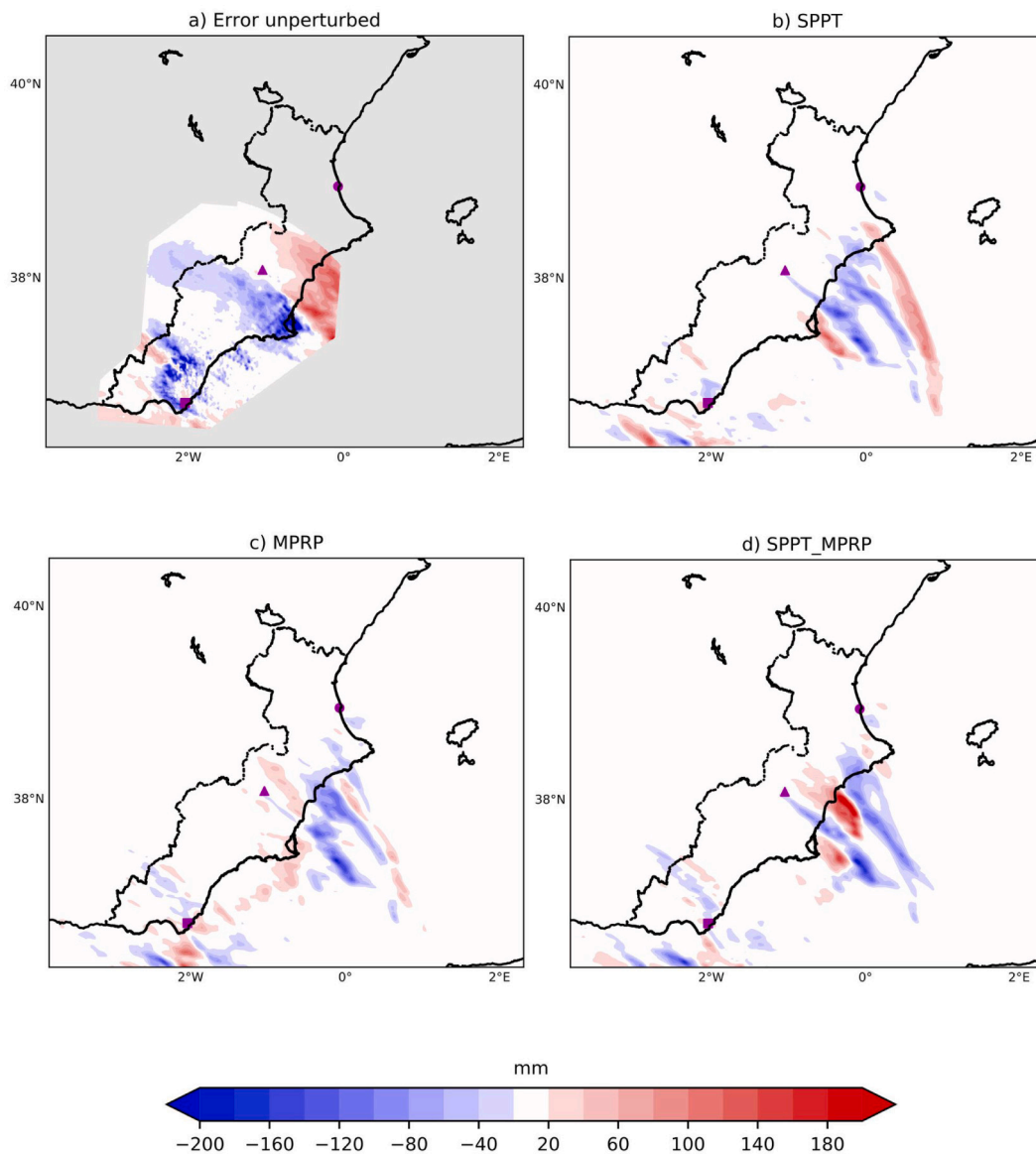
Fig. 10. As in Fig. 9 for 13 September 00–06 UTC.

episode. Although during the last hours of the episode, ensembles containing a combination of stochastic perturbations render the best efficiency - especially for high precipitation thresholds - this behavior is not uniform along the episode. This contrasts with the results of other studies of ensembles based on stochastic parameterizations (e.g, Jankov et al., 2017, 2019; Xu et al., 2020). The nonlinear interactions which operate at convective scale, especially in heavy precipitation cases can counteract the individual positive effects of each perturbation, resulting in a degraded forecast. Moreover, the fast perturbation growth characteristic at small scales can induce rapid saturation, which is not overcome by including additional perturbations. Furthermore, it must be taken into account that conclusions derived from an evaluation over an extended benchmarking period, as carried out in the above mentioned studies, may not be applicable over a particular case study.

4.4. Ensemble features at catchment scale

The quantitative scores provide a general overview of ensemble performance across the verification area. However, the particular characteristics of the watersheds present in the region demand the generation of accurate QPFs at remarkably small scales in order to be

able to estimate the potential flooding risk and taking appropriate protection measures. In this subsection, the forecasted maximum hourly precipitation intensities in 6-h intervals are analyzed and compared to rain gauge values placed inside eight small-to-medium sized catchments (Fig. 15). Owing to the limited number of basins, this evaluation is not centered on the ability of each ensemble to reproduce the probability density function compatible with the observed state, but on the ability of the ensembles to reproduce observed rainfall intensities at very local scales. Station data is used for this evaluation as it is generally more accurate than radar estimates and a smaller dataset is adequate for an analysis at specific basins.

During phase 1, the observed maximum intensity is reproduced by all ensembles at all the considered catchments, except *Beniarrés*, where a remarkable extreme value of almost 70 mm in 1 h was recorded (Fig. 16a). Ensemble diversity is similar in most watersheds, excluding *Moixent* and *Bellús*. At the former, microphysics perturbations increase spread compared to SPPT, while at the latter, stochastic techniques render greater diversity. This is in agreement with the general behavior previously identified. During the first part of phase 2, some members of the stochastic ensembles capture the high intensity registered in *Beniarrés*. Hence, the higher spread of these ensembles has also a
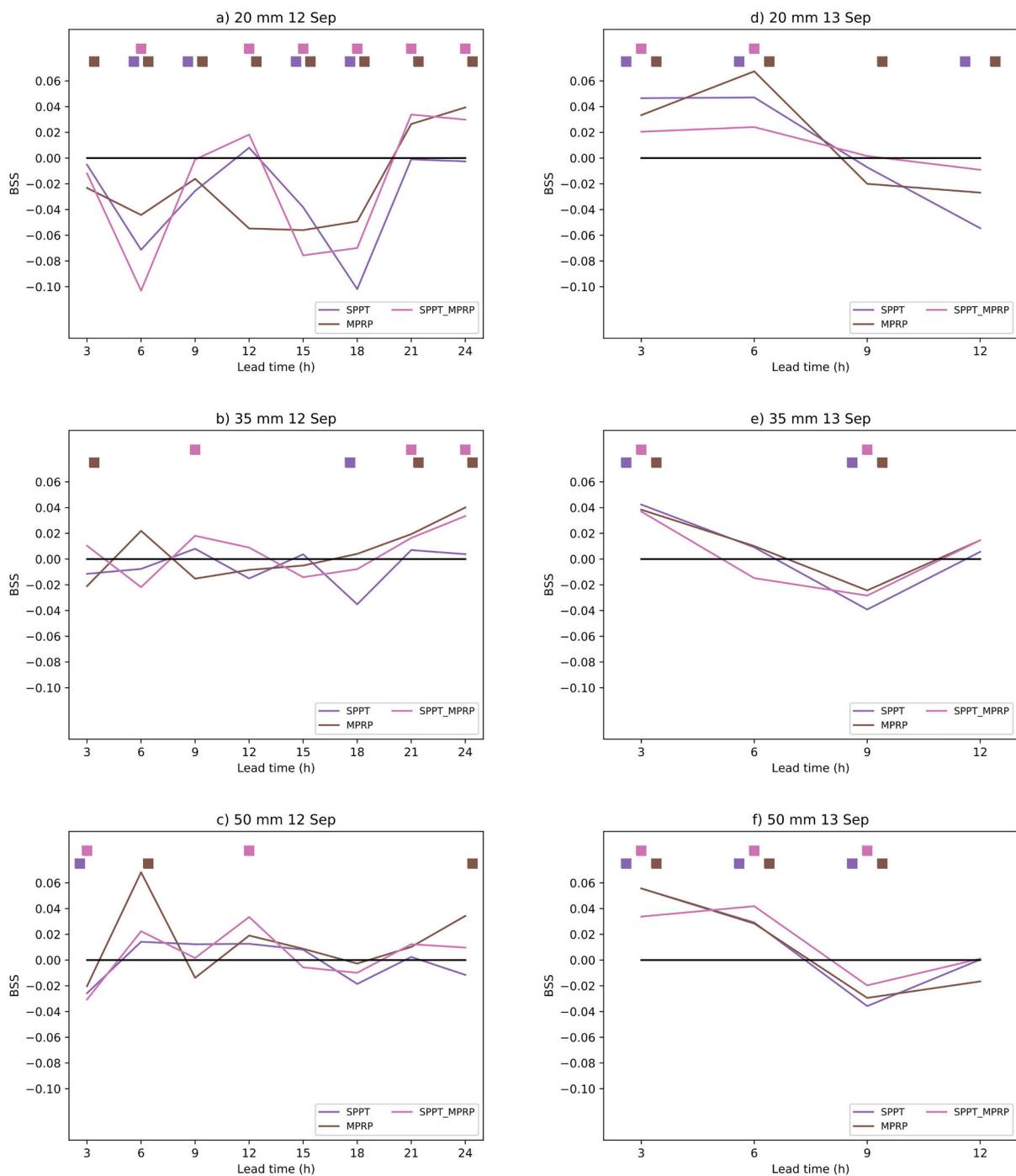
Fig. 11. BSS for 3-h accumulated precipitation as a function of lead time for the first (a-c) and the second simulations (d-f) for 20, 35 and 50 mm thresholds. Colored squares indicate BSS significantly different from 0 (i.e., significant improvement or deterioration with respect to MPS) at the 95% confidence level for the case study considered.

relevant impact at local scale. Nevertheless, diversity is large at some catchments (*Albujón, Benipila, Salada*), which would render some false alarms for this particular episode. In the subsequent evolution, the largest spread is localized over the central and southern basins, although none of the stochastic methods proposed is able to simulate the 15 mm h^{-1} recorded in *Salada*. It should be taken into consideration that this period corresponds to 18–30 h lead time for the first simulation, and for such long lead times, localization errors tend to be high. Thus, the frequency of misses and false alarms increases, especially at the small scales examined in this analysis.

During phase 3 (Fig. 16e, f), observed maximum intensities are

captured by some ensemble members in all the experiments. For the *Albujón, Benipila* and *Beniarrés* basins, the highest ensemble percentiles display significant extreme values > 60 mm h^{-1}, particularly SPPT. Nevertheless, the combination of this perturbation with other stochastic perturbations results in a reduction of these extremes, indicative of counteracting effects produced by each perturbation. Conversely, false alarms are obtained in some catchments where low rainfall rates were registered during the first hours. Stochastic perturbations reduce the excess produced by MPS in *Moixent*, while similar performance is obtained in *Bellús* (Fig. 16e). During this period, the major convective activity is located southward associated with the MCS in *Murcia* and the
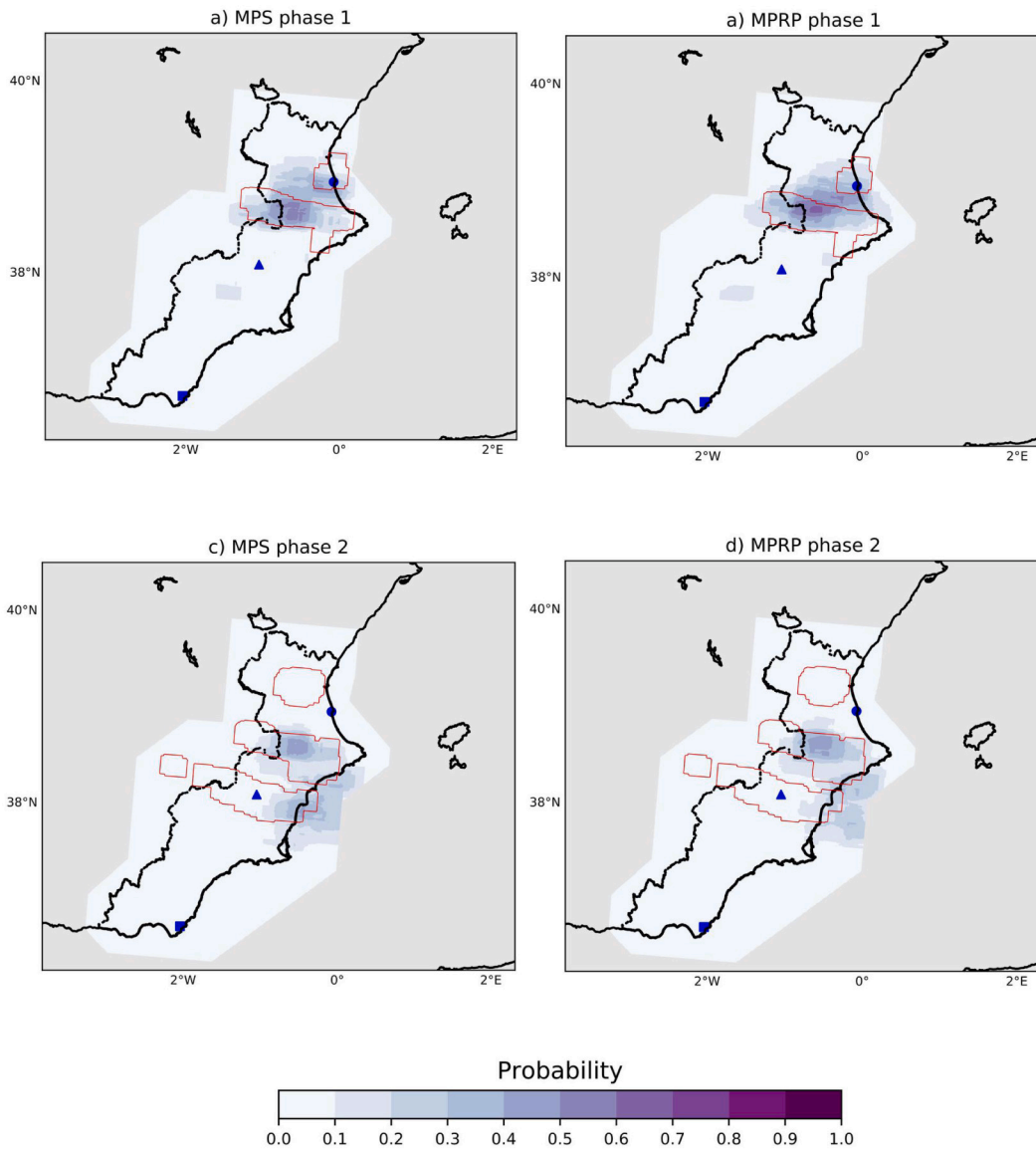
Fig. 12. Probability of 3h accumulated precipitation > 35 mm for MPS (left panels) and MPRP (right panels) on 12 September at a), b) 03–06 UTC (phase 1); c), d) 06–09 UTC (phase 2). Locations of *València*, *Murcia* and *Almería* radars are represented by a blue dot, triangle, and square, respectively. Red contours enclose areas with observed 3-h accumulated precipitation > 35 mm.
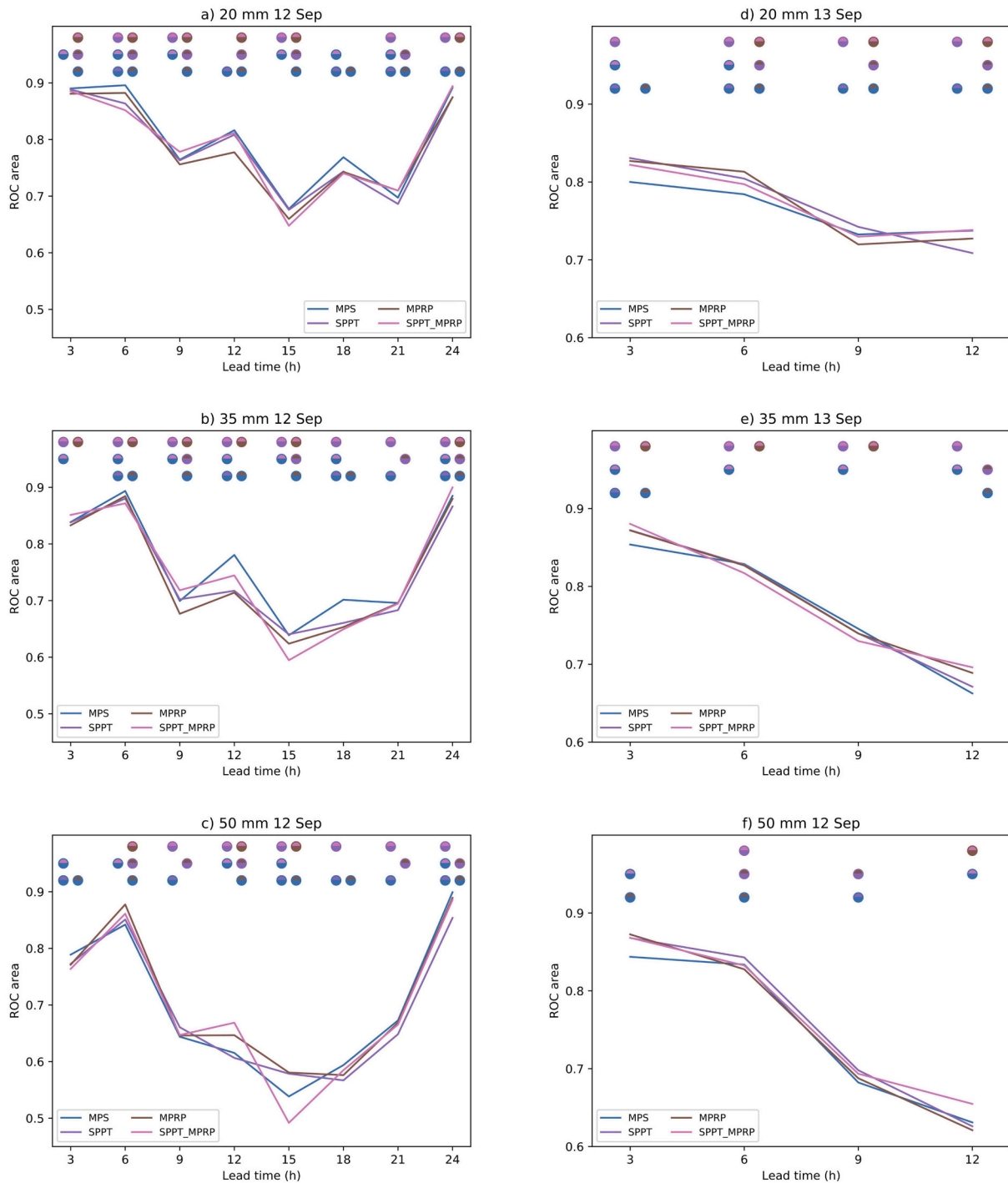
Fig. 13. As in Fig. 11 for the area under the ROC. Bicolored circles represent significant differences between experiments identified by those colors at the 95% confidence level computed as described in the text.
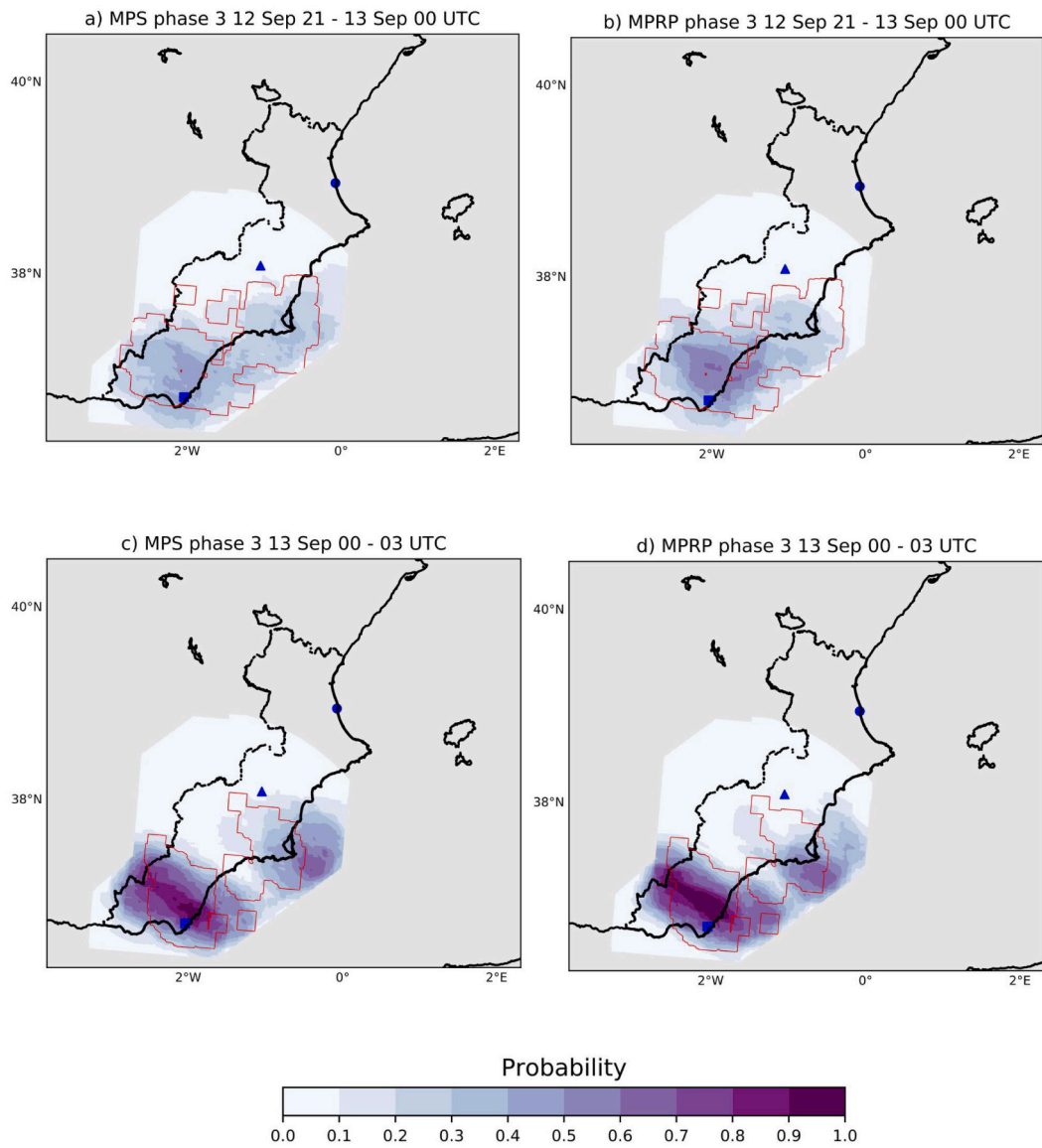
Fig. 14. As in Fig. 12 for phase 3: a), b) 12 September 21–13 September 00 UTC; c) d) 13 September 00–03 UTC.
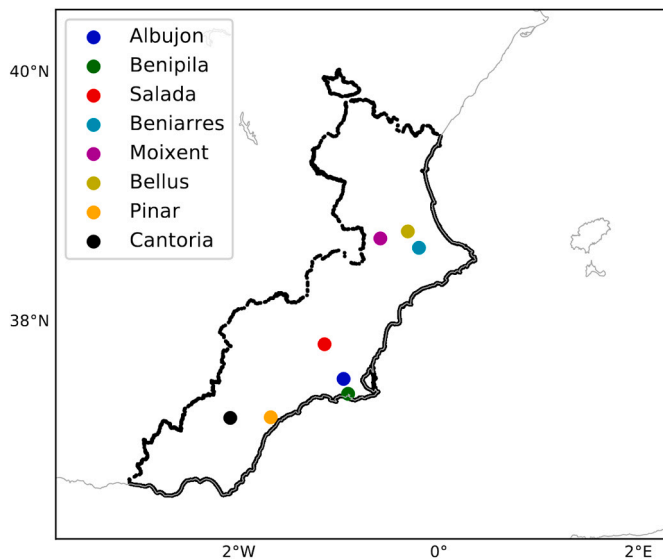
Fig. 15. Location of the rain gauges placed inside the eight catchments used to examine ensemble features at local scale.

convective bands in *Almería*, while only small precipitation structures developed near *Cap de la Nau*, where the *Beniarrés*, *Bellús* and *Moixent* catchments are placed. Slight intensities were registered over *Bellús* and *Moixent*, but due to localization errors, several members generate substantial rainfall on these basins. Stochastic ensembles produce lower onshore precipitation than MPS (Fig. 5a–d), which is consistent with the lower rainfall intensities forecasted in *Moixent*. During the last part of the episode, maximum hourly precipitation intensities over *Beniarrés*, *Moixent* and *Bellús* are captured by some ensemble members. Excluding *Beniarrés*, diversity is larger for stochastic techniques, considering only the basins where substantial intensities were recorded.

In brief, the increased spread displayed by stochastic ensembles at the scale of the study area is also reflected at catchment scale. Overall, the diversity generated by all tested techniques is enough to represent the observed maximum hourly intensities, except for a particularly extreme case of *Beniarrés* during phase 1 with 67 mm h^{-1}.

5. Conclusions

Stimulated by the typical deficiencies found in many convective-scale probabilistic forecasts concerning underdispersion, an examination of multiple methodologies to cope with model uncertainties for a heavy precipitation episode in eastern Spain is conducted. Given the importance of moist processes in such events and the presence of prominent orography near coastal areas in the study region, it is hypothesized that specific parameterized mechanisms, namely microphysics and PBL, played a significant role in the evolution of the episode. Therefore, accounting for inaccuracies in the formulation of these processes is a requirement for the improvement of probabilistic hydrometeorological forecasting chains, which require precise quantitative precipitation forecasts in order to issue valuable civil protection warnings to mitigate personal and material losses.

In this context, the performance of 50-member ensembles based on multiphysics and different stochastic methods is evaluated for the heavy precipitation episode of 12 and 13 September over eastern Spain. Stochastic experiments include the version of SPPT implemented in WRF (Berner et al., 2015), and an adaptation of the Random Parameters scheme of McCabe et al. (2016) applied to the microphysics parameterization (MPRP).

Microphysics processes are found to hold a significant influence in the development of the episode, as ensemble spread is substantially increased when uncertainties in this parameterization are considered. In

this regard, the suitability of the assumptions based on regional specificities and physical considerations, which consider the selected parameters (CCN concentration, fall speed factor and saturation percentage for cloud formation) as influential factors is proven. In general terms, stochastic ensembles improve upon the diversity generated by the multiphysics technique, so resulting in a more comprehensive strategy to sample uncertainties at these scales and improve the forecast of extreme phenomena.

Perturbations generated by the multiple methodologies inspected present diverse characteristics. Microphysics perturbations initially feature low spatial correlation. However, they intensively grow over areas with intense convective instability and gain spatial correlation, while they decay in other areas. Conversely, SPPT perturbations introduce modifications with a spatial correlation determined by the parameter specified in the scheme. As for MPRP, perturbations are intensified in areas with deep moist convection, but also in other regions. Hence, variations produced by MPRP are focused over the most influential areas. The changes introduced produce variations in the precipitation field at the scale of the errors, indicating the potential of stochastic perturbation to improve rainfall forecasts.

Stochastic experiments exhibit better verification scores compared to mutiphysics during the last phase of the episode, characterized by the intense influence of local orography and convective development, mainly initiated over the sea. In this situation better reliability and discrimination - increasing hit rates and reducing false alarms - is obtained for stochastic methods with negligible differences among specific strategies. This finding reveals the benefits of perturbing subgrid processes under circumstances wherein small-scale mechanisms play a key role. By contrast, in the previous phases, which are dominated by larger scale structures differences are lower and multiphysics exhibits better skill, especially for low precipitation thresholds.

The combination of various stochastic perturbations does not present overall significant benefits compared to the use of simple schemes. The nonlinearities that dominate convective-scale dynamics and the fast perturbation saturation may prevent an additive effect when multiple techniques are combined.

The increase in spread caused by stochastic methods is also translated to local scales. All experiments demonstrate the capacity to generate scenarios that capture the maximum hourly precipitation intensity over some small-to-medium size watersheds in the study area. This diagnosis is remarkable for hydrometeorological forecasting, which demands accurate rainfall forecasts at small scales of the order of basin size. Despite this, some extreme observed intensities are not forecasted, which emphasizes the difficulties found to design ensemble strategies that seamlessly sample the entire range of possible scenarios compatible with the uncertainties associated with the initial state and model formulation.

The potential benefits in terms of diversity and skill displayed by the stochastic techniques, and in particular by the microphysics perturbations, present several avenues for further research. Admittedly, the application of these techniques to a single episode limits the significance of the results, and the identified differences in forecast skill may not be representative of ensemble performance in other cases. Nevertheless, the encouraging results obtained in this study lay the basis for a wider investigation including an extended period or a collection of high-impact episodes in order to pursue statistically significant results. In addition, a thorough examination of the particular physical processes modified by stochastic perturbations that lead to an increased ensemble spread could help understanding the behavior of stochastic perturbation methods and making more informed configuration choices that would allow us to extract a greater benefit from stochastic techniques and ensemble prediction systems. Analyzing the impact of perturbations to additional microphysics and PBL parameters could contribute to this goal.

Despite the positive impact of stochastic perturbations to microphysics, the effect cannot be extrapolated in a straightforward manner to
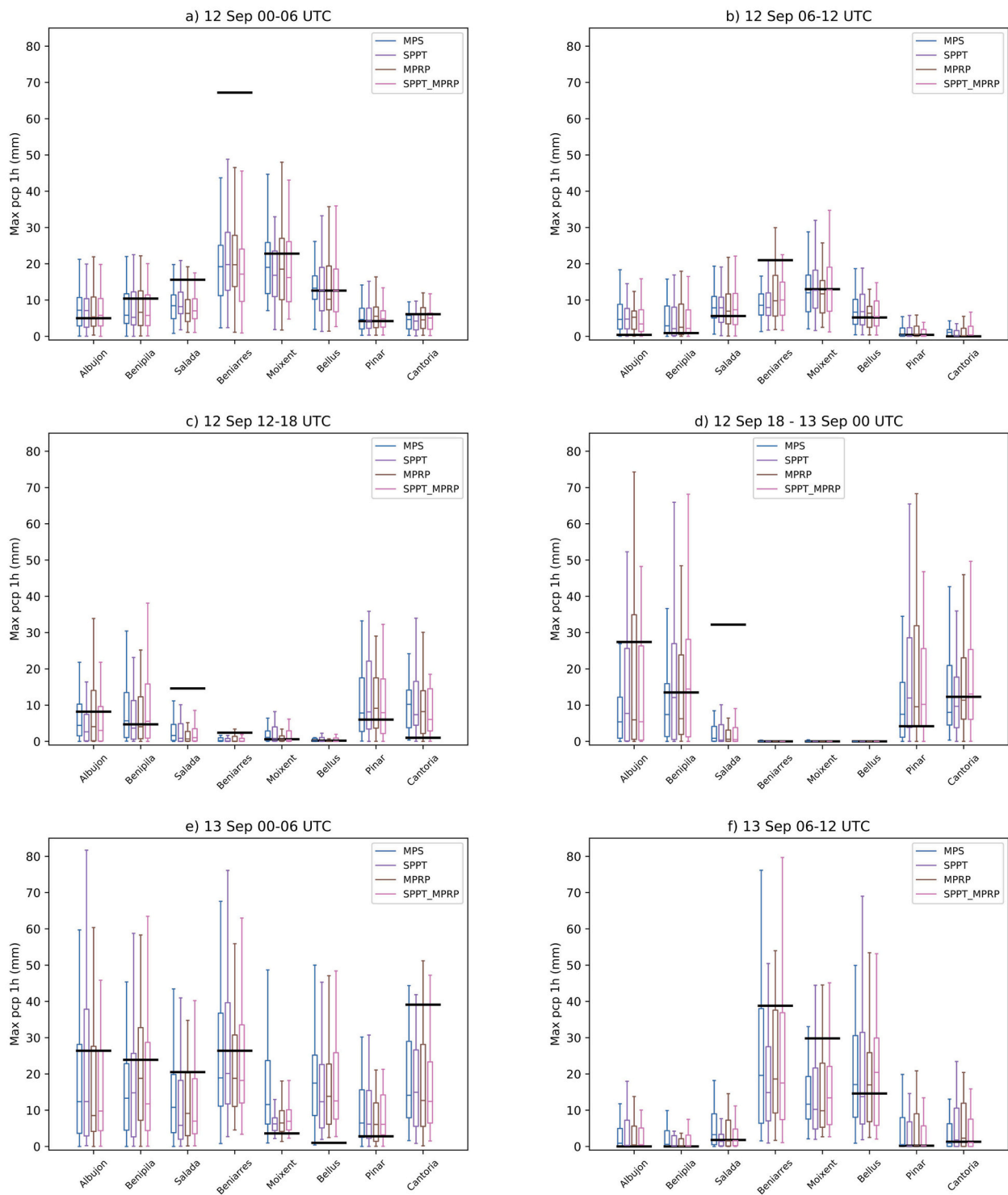
Fig. 16. Maximum hourly precipitation intensities for different 6-h periods over the location of rain gauges depicted in Fig. 15 provided by the four ensemble configurations defined. Boxes represent interquartile ranges and horizontal lines inside boxes indicate ensemble median. Black horizontal lines stand for observed values of maximum hourly rainfall intensity.

other relevant processes or variables for heavy precipitation and flash flooding. The application of stochastic perturbations to sea surface temperature was also attempted, as previous studies have shown its influence on precipitation intensity and location (e.g., Stocchi and Davolio, 2017; Ivatek-Šahdan et al., 2018; Iizuka and Nakamura, 2019), and also its impacts on hydrometeorological forecasts over small-to-medium size catchments (Senatore et al., 2020). However, only marginal effects were obtained either from exclusively perturbing this field

or in combination with the other methods investigated in this study.

Furthermore, initial and boundary conditions should also be included in order to sample all relevant sources of uncertainty. The scarcity of in-situ observations over the sea is extremely influential in the Mediterranean basin. Indeed, many convective systems that ultimately affect coastal areas originate over maritime areas. Therefore, the characterization of uncertainty over the sea and the adoption of techniques that allow the translation of information from land towards

maritime areas is crucial. These error sources are not considered in this study so as to better isolate the impacts of model error sampling strategies. However, the combined effect of initial and boundary condition and model perturbations in a highly nonlinear framework may provide interesting conclusions for the development and improvement of convection-permitting ensembles.

Declaration of Competing Interest

Acknowledgements

References

Amengual, A., Carrió, D., Ravazzani, G., Homar, V., 2017. A comparison of ensemble strategies for flash flood forecasting: the 12 october 2007 case study in Valencia, Spain. J. Hydrometeorol. 18, 1143–1166. https://doi.org/10.1175/JHM-D-16-0281.1.

Bachmann, K., Keil, C., Weissmann, M., 2019. Impact of radar data assimilation and orography on predictability of deep convection. Quart. J. Royal Meteor. Soc. 145, 117–130. https://doi.org/10.1002/qj.3412.

Bachmann, K., Keil, C., Craig, G., Weissmann, M., Welzbacher, C., 2020. Predictability of deep convection in idealized and operational forecasts: Effects of radar data assimilation, orography, and synoptic weather regime. Mon. Weather Rev. 148, 63–81. https://doi.org/10.1175/MWR-D-19-0045.1.

Barthlott, C., Hoose, C., 2018. Aerosol effects on clouds and precipitation over Central Europe in different weather regimes. J. Atmos. Sci. 75, 4247–4264. https://doi.org/10.1175/JAS-D-18-0110.1.

Beck, J., Bouttier, F., Wiegand, L., Gebhardt, C., Eagle, C., Roberts, N., 2016. Development and verification of two convection-allowing multi-model ensembles over Western Europe. Quart. J. Royal Meteor. Soc. 142, 2808–2826. https://doi.org/10.1002/qj.2870.

Berner, J., Ha, S.Y., Hacker, J., Fournier, A., Snyder, C., 2011. Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. Mon. Weather Rev. 139, 1972–1995. https://doi.org/10.1175/2010MWR3595.1.

Berner, J., Fossell, K., Ha, S., Hacker, J.P., Snyder, C., 2015. Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. Mon. Weather Rev. 143, 1295–1320. https://doi.org/10.1175/MWR-D-14-00091.1.

Berner, J., Achatz, U., Batté, L., Bengtsson, L., De La Cámara, A., Christensen, H., Colangeli, M., Coleman, D., Crommelin, D., Dolaptchiev, S., Franzke, C., Friederichs, P., Imkeller, P., Järvinen, H., Juricke, S., Kitsios, V., Lott, F., Lucarini, V., Mahajaajaajan, S., Palmer, T., Penland, C., Sakradzijaja, M., Von Storch, J., Weisheimer, A., Weniger, M., Williams, P., Yano, J.I., 2017. Stochastic parameterization toward a new view of weather and climate models. Bull. Amer. Meteor. Soc. 98, 565–587. https://doi.org/10.1175/BAMS-D-15-00268.1.

Bonavita, M., Trémolet, Y., Holm, E., Lang, S., Chrust, M., Janisková, M., Lopez, P., Laloyaux, P., De Rosnay, P., Fisher, M., Hamrud, M., English, S., 2017. A strategy for data assimilation. Technical Memorandum. https://doi.org/10.21957/tx1epjd2p. ECMWF.

Bouttier, F., Vié, B., Nuissier, O., Raynaud, L., 2012. Impact of stochastic physics in a convection-permitting ensemble. Mon. Weather Rev. 11, 3706–3721. https://doi.org/10.1175/MWR-D-12-00031.1.

Bouttier, F., Vié, B., Raynaud, L., Nuissier, O., Ménétrier, B., 2016. Sensitivity of the AROME ensemble to initial and surface perturbations during HyMeX. Quart. J. Royal Meteor. Soc. 142, 390–403. https://doi.org/10.1002/qj.2622.

Bowler, N., Arribas, A., Mylne, K., Robertson, K., Beare, S., 2008. The MOGREPS short-range ensemble prediction system. Quart. J. Royal Meteor. Soc. 134, 703–722. https://doi.org/10.1002/qj.234.

Buizza, R., Leutbecher, M., 2015. The forecast skill horizon. Quart. J. Royal Meteor. Soc. 141, 3366–3382. https://doi.org/10.1002/qj.2619.

Buizza, R., Miller, M., Palmer, T.N., 1999. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. Quart. J. Royal Meteor. Soc 125, 2887–2908. https://doi.org/10.1256/smsqj.56005.

Buizza, R., Houtekamer, P.L., Toth, Z., Pellerin, G., Wei, M., Zhu, Y., 2005. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. Mon. Weather Rev. 133, 1076–1097. https://doi.org/10.1175/MWR2905.1.

Carrió, D., Homar, V., Wheatley, D., 2019. Potential of an EnKF storm-scale data assimilation system over sparse observation regions with complex orography. Atmos. Res. 216, 186–206. https://doi.org/10.1016/j.atmosres.2018.10.004.

Chen, X., Yuan, H., Xue, M., 2018. Spatial spread-skill relationship in terms of agreement scales for precipitation forecasts in a convection-allowing ensemble. Quart. J. Royal Meteor. Soc. 144, 85–98. https://doi.org/10.1002/qj.3186.

Christensen, H., Moroz, I., Palmer, T., 2015. Stochastic and perturbed parameter representations of model uncertainty in convection parameterization. J. Atmos. Sci. 72, 2525–2544. https://doi.org/10.1111/j.2153-3490.1955.tb01138.x.

Christensen, H., Lock, S., Moroz, I., Palmer, T., 2017. Introducing independent patterns into the Stochastically Perturbed Parametrization tendencies (SPPT) scheme. Quart. J. Royal Meteor. Soc. 143, 2168–2181. https://doi.org/10.1002/qj.3075.

Cintineo, R., Stensrud, D., 2013. On the predictability of supercell thunderstorm evolution. J. Atmos. Sci. 70, 1993–2011. https://doi.org/10.1175/JAS-D-12-0166.1.

Clark, A., Kain, J., Stensrud, D., Xue, M., Kong, F., Coniglio, M.C., Thomas, K.W., Wang, Y., Brewster, K., Gao, J., Wang, X., 2011. Probabilistic precipitation forecast skill as a function of ensemble size and spatial scale in a convection-allowing ensemble. Mon. Weather Rev. 139, 1410–1418. https://doi.org/10.1175/2010mwr3624.1.

Clarke, A., Uehara, T., Porter, J., 1997. Atmospheric nuclei and related aerosol fields over the Atlantic: Clean subsiding air and continental pollution during ASTEX. J. Geophys. Res.-Atmos. 102, 25281–25292. https://doi.org/10.1029/97jd01555.

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

Consorcio de Compensación de Seguros (CCS), 2019. Cuarta nota informativa del consorcio de compensación de seguros sobre las inundaciones producidas entre los días 11 y 15 de septiembre en el sureste. https://www.consorseguros.es/web/noticias/-/asset_publisher/ya2OdYGgbjgX/content/cuarta-not a-informativa-del-consorcio-de-compensacion-de-seguros-sobre-las-inundaciones -producidas-entre-los-dias-11-y-15-de-septiembre-en-el-sureste.

Degelia, S., Wang, X., Stensrud, D., Johnson, A., 2018. Understanding the impact of radar and in situ observations on the prediction of a nocturnal convection initiation event on 25 June 2013 using an ensemble-based multiscale data assimilation system. Mon. Weather Rev. 146, 1837–1859. https://doi.org/10.1175/MWR-D-17-0128.1.

Descamps, L., Labadie, C., Joly, A., Bazile, E., Arbogast, P., Cébron, P., 2015. PEARP, the Météo-France short-range ensemble prediction system. Quart. J. Royal Meteorol. Soc. 141, 1671–1685. https://doi.org/10.1002/qj.2469.

Du, J., DiMego, G., Zhou, B., Jovic, D., Ferrier, B., Yang, B., 2015. Regional ensemble forecast systems at NCEP. In: 27th Conference on Weather Analysis and forecasting/23rd Conference on Numerical Weather Prediction. Amer. Meteor. Soc., Chicago, p. 14. URL: https://ams.confex.com/ams/27WAF23NWP/webprogram/Paper27 3421.html.

Duc, L., Saito, K., Seko, H., 2013. Spatial-temporal fractions verification for high-resolution ensemble forecasts. Tellus A 65, 18171. https://doi.org/10.3402/tellusa.v65i0.18171.

Ducrocq, V., Nuissier, O., Ricard, D., Lebeaupin, C., Thouvenin, T., 2008. A numerical study of three catastrophic precipitating events over southern France. II: Mesoscale triggering and stationarity factors. Quart. J. Royal Meteor. Soc. 134, 131–145. https://doi.org/10.1002/qj.199.

Duda, J., Wang, X., Xue, M., 2017. Sensitivity of convection-allowing forecasts to land surface model perturbations and implications for ensemble design. Mon. Weather Rev. 145, 2001–2025. https://doi.org/10.1175/MWR-D-16-0349.1.

Ebert, E., 2008. Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. Meteorol. Appl. 15, 51–64. https://doi.org/10.1002/met.25.

Ehrendorfer, M., 1994. The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Theory. Mon. Weather Rev. 122, 703–713. https://doi.org/10.1175/1520-0493(1994)122<0703:TLEAIP>2.0.CO;2.

Faccini, F., Paliaga, G., Piana, P., Sacchini, A., Watkins, C., 2016. The Bisagno stream catchment (Genoa, Italy) and its major floods: geomorphic and land use variations in the last three centuries. Geomporphology 273, 14–27. https://doi.org/10.1016/j.geomorph.2016.07.037.

Falk, N., Igel, A., Igel, M., 2019. The relative impact of ice fall speeds and microphysics parameterization complexity on supercell evolution. Mon. Weather Rev. 147, 2403–2415. https://doi.org/10.1175/MWR-D-18-0417.1.

Fiori, E., Comellas, A., Molini, L., Rebora, N., Siccardi, F., Gochis, D., Tanelli, S., Parodi, A., 2014. Analysis and hindcast simulations of an extreme rainfall event in the Mediterranean area: the Genoa 2011 case. Atmos. Res. 138, 13–29. https://doi.org/10.1016/j.atmosres.2013.10.007.

Freud, E., Rosenfeld, D., Kulkarni, J., 2011. Resolving both entrainment-mixing and number of activated CCN in deep convective clouds. Atmos. Chem. Phys. 11, 12887–12890. https://doi.org/10.5194/acp-11-12887-2011.

Frogner, I., Singleton, A., Køltzow, M., Andrae, U., 2019. Convection-permitting ensembles: challenges related to their design and use. Quart. J. Royal Meteor. Soc. 145, 90–106. https://doi.org/10.1002/qj.3525.

García-Moya, J., Callado-Pallarès, A., Escribà, P., Santos, C., Santos-Muñoz, D., Simarro, J., 2011. Predictability of short-range forecasting: a multimodel approach. Tellus A 63, 550–563. https://doi.org/10.1111/j.1600-0870.2010.00506.x.

Gasperoni, N., Wang, X., Wang, Y., 2020. A comparison of methods to sample model errors for convection-allowing ensemble forecasts in the setting of multiscale initial conditions produced by the GSI-based EnVar assimilation system. Mon. Weather Rev. 148, 1177–1203. https://doi.org/10.1175/MWR-D-19-0124.1.

Goger, B., Rotach, M., Gohm, A., Stiperski, I., Fuhrer, O., 2016. Current challenges for numerical weather prediction in complex terrain: Topography representation and parameterizations. In: 2016 International Conference on High Performance Computing and simulation, HPCS 2016, Innsbruck, Austria, pp. 890–894. https://doi.org/10.1109/HPCSim.2016.7568428.

Gowan, T., Steenburgh, W., Schwartz, C., 2018. Validation of mountain precipitation forecasts from the convection-permitting NCAR ensemble and operational forecast systems over the western United States. Wea. Forecasting 33, 739–765. https://doi.org/10.1175/WAF-D-17-0144.1.

Hacker, J., Ha, S., Snyder, C., Berner, J., Eckel, F., Kuchera, E., Pocernich, M., Rugg, S., Schramm, J., Wang, X., 2011a. The U.S. Air Force Weather Agency's mesoscale ensemble: Scientific description and performance results. Tellus A 63, 625–641. https://doi.org/10.1111/j.1600-0870.2010.00497.x.

Hacker, J., Snyder, C., Ha, S., Pocernich, M., 2011b. Linear and non-linear response to parameter variations in a mesoscale model. Tellus A 63, 429–444. https://doi.org/10.1111/j.1600-0870.2010.00505.x.

Hasselmann, K., 1976. Stochastic climate models part I. Theory. Tellus 125, 2487–2515. https://doi.org/10.1256/smsqj.55907.

Hermoso, A., Homar, V., Yano, J.I., 2020a. Exploring the limits of ensemble forecasting via solutions of the Liouville equation for realistic geophysical models. Atmos. Res. 245, 105127. https://doi.org/10.1016/j.atmosres.2020.105127.

Hermoso, A., Homar, V., Greybush, S., Stensrud, D., 2020b. Tailored ensemble prediction systems: Application of seamless scale bred vectors. J. Meteor. Soc. Japan 98. https://doi.org/10.2151/jmsj.2020-053.

Hermoso, A., Homar, V., Amengual, A., 2021. The sequence of heavy precipitation and flash flooding of 12 and 13 September 2019 in eastern Spain. Part I: Mesoscale diagnostic and sensitivity analysis of precipitation. J. Hydrometeor. doi. https://doi.org/10.1175/JHM-D-20-0182.1.

Hohenegger, C., Schär, C., 2007. Predictability and error growth dynamics in cloud-resolving models. J. Atmos. Sci. 64, 4467–4478. https://doi.org/10.1175/2007jas2143.1.

Homar, V., Ramis, C., Alonso, S., 2002. A deep cyclone of African origin over the Western Mediterranean: Diagnosis and numerical simulation. Ann. Geophys. 131, 93–106. https://doi.org/10.5194/angeo-20-93-2002.

Hong, S., Lim, J., 2006. The WRF Single-Moment 6-Class Microphysics Scheme (WSM6). J. Korean Meteor. Soc. 42, 129–151.

Hou, D., Kalnay, E., Droegemeier, K., 2001. Objective verification of the SAMEX '98 ensemble forecasts. Mon. Weather Rev. 129, 73–91. https://doi.org/10.1175/1520-0493(2001)129<0073:OVOTSE>2.0.CO;2.

Houtekamer, P., Mitchell, H., Deng, X., 2009. Model error representation in an operational ensemble Kalman filter. Mon. Weather Rev. 137, 2126–2143. https://doi.org/10.1175/2008MWR2737.1.

Hudson, J., Mishra, S., 2007. Relationships between CCN and cloud microphysics variations in clean maritime air. Geophys. Res. Lett. 34 https://doi.org/10.1029/2007GL030044.

Iacono, M., Delamere, J., Mlawer, E., Shephard, M., Clough, S., Collins, W., 2008. Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. J. Geophys. Res.-Atmos. 113, D13103 https://doi.org/10.1029/2008JD009944.

Iizuka, S., Nakamura, H., 2019. Sensitivity of midlatitude heavy precipitation to SST: a case study in the Sea of Japan area on 9 August 2013. J. Geophys. Res.-Atmos. 124, 4365–4381. https://doi.org/10.1029/2018JD029503.

Isaksen, L., Bonavita, M., Buizza, R., Fisher, M., Haseler, J., Leutbecher, M., Raynaud, L., 2010. Ensemble of data assimilations at ECMWF. In: Technical Memorandum. ECMWF. https://doi.org/10.21957/obke4k60.

Ivatek-Šahdan, S., Stanešić, A., Tudor, M., Odak Plenković, I., Janeković, I., 2018. Impact of SST on heavy rainfall events on eastern Adriatic during SOP1 of HyMeX. Atmos. Res. 200, 36–59. https://doi.org/10.1016/j.atmosres.2017.09.019.

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

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

Jankov, I., Berner, J., Coauthors, 2017. A performance comparison between multiphysics and stochastic approaches within a north American RAP ensemble. Mon. Weather Rev. 145, 1161–1179. https://doi.org/10.1175/MWR-D-16-0160.1.

Jankov, I., Beck, J., Wolff, J., Harrold, M., Olson, J., Smirnova, T., Alexander, C., Berner, J., 2019. Stochastically perturbed parameterizations in an HRRR-based ensemble. Mon. Weather Rev. 147, 153–173. https://doi.org/10.1175/MWR-D-18-0092.1.

Johnson, A., Wang, X., 2016. A study of multiscale initial condition perturbation methods for convection-permitting ensemble forecasts. Mon. Weather Rev. 144, 2579–2604. https://doi.org/10.1175/MWR-D-16-0056.1.

Jolliffe, I.T., Stephenson, D.B., 2003. Forecast Verification. Wiley, A practitioner's guide in atmospheric science.

Kazumori, M., 2014. Satellite radiance assimilation in the JMA operational mesoscale 4DVAR system. Mon. Weather Rev. 1361-1381, 216–230. https://doi.org/10.1175/MWR-D-13-00135.1.

Keil, C., Baur, F., Bachmann, K., Rasp, S., Schneider, L., Barthlott, C., 2019. Relative contribution of soil moisture, boundary-layer and microphysical perturbations on convective predictability in different weather regimes. Quart. J. Royal Meteor. Soc. 145, 3102–3115. https://doi.org/10.1002/qj.3607.

Keresturi, E., Wang, Y., Meier, F., Weidle, F., Wittmann, C., Atencia, A., 2019. Improving initial condition perturbations in a convection-permitting ensemble prediction system. Quart. J. Royal Meteor. Soc. 145, 993–1012. https://doi.org/10.1002/qj.3473.

Klasa, C., Arpagaus, M., Walser, A., Wernli, H., 2018. An evaluation of the convection-permitting ensemble COSMO-E for three contrasting precipitation events in Switzerland. Quart. J. Royal Meteor. Soc. 144, 744–764. https://doi.org/10.1002/qj.3245.

Kober, K., Craig, G., 2016. Physically based stochastic perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. J. Atmos. Sci. 73, 2893–2911. https://doi.org/10.1175/JAS-D-15-0144.1.

Kühnlein, C., Keil, C., Craig, G., Gebhardt, C., 2014. The impact of downscaled initial condition perturbations on convective-scale ensemble forecasts of precipitation. Quart. J. Royal Meteor. Soc. 140, 1552–1562. https://doi.org/10.1002/qj.2238.

Lagasio, M., Parodi, A., Procopio, R., Rachidi, F., Fiori, E., 2017. Lightning potential index performances in multimicrophysical cloud-resolving simulations of a back-building mesoscale convective system: the Genoa 2014 event. J. Geophys. Res. Atmos. 122, 4238–4257. https://doi.org/10.1002/2016JD026115.

Lang, S.K., Lock, S., Leutbecher, M., Bechtold, P., Forbes, R., 2021. Revision of the SPP model uncertainty scheme in the IFS. Quart. J. Royal Meteor. Soc. doi. https://doi.org/10.1002/qj.3978.

Leutbecher, M., Lock, S., Ollinaho, P., Lang, S., Balsamo, G., Bechtold, P., Bonavita, M., Christensen, H., Diamantakis, M., Dutra, E., English, S., Fisher, M., Forbes, R., Goddard, J., Haiden, T., Hogan, R., Juricke, S., Lawrence, H., MacLeod, D., Magnusson, L., Malardel, S., Massart, S., Sandu, I., Smolarkiewicz, P., Subramanian, A., Vitart, F., Wedi, N., Weisheimer, A., 2017. Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. Quart. J. Royal Meteor. Soc. 143, 2315–2339. https://doi.org/10.1002/qj.3094.

Li, Y., Lu, G., Wu, Z., He, H., Shi, J., Ma, Y., Weng, S., 2016. Evaluation of optimized WRF precipitation forecast over a complex topography region during flood season. Atmosphere 7, 145. https://doi.org/10.3390/atmos7110145.

Llasat, M.C., Llasat-Botija, M., Rodríguez, A., Lindbergh, S., 2010. Flash floods in Catalonia: a recurrent situation. Adv. Geosci. 126, 105–111. https://doi.org/10.5194/adgeo-26-105-2010.

Lock, S., Lang, S., Leutbecher, M., Hogan, R., Vitart, F., 2019. Treatment of model uncertainty from radiation by the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme and associated revisions in the ECMWF ensembles. Quart. J. Royal Meteor. Soc. 145, 75–89. https://doi.org/10.1002/qj.3570.

Lorenz, E., 1963. Deterministic nonperiodic flow. J. Atmos. Sci. 20, 130–141. https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Lorenz, E., 1969. The predictability of a flow which possesses many scales of motion. Tellus 21, 289–307. https://doi.org/10.3402/tellusa.v21i3.10086.

Lorenzo-Lacruz, J., Amengual, A., Garcia, C., Morán-Tejeda, E., Homar, V., Maimó-Far, A., Hermoso, A., Ramis, C., Romero, R., 2019. Hydro-meteorological reconstruction and geomorphological impact assessment of the October 2018 catastrophic flash flood at Sant Llorenç, Mallorca (Spain). Nat. Hazards Earth Syst. Sci. 19, 2597–2617. https://doi.org/10.5194/nhess-19-2597-2019.

Lupo, K., Torn, R., Yang, S., 2020. Evaluation of stochastic perturbed parameterization tendencies on convective-permitting ensemble forecasts of heavy rainfall events in New York and Taiwan. Wea. Forecasting 35, 5–24. https://doi.org/10.1175/WAF-D-19-0064.1.

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

Marsigli, C., Montani, A., Paccagnella, T., 2014. Provision of boundary conditions for a convection-permitting ensemble: Comparison of two different approaches. Nonlinear Process. Geophys. 21, 393–403. https://doi.org/10.5194/npg-21-393-2014.

Mason, I., 1982. A model for assessment of weather forecasts. Aust. Meteor. Mag. 30, 291–303.

Mason, S., Graham, N., 2002. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. Quart. J. Royal Meteor. Soc. 128, 2145–2166. https://doi.org/10.1256/003590002320603584.

McCabe, A., Swinbank, R., Tennant, W., Lock, A., 2016. Representing model uncertainty in the Met Office convection-permitting ensemble prediction system and its impact on fog forecasting. Quart. J. Royal Meteor. Soc. 142, 2897–2910. https://doi.org/10.1002/qj.2876.

Melhauser, C., Zhang, F., 2012. Practical and intrinsic predictability of severe and convective weather at the mesoscales. J. Atmos. Sci. 69, 3350–3371. https://doi.org/10.1175/JAS-D-11-0315.1.

Molteni, F., Buizza, R., Palmer, T.N., Petroliagis, T., 1996. The ECMWF ensemble prediction system: Methodology and validation. Quart. J. Royal Meteor. Soc. 122, 73–119. https://doi.org/10.1256/smsqj.52904.

Morales, A., Posselt, D., Morrison, H., He, F., 2019. Assessing the influence of microphysical and environmental parameter perturbations on orographic precipitation. J. Atmos. Sci. 76, 1373–1395. https://doi.org/10.1175/JAS-D-18-0301.1.

Nakanisi, M., Niino, H., 2006. An improved Mellor-Yamada level-3 model: its numerical stability and application to a regional prediction of advection fog. Boundary-Layer Meteor. 119, 397–407. https://doi.org/10.1007/s10546-005-9030-8.

Nuissier, O., Ducrocq, V., Ricard, D., Lebeaupin, C., Anquetin, S., 2008. A numerical study of three catastrophic precipitating events over southern France. I: Numerical framework and synoptic ingredients. Quart J. Royal Meteor. Soc. 134, 111–130. https://doi.org/10.1002/qj.200.

Okamoto, K., Sawada, Y., Kunii, M., 2019. Comparison of assimilating all-sky and clear-sky infrared radiances from Himawari-8 in a mesoscale system. Quart. J. Royal Meteor. Soc. 145, 745–766. https://doi.org/10.1002/qj.3463.

Ollinaho, P., Lock, S., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., Forbes, R., Haiden, T., Hogan, R., Sandu, I., 2017. Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ECMWF ensemble. Quart. J. Royal Meteor. Soc. 143, 408–422. https://doi.org/10.1002/qj.2931.

Palmer, T., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G., Steinheimer, M., Weisheimer, A., 2009. Stochastic Parametrization and Model uncertainty. Technical Memorandum. ECMWF. https://doi.org/10.21957/ps8gbwbdv.

Peralta, C., Ben Bouallègue, Z., Theis, S.E., Gebhardt, C., Buchhold, M., 2012. Accounting for initial condition uncertainties in COSMO-DE-EPS. J. Geophys. Res.-Atmos. 117, 1–13. https://doi.org/10.1029/2011JD016581.

Phillips, V., Donner, L., Garner, S., 2007. Nucleation processes in deep convection simulated by a cloud-system-resolving model with double-moment bulk microphysics. J. Atmos. Sci. 64, 738–761. https://doi.org/10.1175/JAS3869.1.

Posselt, D., He, F., Bukowski, J., Reid, J., 2019. On the relative sensitivity of a tropical deep convective storm to changes in environment and cloud microphysical parameters. J. Atmos. Sci. 76, 1163–1185. https://doi.org/10.1175/JAS-D-18-0181.1.

Primo, C., Rodríguez, M., Gutiérrez, J., 2008. Logarithmic bred vectors. A new ensemble method with adjustable spread and calibration time. J. Geophys. Res.-Atmos. 113, 1–11. https://doi.org/10.1029/2007JD008998.

Prospero, J., 1996. Saharan dust transport over the North Atlantic Ocean and Mediterranean: an overview. In: Guerzoni, S., Chester, R. (Eds.), The Impact of Desert Dust Across the Mediterranean. Environmental Science and Technology Library, 11. Springer, Dordrecht, pp. 133–151. https://doi.org/10.1007/978-94-017-3354-0_13.

Qiao, X., Wang, S., Min, J., 2018. The impact of a stochastically perturbing microphysics scheme on an idealized supercell storm. Mon. Weather Rev. 146, 95–118. https://doi.org/10.1175/MWR-D-17-0064.1.

Ramis, C., Romero, R., Homar, V., Alonso, S., Alarcón, M., 1998. Diagnosis and numerical simulation of a torrential precipitation event in Catalonia (Spain). Meteorog. Atmos. Phys. 69, 1–21. https://doi.org/10.1007/BF01025180.

Romero, R., Sumner, G., Ramis, C., Genovés, A., 1999. A classification of the atmospheric circulation patterns producing significant daily rainfall in the Spanish Mediterranean area. Int. J. Climatol. 19, 765–785. https://doi.org/10.1002/(SICI)1097-0088(19990615)19:7<765::AID-JOC388>3.0.CO;2-T.

Romine, G., Scwartz, C., Snyder, C., Anderson, J., M.L. W, 2013. Model bias in a continuously cycled assimilation system and its influence on convection-permitting forecasts. Mon. Weather Rev. 141, 1263–1284. https://doi.org/10.1175/MWR-D-12-00112.1.

Romine, G., Scwartz, C., Berner, J., Fossell, K., Snyder, C., Anderson, J., Weisman, M., 2014. Representing forecast error in a convection-permitting ensemble system. Mon. Weather Rev. 142, 4519–4541. https://doi.org/10.1175/MWR-D-14-00100.1.

Roux, H., Amengual, A., Romero, R., Bladé, E., Sanz-Ramos, M., 2020. Evaluation of two hydrometeorological ensemble strategies for flash-flood forecasting over a catchment of the eastern Pyrenees. Nat. Hazards Earth Syst. Sci. 20, 425–450. https://doi.org/10.5194/nhess-20-425-2020.

Schellander-Gorgas, T., Wang, Y., Meier, F., Weidle, F., Wittmann, C., Kann, A., 2017. On the forecast skill of a convection-permitting ensemble. Geosci. Model Dev. 10, 35–56. https://doi.org/10.5194/gmd-10-35-2017.

Schwartz, C., Sobash, R., 2017. Generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: a review and recommendations. Mon. Weather Rev. 145, 3397–3418. https://doi.org/10.1175/MWR-D-16-0400.1.

Schwartz, C., Romine, G., Smith, K., Weisman, M., 2014. Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble kalman filter. Wea. Forecasting 29, 1295–1318. https://doi.org/10.1175/WAF-D-13-00145.1s.

Selz, T., Craig, G., 2015. Upscale error growth in a high-resolution simulation of a summertime weather event over Europe. Mon. Weather Rev. 143, 813–827. https://doi.org/10.1175/MWR-D-14-00140.1.

Senatore, A., Furnari, L., Mendicino, G., 2020. Impact of high-resolution sea surface temperature representation on the forecast of small Mediterranean catchments' hydrological responses to heavy precipitation. Hydrol. Earth Syst. Sci. 24, 269–291. https://doi.org/10.5194/hess-24-269-2020.

Shutts, G., 2005. A kinetic energy backscatter algorithm for use in ensemble prediction systems. Quart. J. Royal Meteor. Soc. 131, 3079–3102. https://doi.org/10.1256/qj.04.106.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D.O., Barker, D., Duda, M.G., Huang, X.Y., Wang, W., Powers, J.G., 2008. A Description of the Advanced Research WRF Version 3. NCAR Tech. Note NCAR/TN-475+STR. https://doi.org/10.5065/D6DZ069T.

Smirnova, T., Brown, J., Benjamin, S., Kenyon, J., 2016. Modifications to the Rapid Update Cycle land surface model (RUC LSM) available in the weather research and forecasting (WRF) model. Mon. Weather Rev. 144, 1851–1865. https://doi.org/10.1175/MWR-D-15-0198.1.

Stensrud, D., Brooks, H., Du, J., Tracton, M., Rogers, E., 1999. Using ensembles for short-range forecasting. Mon. Weather Rev. 127, 433–446. https://doi.org/10.1175/1520-0493(1999)127<0433:UEFSRF>2.0.CO;2.

Stocchi, P., Davolio, S., 2017. Intense air-sea exchanges and heavy orographic precipitation over Italy: the role of Adriatic Sea surface temperature uncertainty. Atmos. Res. 196, 62–82. https://doi.org/10.1016/j.atmosres.2017.06.004.

Sun, Y., Zhang, F., 2016. No access intrinsic versus practical limits of atmospheric predictability and the significance of the butterfly effect. J. Atmos. Sci. 73, 1419–1438. https://doi.org/10.1175/JAS-D-15-0142.1.

Surcel, M., Zawadzki, I., Yau, M., 2015. A study on the scale dependence of the predictability of precipitation patterns. J. Atmos. Sci. 72, 216–235. https://doi.org/10.1175/JAS-D-14-0071.1.

Tennant, W., Beare, S., 2014. New schemes to perturb sea-surface temperature and soil moisture content in MOGREPS. Quart. J. Royal Meteor. Soc. 140, 1150–1160. https://doi.org/10.1002/qj.2202.

Thanh, C., Tien, T., Chanh, K., 2016. Application of breeding ensemble to tropical cyclone track forecasts using the Regional Atmospheric Modeling System (RAMS) model. Appl. Math. Model. 40, 8309–8325. https://doi.org/10.1016/j.apm.2016.04.010.

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

Tian, J., Liu, J., Yan, D., Ding, L., Li, C., 2019. Ensemble flood forecasting based on a coupled atmospheric-hydrological modeling system with data assimilation. Atmos. Res. 224, 127–137. https://doi.org/10.1016/j.atmosres.2019.03.029.

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

Velden, C., Lewis, W., Bresky, W., Stettner, D., Daniels, J., Wanzong, S., 2017. Assimilation of high-resolution satellite-derived atmospheric motion vectors: Impact on HWRF forecasts of tropical cyclone track and intensity. Mon. Weather Rev. 145, 1107–1125. https://doi.org/10.1175/mwr-d-16-0229.1.

Vié, B., Nuissier, O., Ducrocq, V., 2011. Cloud-resolving ensemble simulations of Mediterranean heavy precipitating events: uncertainty on initial conditions and lateral boundary conditions. Mon. Weather Rev. 139, 403–423. https://doi.org/10.1175/2010mwr3487.1.

Wang, S., Qiao, X., Min, J., 2020. Impact of stochastically perturbed terminal velocities on convective-scale ensemble forecasts of precipitation. Adv. Meteorol. 2020 https://doi.org/10.1155/2020/4234361.

Wang, Y., Wang, X., 2017. A pseudorandom number generator based on piecewise logistic map. Mon. Weather Rev. 145, 1447–1471. https://doi.org/10.1175/MWR-D-16-0231.1.

Wang, Y., Bellus, M., Wittmann, C., Steinheimer, M., Weidle, F., Kann, A., Ivatek-Šahdan, S., Tian, W., Ma, X., Tascu, S., Bazile, E., 2011. The central European limited-area ensemble forecasting system: ALADIN-LAEF. Quart. J. Royal Meteor. Soc. 137, 483–502. https://doi.org/10.1002/qj.751.

Wastl, C., Wang, Y., Atencia, A., Wittmann, C., 2019a. Independent perturbations for physics parametrization tendencies in a convection-permitting ensemble (pSPPT). Geosci. Model Dev. 12, 261–273. https://doi.org/10.5194/gmd-12-261-2019.

Wastl, C., Wang, Y., Atencia, A., Wittmann, C., 2019b. A hybrid stochastically perturbed parametrization scheme in a convection-permitting ensemble. Mon. Weather Rev. 147, 2217–2230. https://doi.org/10.1175/MWR-D-18-0415.1.

Wheatley, D., Yussouf, N., Stensrud, D., 2014. Ensemble Kalman filter analyses and forecasts of a severe mesoscale convective system using different choices of microphysics schemes. Mon. Weather Rev. 142, 3243–3263. https://doi.org/10.1175/MWR-D-13-00260.1.

Whitaker, J., Hamill, T., Wei, X., Song, Y., Toth, Z., 2008. Ensemble data assimilation with the NCEP global forecast system. Mon. Weather Rev. 136, 463–482. https://doi.org/10.1175/2007MWR2018.1.

White, B., Gryspeerdt, E., Stier, P., Morrison, H., Thompson, G., Kipling, Z., 2017. Uncertainty from the choice of microphysics scheme in convection-permitting models significantly exceeds aerosol effects. Atmos. Chem. Phys. 17, 12145–12175. https://doi.org/10.5194/acp-17-12145-2017.

Xu, Z., Chen, J., Jin, Z., Li, H., Chen, F., 2020. Representing model uncertainty by multi-stochastic physics approaches in the GRAPES ensemble. Adv. Atmos. Sci. 37, 328–346. https://doi.org/10.1007/s00376-020-9171-1.

Yáñez Morroni, G., Gironás, J., Caneo, M., Delgado, R., Garreaud, R., 2018. Using the Weather Research and forecasting (WRF) model for precipitation forecasting in an Andean region with complex topography. Atmosphere 9, 304. https://doi.org/10.3390/atmos9080304.

Yussouf, N., Stensrud, D., 2012. Comparison of single-parameter and multiparameter ensembles for assimilation of radar observations using the Ensemble Kalman Filter. Mon. Weather Rev. 140, 562–586. https://doi.org/10.1175/MWR-D-10-05074.1.

Zhang, S., Pu, Z., Velden, C., 2018. Impact of enhanced atmospheric motion vectors on HWRF hurricane analyses and forecasts with different data assimilation configurations. Mon. Weather Rev. 146, 1549–1569. https://doi.org/10.1175/MWR-D-17-0136.1.

Zhang, X., 2019. Multiscale characteristics of different-source perturbations and their interactions for convection-permitting ensemble forecasting during SCMREX. Mon. Weather Rev. 147, 291–310. https://doi.org/10.1175/MWR-D-18-0218.1.

Zhang, Y., Zhang, F., Stensrud, D., Meng, Z., 2015. Practical predictability of the 20 May 2013 tornadic thunderstorm event in Oklahoma: Sensitivity to synoptic timing and topographical influence. Mon. Weather Rev. 143, 2973–2997. https://doi.org/10.1175/MWR-D-14-00394.1.

Zhang, Y., Zhang, F., Stensrud, D., Meng, Z., 2016. Intrinsic predictability of the 20 May 2013 tornadic thunderstorm event in Oklahoma at storm scales. Mon. Weather Rev. 144, 1273–1298. https://doi.org/10.1175/MWR-D-15-0105.1.