Atmospheric Research 281 (2023) 106479

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

# Characterization of ensemble generation strategies: Application to three illustrative examples of Mediterranean high-impact weather

Alejandro Hermoso [a,b,*], Victor Homar [a], Romualdo Romero [a]

[a] Meteorology Group, Physics Department, Universitat de les Illes Balears, Palma Mallorca, Spain
[b] Institute for Atmospheric and Climate Science, ETH Zürich, Zürich, Switzerland

ARTICLE INFO

*Keywords:*
Ensemble prediction systems
Stochastic parameterization
Bred vectors
Microphysics
Western Mediterranean
Heavy precipitation episodes

ABSTRACT

Predictability of high-impact weather constitutes a major challenge. In this study, we contribute to this relevant topic by analyzing the characteristics of multiple ensemble prediction systems aimed at sampling initial and/or model uncertainties for three exemplary heavy precipitation systems that occurred over the western Mediterranean. In particular, initial condition perturbations are generated by (i) directly downscaling the ensemble members from a global ensemble and (ii) creating perturbations by sampling a wide range of scales using an adaptation of the breeding methodology. Model error is sampled by applying stochastic perturbations to physical parameterizations and to microphysics parameters. A positive impact in terms of ensemble diversity is obtained when initial condition perturbations across a broad range of scales are applied, especially at low levels and for cases in which local factors are more relevant. Regarding stochastic model perturbations, they display very localized perturbations with low amplitude, which are insufficient to capture extreme scenarios, except when local factors play a dominant role. However, when combined with initial condition perturbations, they generally increase diversity and perturbation amplitude over areas characterized by deep moist convection. This outcome exhibits the positive impact of sampling both uncertainty sources, even in the extremely nonlinear regime that defines convective-scale phenomena.

## 1. Introduction

Extreme weather events, such as intense cyclones, heavy precipitation, hailstorms, or convective winds are among the most frequent and damaging natural disasters. Indeed, economic losses near 150 billion USD and 8000 casualties were caused by floods and storms worldwide in 2020 alone (CRED, 2021). Although many of these events occur in tropical areas, the Mediterranean region contains a sizable part of them (see `https://www.munichre.com`). As a consequence, large socio-economic losses also occur in Mediterranean countries. Indeed, according to the Emergency Events Database (EM-DAT) (`https://www.emdat.be/`), more than 600 fatalities have been registered since 2000, while economic costs have amounted to 23 billion EUR in Mediterranean countries. Focusing on the western Mediterranean, heavy precipitation episodes are recurrent in eastern Iberian Peninsula (e.g., Llasat and Puigcerver, 1994; Homar et al., 2002; Hermoso et al., 2021a), Balearic Islands (e.g., Martín et al., 2006; Romero et al., 2015; Lorenzo-Lacruz et al., 2019), southeastern France (e.g., Nuissier et al., 2008; Lagouvardos et al., 2013; Caumont et al., 2021), northwestern Italy (e.g., Davolio et al., 2013; Fiori et al., 2014; Davolio et al., 2020), and Corsica and Sardinia (e.g., Lambert and Argence, 2008; Torcasio et al., 2020).

The factors contributing to the development of convective systems in the Mediterranean region have been extensively documented in the available literature (cf. Michaelides et al., 2018). The complex topography of the area, characterized by prominent mountain ranges and the presence of multiple islands, peninsulas or gulfs modifies the mesoscale flow and also acts as a direct lifting mechanism, exercising a substantial impact on the genesis and evolution of these systems. In addition, the intrusion of upper-level disturbances combined with warm and moist air at low levels produces the necessary instability for the development of deep moist convection. Indeed, the presence of low-level jets with long pathways over the sea is crucial to create and sustain instability and provide an adequate moisture feed for high rainfall rates. The relatively warm sea surface temperature intensifies evaporation, especially during late summer and autumn, when these events are more likely, according to climatology (Llasat et al., 2010; Insua-Costa et al., 2021).

Accurate numerical weather prediction of these extreme phenomena

* Corresponding author at: Physics Department, Universitat de les Illes Balears, 07122 Palma de Mallorca, Spain.
*E-mail address:* alejandro.hermoso@env.ethz.ch (A. Hermoso).

https://doi.org/10.1016/j.atmosres.2022.106479
Received 25 February 2022; Received in revised form 23 September 2022; Accepted 16 October 2022
Available online 29 October 2022
0169-8095/

is essential to mitigate their associated personal and material losses. Owing to the chaotic nature of the atmospheric system, recognized since the seminal works of Lorenz (Lorenz, 1963), and the multiple uncertainties involved in the forecasting process, namely related to the initial state of the system and model formulation, the degree of uncertainty of the numerical weather prediction must be quantified. For realistic applications, this process is tackled by means of ensemble forecasting due to the unfeasibility of computing the full probability density function (PDF) by solving the Liouville or Fokker-Planck equations. Under the recurring limitation of computational resources, the sampling of the PDF should aim at identifying fast-growing error modes, as proposed in the first operational ensemble prediction systems (EPS), such as the singular vectors developed at the European Centre for Medium-Range Forecasts (ECMWF; Molteni et al., 1996) or the bred vectors designed at the National Centers for Environmental Prediction (NCEP; Toth and Kalnay, 1993). For the particular case of short-range mesoscale and convective-scale forecasting, required to usefully anticipate severe weather events, the initial condition sampling strategy must account for errors across a wide range of scales, in order to cover all relevant uncertainties necessary to capture all plausible scenarios. In this sense, previous research has been focused on the impacts of multi-scale initial condition perturbations on the mesoscale by means of data assimilation (e.g., Johnson and Wang, 2016, 2020) or bred vectors (Hermoso et al., 2020).

Concerning model error, although some ensembles are built by combining different models (e.g., García-Moya et al., 2011; Beck et al., 2016), the most general approach consists of exploring uncertainties associated with the physical parameterizations. In this regard, either physics schemes (e.g., Hacker et al., 2011; Wang et al., 2011; Du et al., 2015) or specific parameters within these parameterizations (e.g., Gebhardt et al., 2011; Duda et al., 2017) are varied across different ensemble members. Moreover, stochastic perturbations can be applied for this task (Berner et al., 2017). A popular approach includes perturbations to the full physics tendency by applying a random pattern with preset spatial and temporal correlation, known as Stochastically Perturbed Physics Tendencies (SPPT; Buizza et al., 1999; Berner et al., 2015). Additional methods are focused at the process level by perturbing specific and relevant parameters within particular schemes (Jankov et al., 2017; Lang et al., 2021).

The attainment of dependable predictions of severe weather events is challenging due to the dominance of nonlinear dynamics by which small perturbations rapidly amplify, especially in areas affected by deep moist convection (Zhang et al., 2007; Selz and Craig, 2015). Under these conditions, the predictability of socially relevant features of convective systems, such as their location, timing and intensity, has generally a short predictability horizon. This is caused by errors in either the initial condition or model formulation, which restrict practical predictability (Melhauser and Zhang, 2012; Zhang et al., 2015; Flora et al., 2018), but also by the nature of the system, which has an intrinsic predictability limit that cannot be overcome by increasing accuracy (Zhang et al., 2016; Sun and Zhang, 2016; Markowski, 2020). However, the predictability horizon is strongly case dependent (Frogner et al., 2019), and affected by additional factors, such as the orography (Bachmann et al., 2020).

In this general context, we focus on the performance of multiple ensemble generation strategies based on perturbations to initial conditions and/or subgrid parameterizations for convective-scale short-range forecasting of heavy precipitation episodes in the western Mediterranean. In particular, we analyze the impact of introducing large-scale perturbations by downscaling from a global model, and investigate the effect of initial condition perturbations tailored to cover a wide range of spatial scales following the recently developed methodology of Hermoso et al. (2020). The influence of model error is inspected by means of stochastic perturbations to relevant physical parameterizations involved in convective-scale forecasting, especially microphysics (Hermoso et al., 2021b). The main objective consists thus of testing these novel sampling strategies in the extremely challenging context of high-impact weather forecasting. In order to obtain broader results and provide recommendations for ensemble design, three different episodes, representing diverse atmospheric settings leading to heavy precipitation systems, are selected to investigate the performance of each type of perturbation. Although the number of cases is limited, thus reducing statistical significance, this design allows us to emphasize the singularities of each technique for each atmospheric configuration considered and provide a detailed and physically based analysis. More specifically, we aim at testing (i) whether the introduction of smaller-scale initial condition perturbations produces improvements in terms of diversity and skill, which can be useful when forecasting extreme events, (ii) to assess the positive impact of the combination of perturbations applied to multiple error sources in this particular framework, and (iii) assess the performance of recently developed techniques compared to more simple methods (e.g., downscaling). In addition, the evolution of the perturbations is investigated to relate the impacts of each method to the characteristics of each event.

The most characteristic features of these episodes are described in section 2, while the design of the EPSs is detailed in section 3. Results are presented and discussed in section 4 and main conclusions are summarized in section 5.

## 2. Case studies

The choice of three recent and illustrative case studies to investigate the characteristics of multiple ensemble generation strategies is based on the severity of their socioeconomic impacts. The selected case studies exhibit diverse atmospheric settings, which enables the assessment of the benefits of each strategy in terms of the most important driving mechanisms of each event.

### 2.1. Case 1: 19–20 December 2016

This episode mainly impacted Corsica and Sardinia, with accumulations exceeding 200 mm in 24 h in southern Corsica and northwestern Sardinia (Fig. 1a). The intense rainfall and subsequent floods caused infrastructure and property damage, as well as roadblocks.

The upper-level situation on 17 December was dominated by a cold cut-off low over north Africa and southern Iberian Peninsula. Along this area, a negatively tilted trough developed over northern Europe on 18 December and moved southwestwards producing an additional cut-off low over central Europe. At low levels, a cyclone formed over north Africa during the first hours of 19 December, subsequently intensifying, advancing northwards and positioning between Balearic Islands and Corsica and Sardinia (Fig. 1b). This cyclonic circulation advected warm and moist air towards Corsica and Sardinia, characterized by high equivalent potential temperature ($\theta_e$) (Fig. 1b), providing a favorable environment for the development of the heavy precipitation systems that affected this area. Therefore, this event allows us to identify the benefits of each ensemble generation strategy when larger-scale features are more influential.

### 2.2. Case 2: 09–10 September 2017

The 9–10 September event produced substantial rainfall with accumulations above 300 mm in 24 h in northern and central Italy, especially over Livorno (Fig. 1c). The devastating effects of this episode resulted in 9 casualties and economic losses above 200 million EUR, according to the EM-DAT data.

The synoptic setting was characterized by a prominent trough, extending from Scandinavia to Iberian Peninsula on 9 September, and moving eastwards (Fig. 1d). At low levels, a low pressure system formed between the Alps and the Ligurian Sea (Fig. 1d). Indeed, this phenomenon of lee cyclogenesis (Buzzi and Tibaldi, 1978) is the predominant mechanism for cyclone formation in the area (Buzzi et al., 2020). In

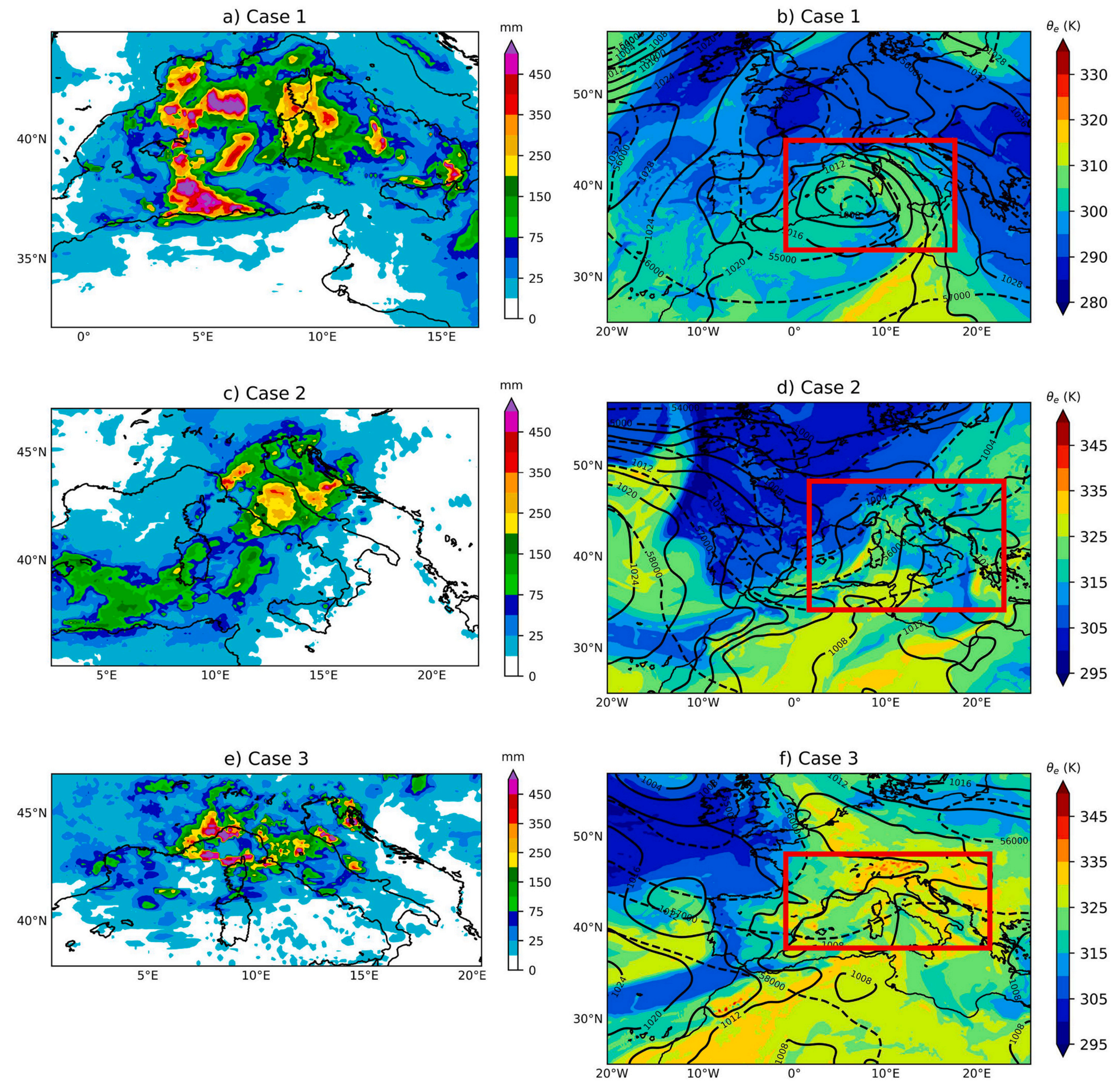


**Fig. 1.** Left column: Global Precipitation Measurement Integrated Multi-satellite Retrievals (GPM IMERG) 24-h accumulated precipitation for a) Case 1 (from 19 December 2016 at 1800 UTC to 20 December 2016 at 1800 UTC), c) Case 2 (from 9 September 2017 at 1800 UTC to 10 September 2017 at 1800 UTC) and e) Case 3 (from 27 July 2019 at 0600 UTC to 28 July 2019 at 0600 UTC). Right column: ECMWF analysis of sea level pressure (hPa, solid lines), geopotential at 500 hPa ($m^2s^{-2}$, dashed lines) and equivalent potential temperature at 850 hPa (K, shaded) valid for b) Case 1 (20 December 2016 at 0000 UTC), d) Case 2 (10 September 2017 at 0000 UTC) and f) Case 3 (27 July 2019 at 0000 UTC). The areas of concern, where the highest rainfall amounts were registered and represented in panels a), c), e) are indicated by the red rectangles in panels b), d), f). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

addition, inspection of $\theta_e$ reveals the contrast between a warm moist air mass, progressing westwards, and a cooler and drier air mass extending from southwestern France (Fig. 1d). This warm and moist advection contributed to production and sustenance of convective instability and favored high rainfall rates in northern Italy. A convective system developed over the area of $\theta_e$ contrast and progressed towards the west central Italian coast. This system interacted with the lee cyclone, resulting in a mesoscale convective system that affected central Italy. Therefore, this event is useful for investigating the performance of the different ensemble generation strategies on mesoscale convection and to analyze the influence of the orography on this convective systems. Further details of this case are provided in Federico et al. (2019) and Capecchi et al. (2021).

### 2.3. *Case 3: 27 July 2019*

The event of 27 July 2019 produced flash floods and large hail over southwestern France and northern Italy. Precipitation accumulations surpassed 450 mm in 24 h over specific locations, but contrary to the previous cases, rainfall affected more localized areas (Fig. 1e). The case was extremely disruptive, involving considerable personal and material losses.

A positively oriented trough extending from Iceland to the western Mediterranean defined the upper-level setting on 27 July (Fig. 1f), while at low levels, a low pressure system formed over the Ligurian Sea (Fig. 1f). Convective systems favored by the orography of the area formed over southwestern France during the first hours of 27 July in an area with high values of convective available potential energy. Additional convective systems developed over northern Italy, favored by the advection of the unstable air mass and a southerly flow at low levels, which supplied moisture to feed these systems. Local scale features, such as the topography, seem to play a more relevant role in this episode, which is characterized by a weaker synoptic forcing. Thus, this case especially constitutes an appropriate testbed to assess the impacts of ensemble perturbations on small-scale aspects of the flow and the environment, complementing the analysis of larger scale structures of the previous cases.

## 3. Methodology

The numerical simulations to investigate the effect of different ensemble perturbations are performed with the Weather Research and Forecasting (WRF) model version 3.9.1.1 (Skamarock et al., 2008). The domains adopted for each case study have 750 $\times$ 500 grid points with a horizontal grid spacing of 2.5 km and 50 vertical levels. The region covered by each domain is centered to capture the main features of each episode, so that the influence of boundary conditions is lessened (Fig. 2). The simulations for each case study extend over 30 h, leaving the first 6 h as a spin-up period, which is not considered for verification purposes. Simulation start times are chosen to cover the initiation and evolution of the relevant systems identified in the diagnostic of each event and are indicated in Table 1. Physical parameterizations are the same for all experiments and case studies and include the National Severe Storms Laboratory 2-moment microphysics scheme (Mansell et al., 2010), Rapid Radiative Transfer Model radiation (Iacono et al., 2008), Rapid Update Cycle land-surface model (Smirnova et al., 2016) and Mellor-Yamada and Nakanishi Niino planetary boundary layer (Nakanisi and Niino, 2006). Although some previous works have proven that horizontal resolutions of the order of 100 m are needed to reproduce all relevant aspects of convection (Bryan et al., 2003), this process is assumed to be explicitly resolved and thus, it is not parameterized in these simulations, since no universal benefits have been identified when this parameterization is applied to gray-zone resolutions (Han and Hong, 2018). Five 50-member (+1 unperturbed control member) ensemble configurations including initial condition/lateral boundary condition (IC/LBC) and/or model perturbations are assessed for each case study. These configurations are described below.

### 3.1. Dynamical downscaling (DOWN)

This ensemble accounts for large-scale uncertainty in the initial state and at the boundaries of the limited area domain by dynamically downscaling the 50 ensemble members from the ECMWF-EPS. The latest resolution update was conducted in March 2016, when a truncated cubic octahedral grid $T_{CO}639$, corresponding to approximately 18 km in mid-latitudes, was implemented (Haiden et al., 2016). In the ECMWF-EPS, uncertainty in the initial atmospheric state is explored by means of a combination of singular vectors and ensemble data assimilation (Bonavita et al., 2017; ECMWF, 2020), which account for both fast-growing errors in the near future sampled by the singular vectors and dominant growing modes in the analysis cycle. Conversely, model error is sampled with SPPT using a combination of random patterns covering multiple scales and including a tapering to reduce perturbations near the model surface and top levels (Leutbecher et al., 2017).

The 50 different initial conditions provided by the ECMWF-EPS are applied to define the atmospheric initial state for each member of DOWN, while boundary conditions in the limited area domain are updated every 3 h using the corresponding member of the global ensemble.

### 3.2. Tailored bred perturbations (BRED)

The potential of introducing IC perturbations covering a wide range of relevant scales is investigated with this ensemble. The methodology to build these perturbations is based on the breeding technique created by Toth and Kalnay (1993). A breeding cycle is produced by running the full nonlinear model using an unperturbed and a randomly perturbed initial state for a short time interval. The difference between the two forecasts is rescaled and added to the next analysis to continue the cycle. The method used for this study follows the modifications introduced in Hermoso et al. (2020), including an orthogonalization of the perturbations before applying the rescaling during the breeding cycle. This procedure was shown to significantly increase the perturbations diversity compared to the traditional rescaling, preventing the common collapse of bred cycles. In addition, the scale of the resulting bred vectors is modified by introducing an exponential transformation that allows us to tailor the scale of the perturbation without changing the characteristics of the breeding cycle, namely the interval between rescalings. In order to generate 50 perturbations with this technique, 25 breeding cycles are initiated 10 days before the forecast start time, allowing for a spin-up period to reduce the effect of the initial random perturbation. Bred vectors are rescaled and orthogonalized every 6 h, and the exponential transformation is applied to the bred vectors valid on the forecast start times. The resulting perturbations are added and subtracted to the analysis to create an ensemble of 50 twin perturbations. Uncertainty in boundary conditions is provided by the ECMWF-EPS, as for the DOWN ensemble.

Bred vectors aim at identifying dynamically unstable modes, which arguably increases the probability of finding extremes. This makes this ensemble generation strategy adequate for the purpose of this study. Furthermore, this method is less computationally demanding and easier to apply to different case studies than data assimilation, which allows us to investigate the feasibility of this technique to be implemented as an alternative or even a complement to ensemble data assimilation.

### 3.3. Stochastic model perturbations (STO)

The influence of sampling model uncertainties is examined by means of stochastic perturbations to parameterizations affecting relevant processes for convective development and evolution. Due to the fact that convection is not parameterized in these simulations, planetary boundary layer (PBL) and microphysics processes have the strongest effects at the scales of interest. Perturbations to these schemes are generated by combining the SPPT scheme available in WRF (Berner et al., 2015) with a method designed to modify some key parameters within the NSSL 2-moment microphysics scheme (see Hermoso et al., 2021b). Indeed, this combination allows perturbations to be applied to all physical parameterizations, since the WRF SPPT does not perturb microphysics tendencies to avoid double-counting the effect of these perturbations (Jankov et al., 2019). The random patterns used in SPPT have a spatial correlation of 100 km, a temporal correlation of 1 h and a variance of 0.25, the maximum value that ensures that the sign of physical tendencies is not reversed, which could produce numerical instabilities. The method to perturb microphysics is based on the random parameters approach used at UK Met Office (McCabe et al., 2016) and consists of perturbing various important parameters, so that they vary along forecast lead time following a first-order autoregressive process with a decorrelation time of 0.5 h. In this method no spatial variations are applied, and the values of the parameters are kept within predefined ranges determined by physical considerations, which are

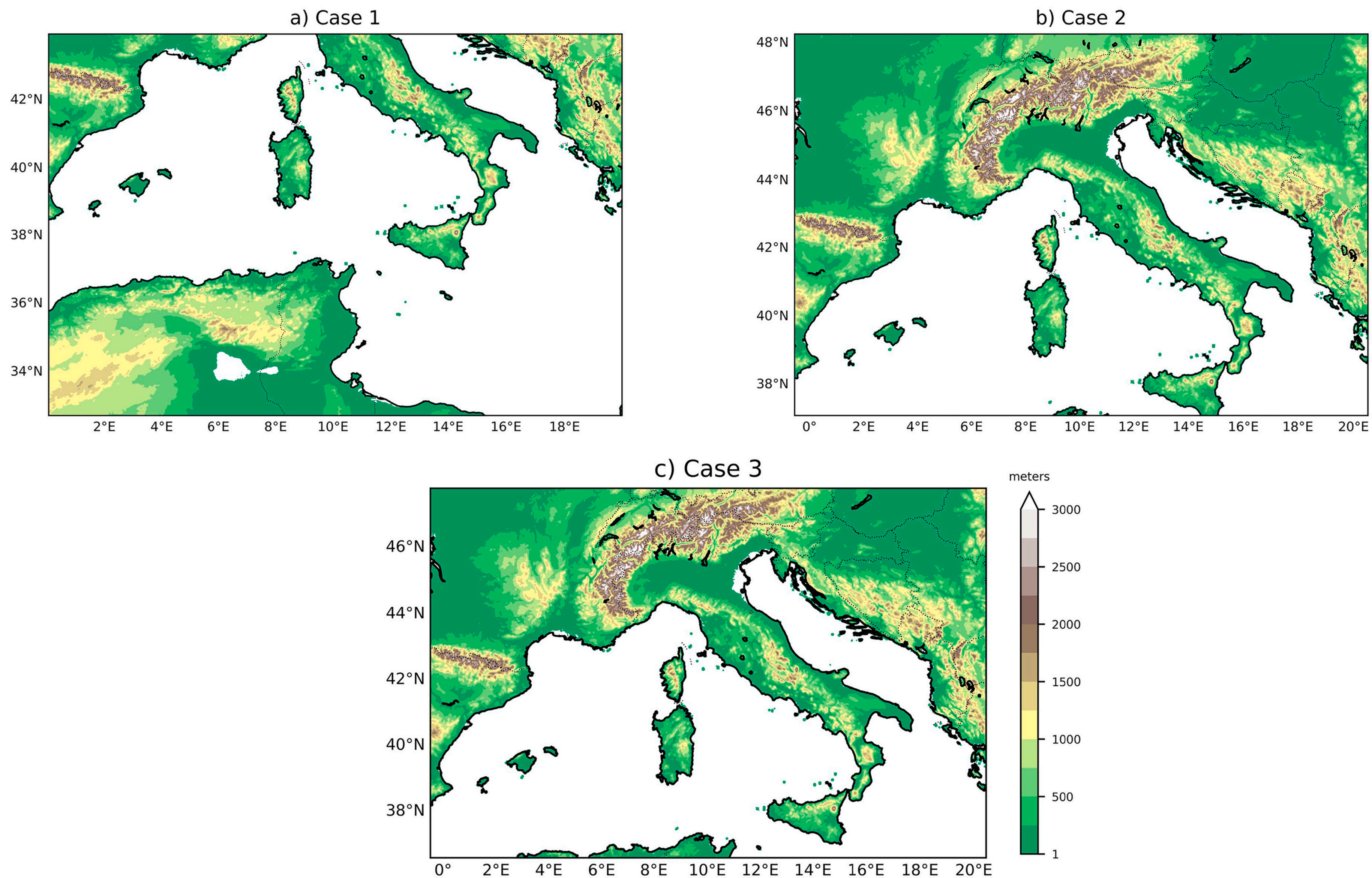


**Fig. 2.** Computational domains used for the numerical experiments of a) Case 1, b) Case 2 and c) Case 3.

**Table 1**
Summary of the three selected case studies.

| Case | Main area affected | Date | Simulation start time |
|---|---|---|---|
| Case 1 | Corsica and Sardinia | 19–20 December 2016 | 19 December 1200 UTC |
| Case 2 | Northern and central Italy | 9–10 September 2017 | 9 September 1200 UTC |
| Case 3 | Southwestern France and northern Italy | 27 July 2019 | 27 July 0000 UTC |

indicated in Table 2. Perturbed parameters include cloud condensation nuclei (CCN), due to the geographical characteristics of the Mediterranean basin, favorable for the presence of air masses with different origins (Prospero, 1996) and the relevance of this quantity to precipitation (Barthlott and Hoose, 2018; Keil et al., 2019); graupel and hail fall speed factors, which influence cold pool intensity (Falk et al., 2019), and saturation percentage for initial cloud formation (ssmxinit), which regulates the number of activated CCN (Phillips et al., 2007). This ensemble is driven by the control member of the ECMWF-EPS and does not include IC/LBC perturbations. The value of the SPPT stochastic parameters follows the set-up of Hermoso et al. (2021b), where suitable performance was obtained for a heavy precipitation episode. This parameter selection is also consistent with other studies focusing on the convective scale (e.g., Romine et al., 2014).

### 3.4. *Combination of IC/LBC and model perturbations (DOWN+STO and BRED+STO)*

The purpose of these experiments is to investigate the impact of combined IC/LBC and model stochastic perturbation perturbations. Both downscaling from ECMWF-EPS and tailored bred vectors are used for this task. Considering the highly nonlinear regime in which convective systems develop, these experiments can shed light on the effects of sampling both sources uncertainties, which can be significantly different from the sum of individual impacts.

### 3.5. *Metrics*

We aim at evaluating the spread of the precipitation forecasts and the actual number of degrees of freedom spanned by each ensemble strategy to determine which method is able to perform a wider sampling of the phase space. Correspondence ratio and ensemble dimension constitute appropriate metrics to perform these tasks. Furthermore, the amplitude and scale of ensemble perturbations is analyzed by means of two quantities (log$\rho$ and $\omega^2$) that measure their mean and variance. The term perturbation is defined as the difference between an ensemble member and a control (unperturbed) forecast. Perturbations are computed using a moist total energy norm in order to include effects from all state variables (i.e., temperature, wind components and specific humidity):

$$TE = U^2 + V^2 + \frac{c_p}{T_{ref}}T^2 + w_q\frac{{L_v}^2}{c_pT_{ref}}Q^2, \quad (1)$$

where $U$ and $V$ and are the zonal and meriodional wind component, $T$ represents the temperature, $Q$ stands for the specific humidity, $c_p$ is the gas specific heat for constant pressure, $L_v$ is the latent heat of vaporization $T_{ref} = 300$ K and $w_q = 0.1$.

**Table 2**
Microphysics parameters perturbed in STO.

| 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 |

#### 3.5.1. *Verification: receiver operating characteristic (ROC)*

Precipitation ensemble forecasts are verified with the Receiver Operating Characteristic (Mason, 1982), which assesses the discrimination capacity of probabilistic forecasts by looking at multiple decision thresholds for a particular event of interest. ROCs are built by representing hit rate, defined as the fraction of correct positives forecasted for each probability threshold considered, versus false alarm rate, characterized by the fraction of false positives with respect to the number of non occurrences of the event. The area under the ROC is interpreted as the forecast discrimination skill. A value of 0.5 represents no skill compared to a random forecast, while a value of 1 corresponds to a perfect forecast. The achievement of a good hit rate against false alarm rate is extremely demanding in the context of extreme precipitation events. Therefore, it constitutes a suitable metric to examine the performance of the different ensemble generation strategies tested.

The verification is applied to 3-h accumulated rainfall using the Global Precipitation Measurement Integrated Multi-satellite Retrievals (GPM IMERG) product of the Goddard Earth Sciences Data and Information Services Disc (GES DISC) as observations. This dataset contains global precipitation information derived from satellite measurements with a spatial resolution of 0.1° × 0.1° and a temporal resolution of 30 min (Huffman et al., 2019). These data have been evaluated over the area of interest for this study using the ENSEMBLES OBServation (*E-OBS*) dataset (Navarro et al., 2019) and precipitation gridded products derived from rain gauge measurements (e.g., Tapiador et al., 2020 for Spain), and showed general adequate performance with the worst behavior occurring in mountainous locations. The homogeneous coverage provided by the dataset facilitates the comparison of results between multiple case studies affecting different areas in the western Mediterranean basin. Precipitation verification is affected by double-penalty, whereby small displacements of the rainfall field can produce both a miss and a false alarm (Mittermaier et al., 2013). This effect could be minimized by applying neighboring approaches (e.g., Schwartz and Sobash, 2017), which have been proven to be as effective as a large ensemble (Flack et al., 2021). However, this technique relies on various subjective choices such as the size and definition of the neighborhoods, which can affect the ROC result. For this reason, we only consider the grid-scale resolution in this study and thus the result are only representative for these small scales.

#### 3.5.2. *Correspondence ratio*

The degree of diversity in the precipitation fields is evaluated by means of the correspondence ratio (CR; Stensrud and Wandishin, 2000). It is defined as the ratio between the area of intersection where all the fields satisfy a condition (in this case, precipitation larger than a specified threshold) and the area wherein the union of all fields satisfies the condition. In this study, a probabilistic version of this metric is adopted. Thus, CR is defined herein as the ratio of the area where at least a fraction *P* of the ensemble members produce a precipitation higher than a specified threshold, and the area where at least one member surpasses the precipitation threshold. The correspondence ratio is defined as

$$CR = \frac{\sum_{i=1}^{N_{CR}}\left[\frac{1}{M}\sum_{j=1}^{M} d\left(f_{j,i}\right)\right] \geq P}{\sum_{i=1}^{N_{CR}} f_{1,i}\cup f_{2,i}\cup\ldots f_{M,i}}, \quad (2)$$

where $N_{CR}$ is the total number of grid points, $M$ is the number of ensemble members (50), $f_{j,\ i}$ is the binary forecast of the ensemble member $j$ at grid point $i$ and $d(f_{j,\ i})$ is 1 if the rainfall threshold is exceeded by the ensemble member $j$ at gridpoint $i$ and 0 otherwise. The original definition of CR is recovered when $P = 1$. CR is bounded between 0, which represents complete divergence of the fields, and is obtained if forecast probabilities at all grid points are lower than $P$, and 1, which is associated with perfect correspondence between all the ensemble members. CR is computed for 3-h accumulated precipitation for 20 and 50 mm thresholds. Different values of $P$ have been considered, but for the sake of conciseness only $P = 0.2$ and $P = 0.6$ for the 20 mm threshold and $P = 0.2$ and $P = 0.4$ for the 50 mm threshold are presented and discussed.

#### 3.5.3. Ensemble dimension

The degree of diversity in the output of each ensemble is quantified through the ensemble dimension. This quantity is derived from the eigenvalues of the perturbations covariance matrix, and measures the linear independence between a set of perturbations (Bretherton et al., 1999; Patil et al., 2001). It ranges between 1 if all perturbations are linearly dependent and the number of ensemble members if all perturbations are orthogonal. Contrary to other popular metrics to quantify spread, ensemble dimension is invariant to modifications of perturbation amplitude, providing a genuine assessment of diversity, which is insensitive to perturbation inflation.

#### 3.5.4. Perturbation amplitude and localization ($log\rho - \omega^2$)

Given a perturbation $\delta x_i$, its amplitude and scale can be characterized by means of the following parameters as proposed by Primo et al. (2008):

$$log\rho = log\left(\prod_{i=1}^{N_{log\rho}} |\delta x_i|^{\frac{1}{N_{log\rho}}}\right), \tag{3}$$

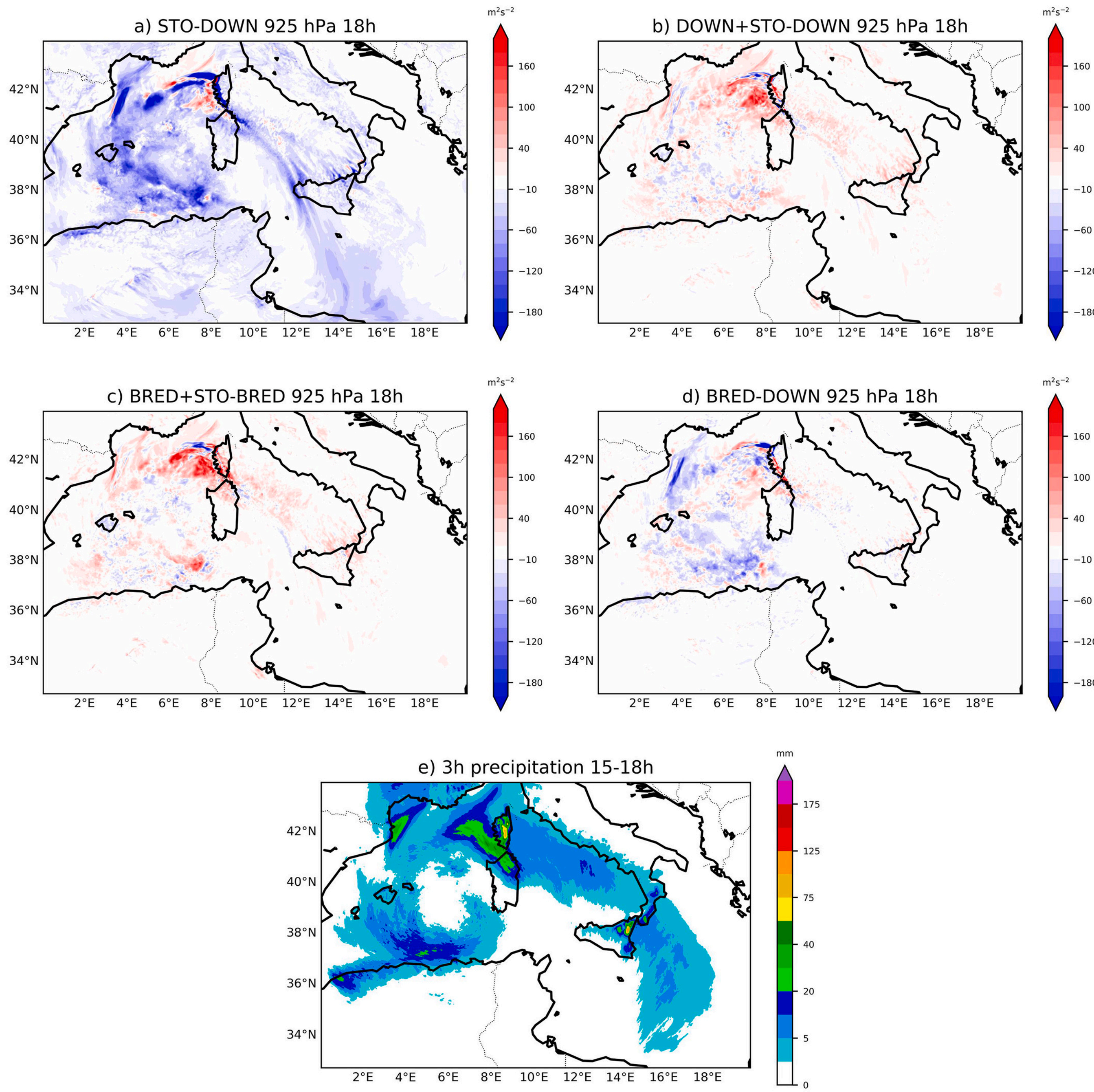


**Fig. 3.** Difference between moist total energy perturbations for multiple experiment pairs indicated in each panel for Case 1: a) STO - DOWN, b) DOWN+STO - DOWN, c) BRED+STO - BRED, d) BRED - DOWN. The displayed field corresponds to the mean along the top decile (i.e., 5 ensemble members) of perturbation differences, e) 3-h accumulated precipitation valid at the time indicated in panels a)-d).

$$\omega^2 = \frac{1}{N_{log\rho}} \sum_{i=1}^{N_{log\rho}} (log\delta x_i - log\rho)^2, \tag{4}$$

where $N_{log\rho}$ denotes the number of grid points where the perturbation is not zero. The first quantity mathematically represents the logarithmic mean of the perturbation and provides a measure of its amplitude, while $\omega^2$ is the variance and can be interpreted as a quantification of the scale of the perturbations. Admittedly, a value of $\omega^2$ is not categorically associated with a specific scale, since the value depends on the domain characteristics. For instance, a value close to 0 would imply low variance, which results in a perturbation covering the whole domain, although the scale of this perturbation depends on domain extension. If, as in the present case, all domains have the same characteristics, $\omega^2$ can be interpreted as a measure of perturbation localization, so that small (large) values represent low (high) variability and thus, large (small) perturbation spatial scales.

## 4. Results

### 4.1. Ensemble perturbation characteristics

In order to distill information from the ensemble experiments before analyzing specific metrics, the differences between forecast perturbations (total moist energy, Eq. 1) and rainfall accumulations obtained from different generation strategies are examined (Fig. 3-8)). Perturbations are computed as the difference between each ensemble member and the unperturbed forecast at model level 4 (approximately 925 hPa) after 18 h lead time for cases 1 and 3 and 24 h for case 2. At these specific forecast times the characteristics highlighted below are easier to distinguish. In order to focus on the largest perturbations, only the top decile (i.e., 5 ensemble members) is averaged at each grid point to obtain the perturbation maps displayed in Figs. 3, 5 and 7. We analyze the extreme values of the distribution rather than the mean in order to determine whether the different ensembles are able to capture the events.

The changes between DOWN or BRED and STO, that is the difference between perturbations in the ensembles based only on initial condition perturbations (DOWN or BRED) and the perturbations produced by the stochastic scheme, highlight the lower amplitude of the latter type of perturbation compared to IC/LBC modifications, which is common to all case studies. However, areas affected by precipitation systems are emphasized by STO (Figs. 3a,e, 5a,e and 7a,e), indicating the potential of these perturbations to sample relevant uncertainties influential for rainfall intensity and location (Figs. 3a, 5a). For precipitation, STO produces lower accumulations over most locations, seemingly owing to the smaller growth of the perturbations leading to heavy precipitation systems (Figs. 4a, 6a). This behavior could also be produced by the fact that this ensemble does not produce spurious rainfall as could be the case during the spin-up period in the experiments including downscaling from a coarser resolution model. For Case 3, in which the influence of local factors is greater, model perturbations are also more relevant than for the other cases, for which their impact is restricted to specific locations (Fig. 7a). STO perturbations are more dominant at low levels and for intermediate lead times, when the initial random perturbations have achieved larger spatial correlation. This effect influences rainfall accumulations so that a general decrease is not so appreciable, in contrast to the other cases (Fig. 8a).

The combination of perturbations applied to IC/LBC and physical parameterizations for both downscaling and bred perturbations (i.e., DOWN+STO and BRED+STO) exhibits a prevailing increase of the amplitude of the perturbations compared to the experiments including only initial condition perturbations (DOWN and BRED), especially over the areas of more intense convective development highlighted by STO, and at lower levels (Figs. 3b,c,e, 5b,c,e, 7b,c,e). However, shifts in the location of highest perturbations can also be seen, indicating the potential of subgrid perturbations to introduce larger spatial variability. It should be noted that precipitation location is of utmost importance in terms of flash flood forecasting, as small shifts can be extremely consequential. In this sense, the introduction of stochastic perturbations can enhance spatial diversity of the rainfall fields. This result is more appreciable for Cases 2 and 3 than for Case 1, in which larger-scale dynamics dominates and differences between precipitation accumulations for DOWN and DOWN+STO or BRED and BRED+STO display a random pattern (Figs. 4b,c,6b,c,8b,c). Regarding the comparison between DOWN and BRED, large-scale perturbations create greater divergence for Case 1 and Case 2, in which larger-scale systems are dominant (Figs. 3d,5d). The differences decrease with lead time, influenced by the common boundary condition perturbations in both experiments. This situation is reversed for Case 3, illustrating the relevance of smaller-scale structures, which are better represented in BRED (Fig. 7d). In addition, the effect of sampling across a wider range of scales in BRED generates shifts in the precipitation fields (Fig. 8d), as can be observed in STO.

### 4.2. Ensemble verification

The precipitation verification by means of the ROC area at grid-scale, considering a moderate threshold of 20 mm in 3 h, exhibits slight differences among experiments, except for STO in cases 1 and 2 (Fig. 9a,c, e). For these events, stochastic perturbations generate lower spatial spread than IC perturbations. As a consequence, high rainfall probabilities are concentrated over narrow areas, much smaller than the region that was affected by intense precipitation (Fig. 10b). Conversely, other strategies produce more widespread rainfall patterns, which leads to a larger area with non-zero probabilities in the area where precipitation was observed (Fig. 10b), resulting in higher hit rates for low probability thresholds and thus, greater ROC areas. However, for Case 3, distinguished by scattered convective precipitation, applying perturbations to subgrid parameterizations is as effective as introducing IC perturbations, especially during the last part of the episode. In contrast to the other events considered, the regions with high probability are distributed over wider areas (Fig. 10d). This result highlights the potential of model error sampling in specific situations in which local-scale factors, such as the orography play a meaningful role, although in general these perturbations are not sufficient to represent risk scenarios compatible with the inherent uncertainties. Small differences among techniques including IC/LBC perturbations can be attributed to slight shifts of rainfall probability.

When a larger threshold of 50 mm in 3 h is analyzed, differences between STO and ensembles including IC/LBC perturbations are reduced, especially for cases 1 and 3 (Fig. 9b,d,f). In the analyzed cases, experiments exclusively based on model perturbations have similar performance than IC based ensembles when forecasting extreme precipitation. However, hit rates, even for low probability thresholds are low for all experiments, indicating that the different perturbations included do not adequately sample all relevant uncertainties necessary to capture extreme events, especially for Case1. Only for Case 3, in which the orographic forcing is more influential, the skill of the different techniques tested is slightly higher. Efforts to increase statistical resolution (i.e., the number of ensemble members) should be considered to better capture convective-scale extreme phenomena (Raynaud and Bouttier, 2017), despite the additional computational cost, as well as the improvement of the representation of relevant physical processes at these scales, considering the substantial effect of model perturbations in areas where deep convection is developing.

### 4.3. Correspondence ratio

The spread of the precipitation fields is analyzed by means of the correspondence ratio. The five different ensemble generation strategies considered exhibit common characteristics in terms of correspondence

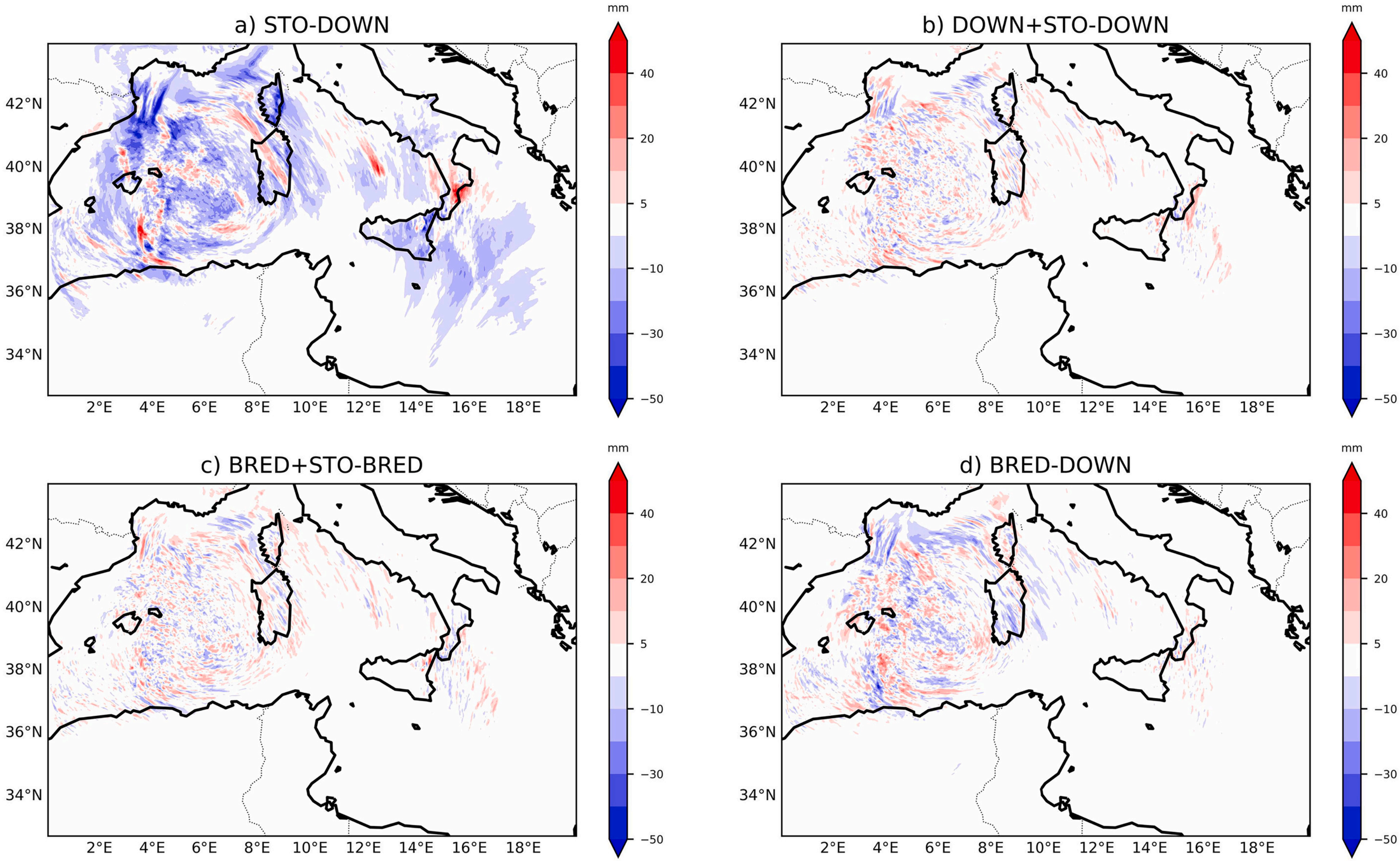


**Fig. 4.** As in Fig. 3 a)-d) for 24-h accumulated precipitation between 19 and 20 December 2016 at 1800 UTC.

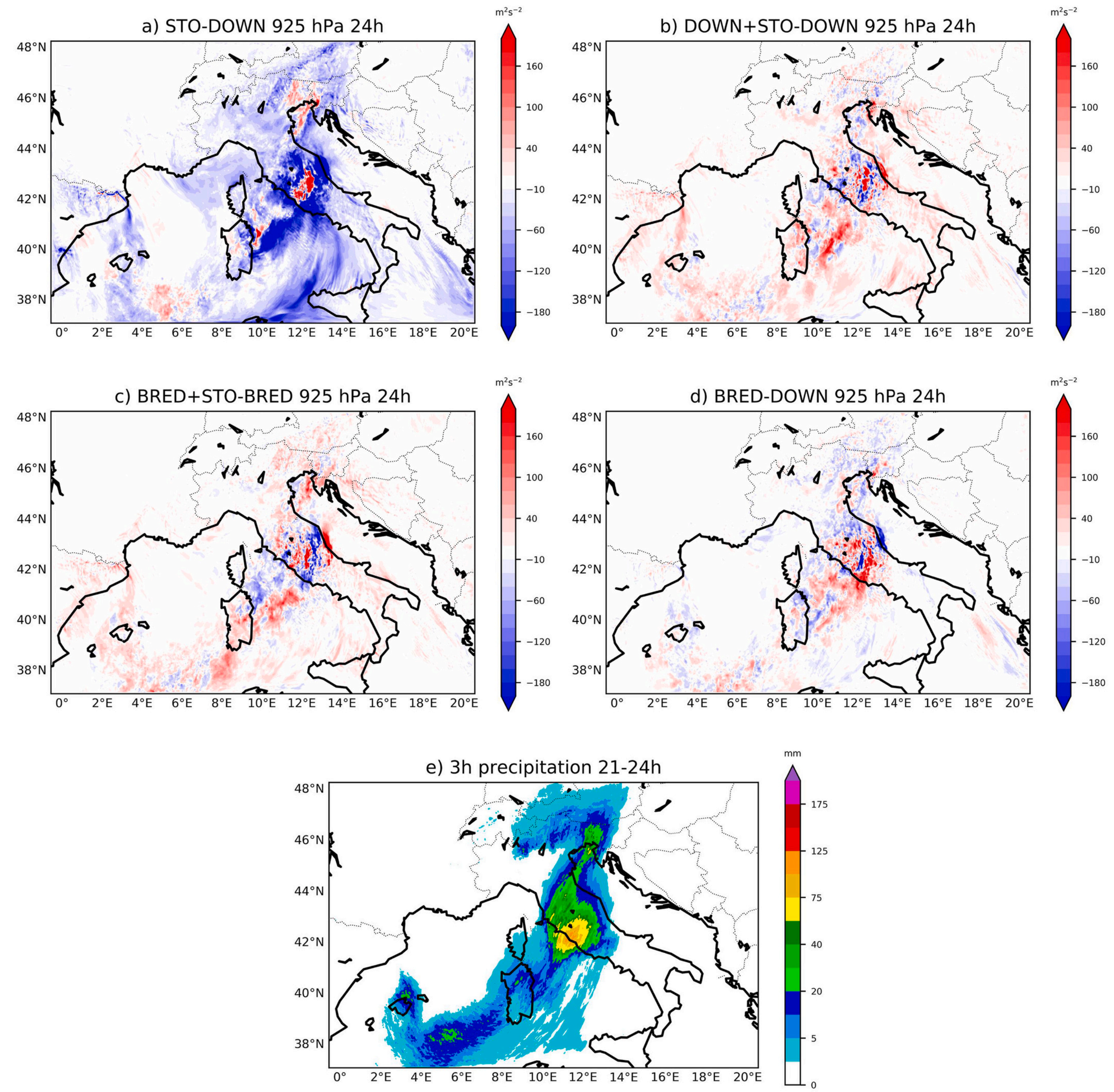


**Fig. 5.** As in Fig. 3 for Case 2.

ratio. This metric and thus the ensemble spread generally decreases with lead time as expected for a chaotic system (Fig. 11). Concerning the characteristics displayed by each ensemble generation strategy, STO has the largest CR for the three case studies analyzed and for both selected precipitation thresholds, which is consistent with the perturbation characteristics inspected in the previous section. This indicates that these perturbations generate lower spread compared to IC/LBC perturbations. These differences are slightly reduced in Case 3 (Fig. 11e,f), which is characterized by more localized convection, suggesting that the impact of model error sampling is larger in this situation. While the dispersion of STO is smaller, when stochastic perturbations are combined with IC/LBC perturbations a reduction in CR is obtained for all case studies and lead times. As a consequence of the enhanced relevance of stochastic perturbations in Case 3, this effect is especially appreciable in this event. By contrast, these differences are minor in Case 1 (Fig. 11a, b), in which larger scale structures are more relevant and thus, initial condition sampling has a higher impact on the ensemble spread. A decrease in CR is produced when $P$ is increased, indicating that the probability of exceeding the selected rainfall threshold is low, which indicates a certain diversity in terms of precipitation intensity and/or location among the ensemble. This diversity is required to successfully capture extreme risk scenarios. Focusing on the differences between the case studies considered, although a similar pattern is obtained for all of them, specific differences are found. In Case 3, CR is generally lower than for the other two case studies, especially for the 50 mm threshold which illustrates the larger uncertainties associated with this type of heavy precipitation event.

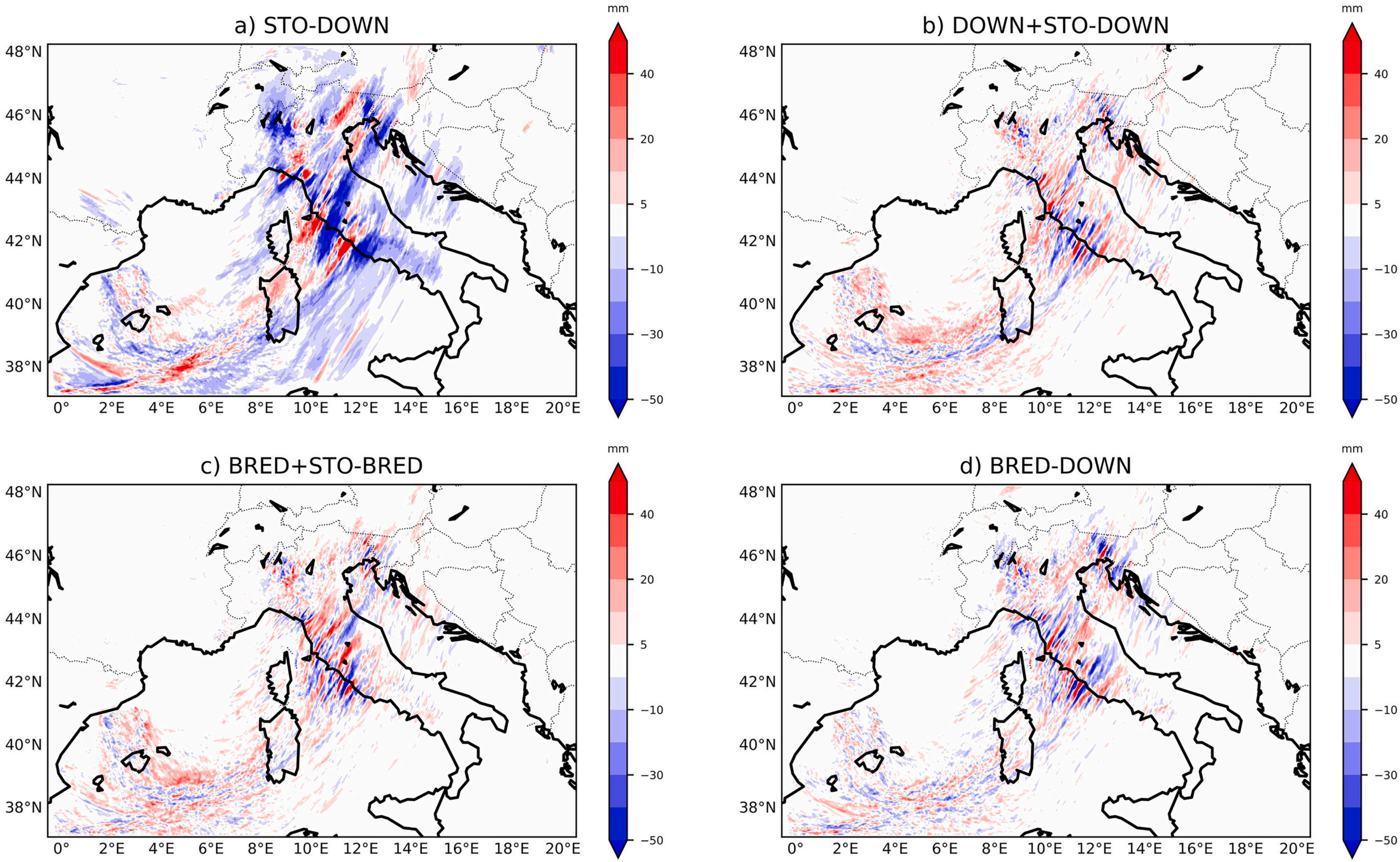


**Fig. 6.** As in Fig. 4. Accumulation period between 9 and 10 September 2017 at 1800 UTC.

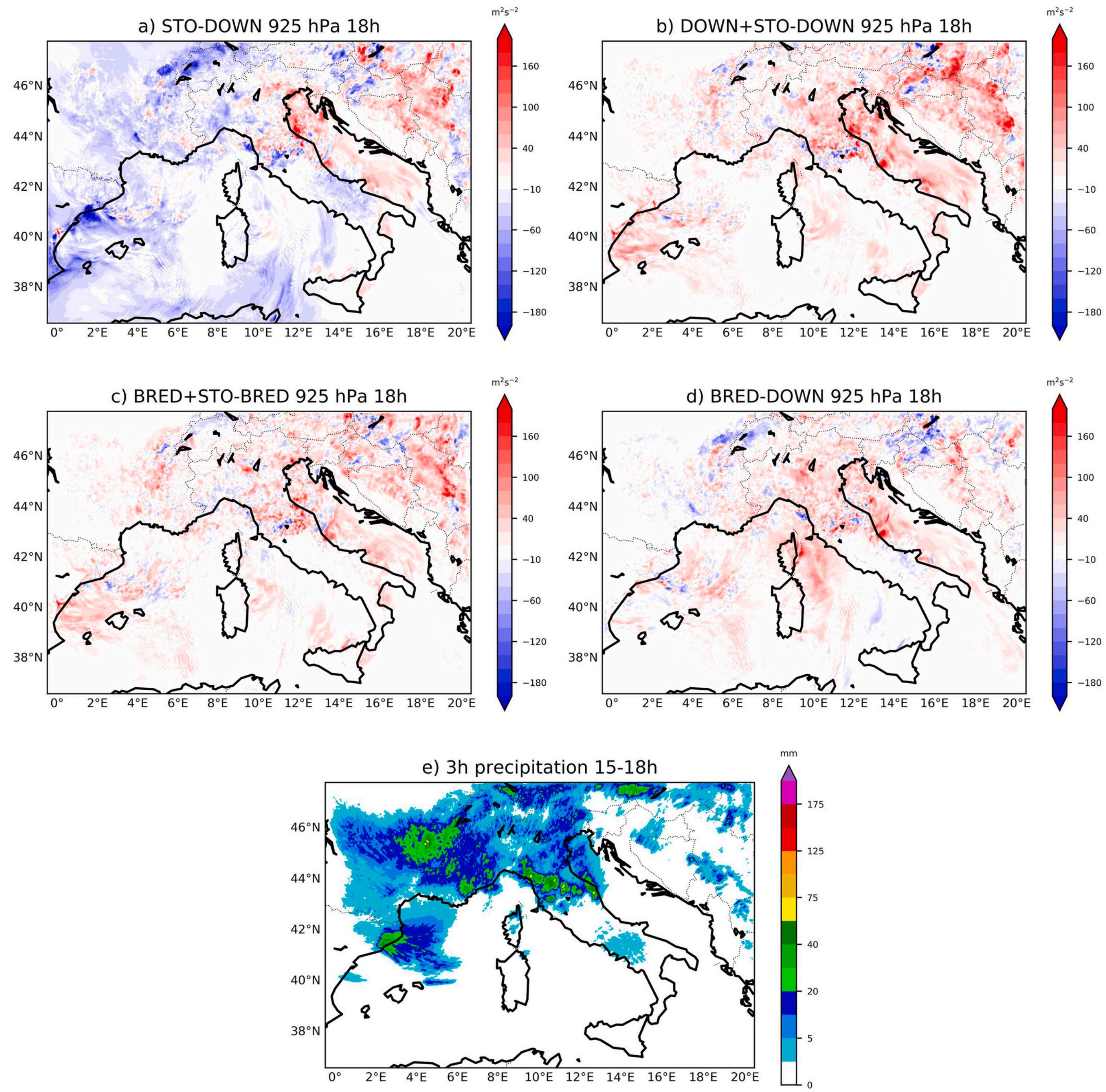


**Fig. 7.** As in Fig. 3 for Case 3.

### 4.4. *Ensemble dimension*

The ensemble dimension quantifies the degrees of freedom achieved by each ensemble regardless of the perturbation amplitude. Stochastic perturbations substantially contribute to increasing diversity at low levels after the spin-up period (Fig. 12a,c,e). This is a natural consequence of not introducing perturbations at the forecast start time (i.e., ensemble dimension is 1), so some time is needed for the perturbations to generate diversity. The fact that at low levels STO substantially features higher dimension than the ensembles including IC/LBC perturbations, but has a lesser impact at higher levels is related to the process being perturbed. The effect of perturbing PBL tendencies through the SPPT scheme is largely concentrated on the lowest 1000 m, while perturbations to microphysics can have impacts at higher levels, although this influence is focused on particular areas affected by deep moist convection. This circumstance contributes to maintaining high dimension at upper levels, especially during the second half of the forecast (Fig. 12b,d,f). Regarding the combination of IC/LBC and model perturbations, these results show an increase in ensemble dimension when perturbations to physical parameterizations are added to IC/LBC sampling for most lead times in the three case studies, especially at low levels, while at higher levels differences become smaller. This result is consistent with the increase in diversity obtained for the correspondence ratio. However, this effect is not uniform, as for large lead times in Case 3, which are characterized by more localized convection, the combination of both perturbations results in a reduction of ensemble dimension. It must be noted that in the highly nonlinear regime in which we are focusing, combining perturbations does not necessarily produce larger

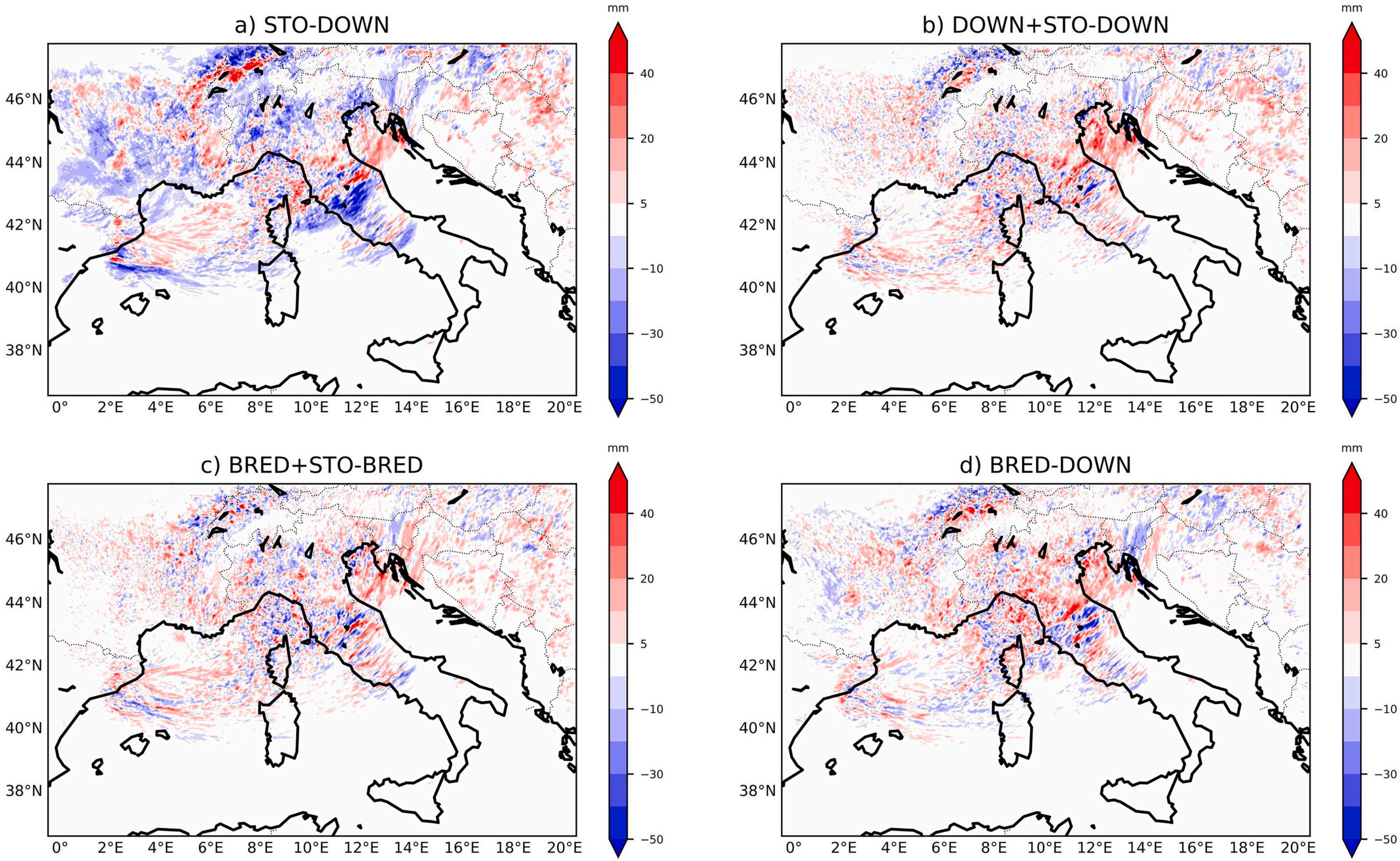


**Fig. 8.** As in Fig. 4. Accumulation period between 27 and 28 July 2019 at 0600 UTC.

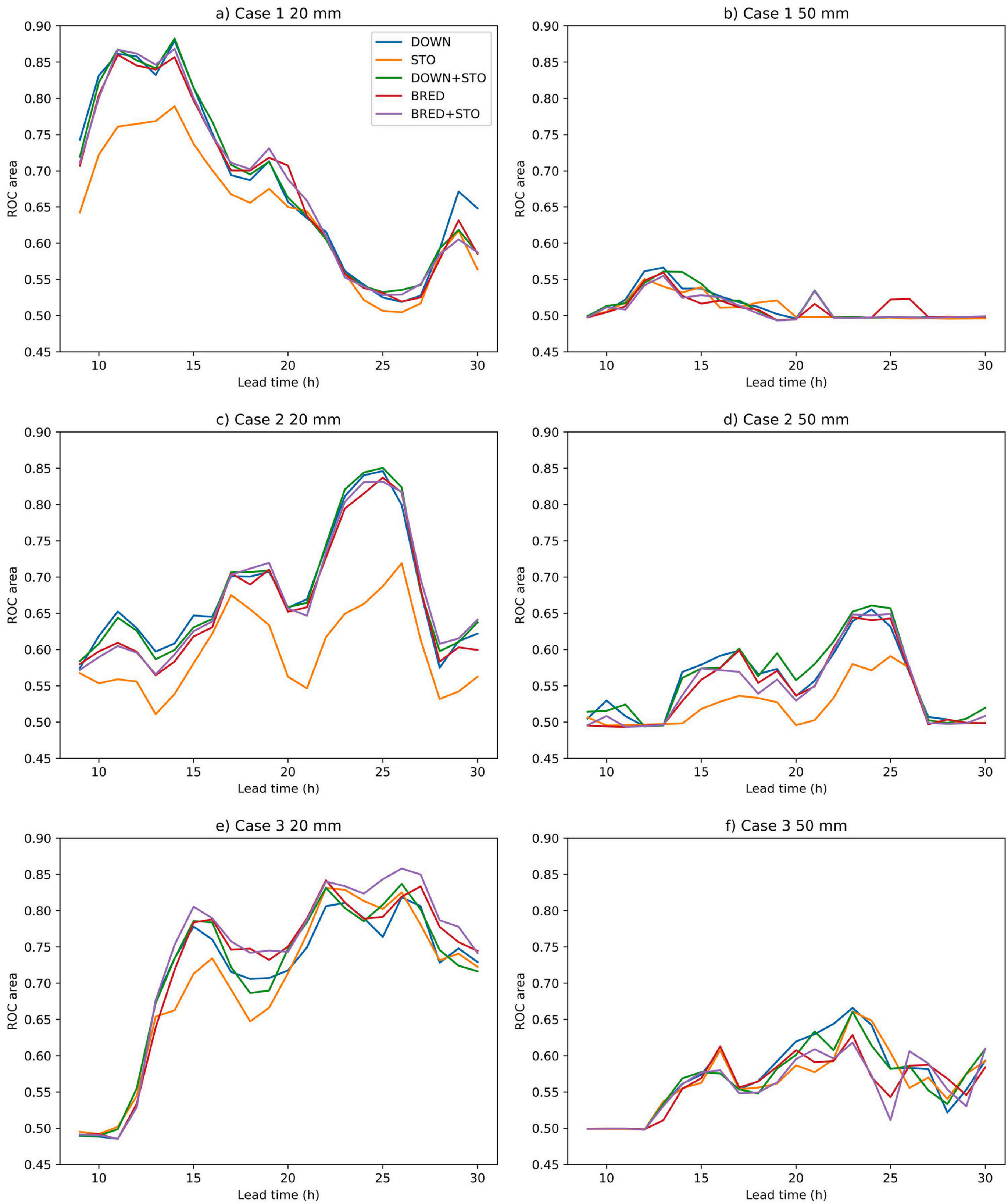


**Fig. 9.** Evolution of the area under the ROC for 3-h accumulated precipitation as a function of lead time for 20 mm (left column) and 50 mm (right column) thresholds.

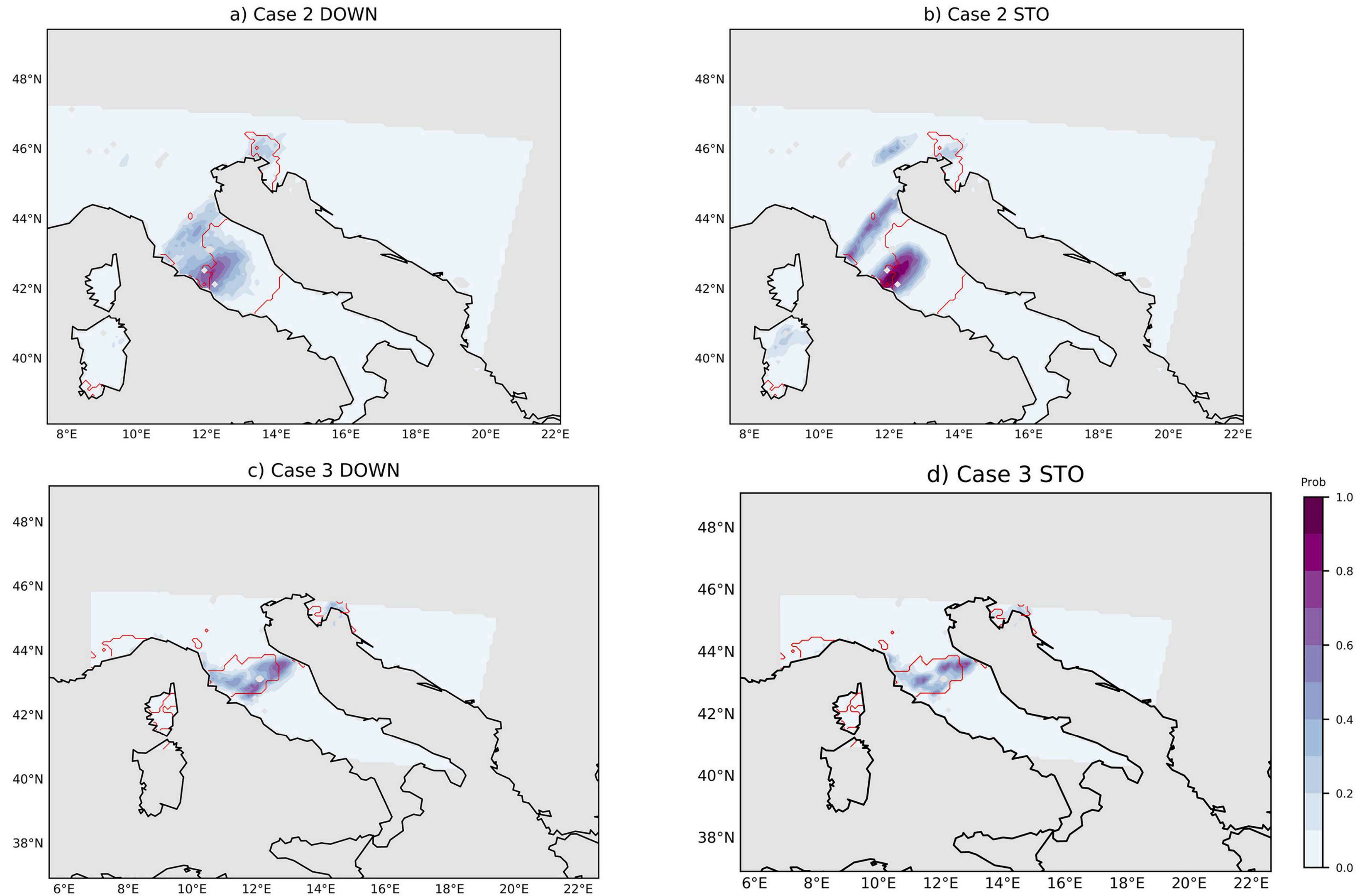


**Fig. 10.** Probability of 3-h accumulated precipitation > 20 mm for DOWN (left column) and STO (right column) for Case 2 (between 0900 and 1200 UTC; upper panels) and Case 3 (between 0700 and 1000 UTC; lower panels). The region that surpass the threshold in the GPM IMERG observations is indicated by red contours and the area not considered for the verification is grayed. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

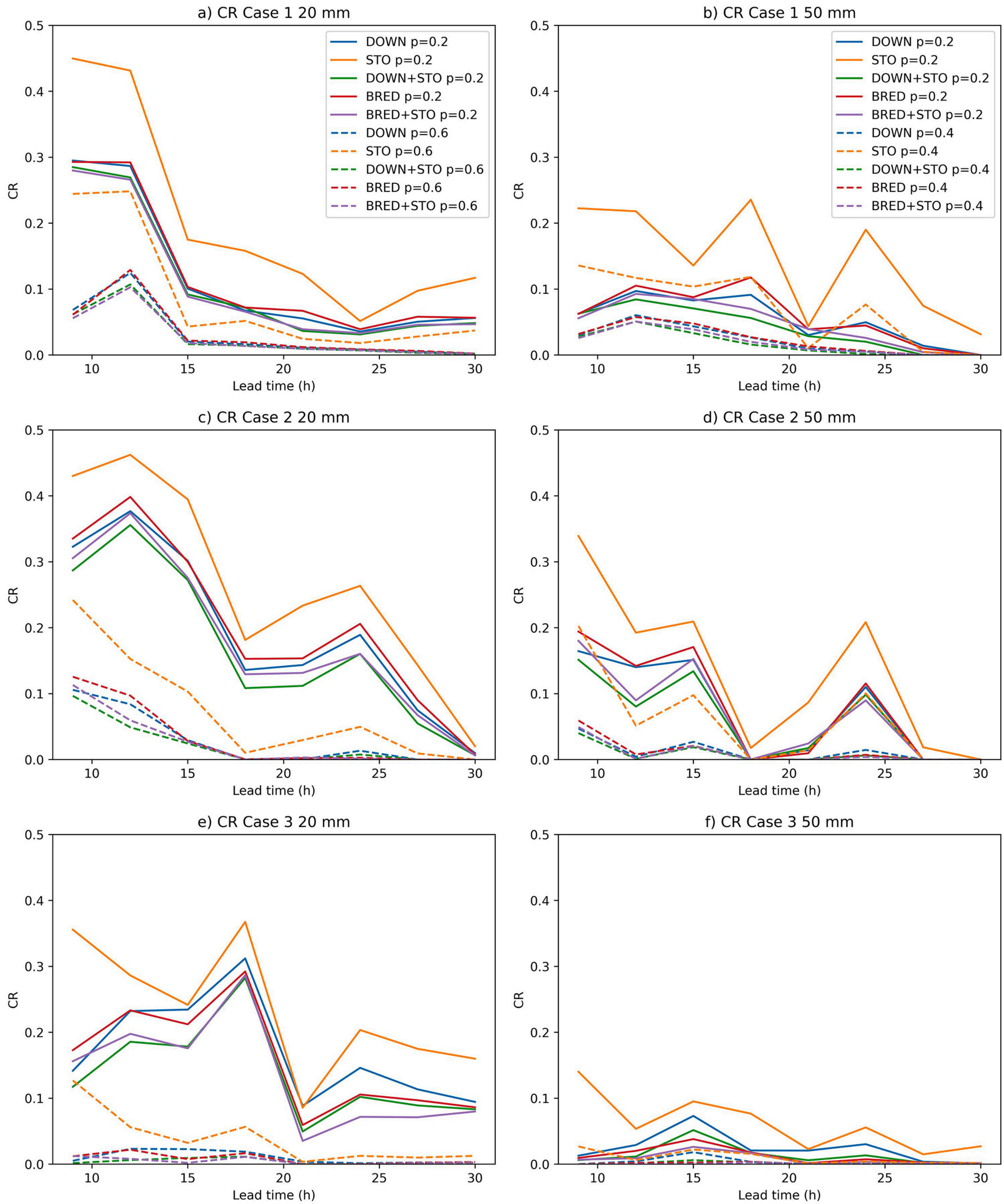


**Fig. 11.** Correspondence ratio for 3-h accumulated precipitation with a threshold of a), c), e) 20 mm and b), d), f) 50 mm for the three case studies. Solid lines represent a probability threshold $P$ of 0.2 in all panels and dashed lines depict $P = 0.6$ in the left column panels and $P = 0.4$ in the right column.

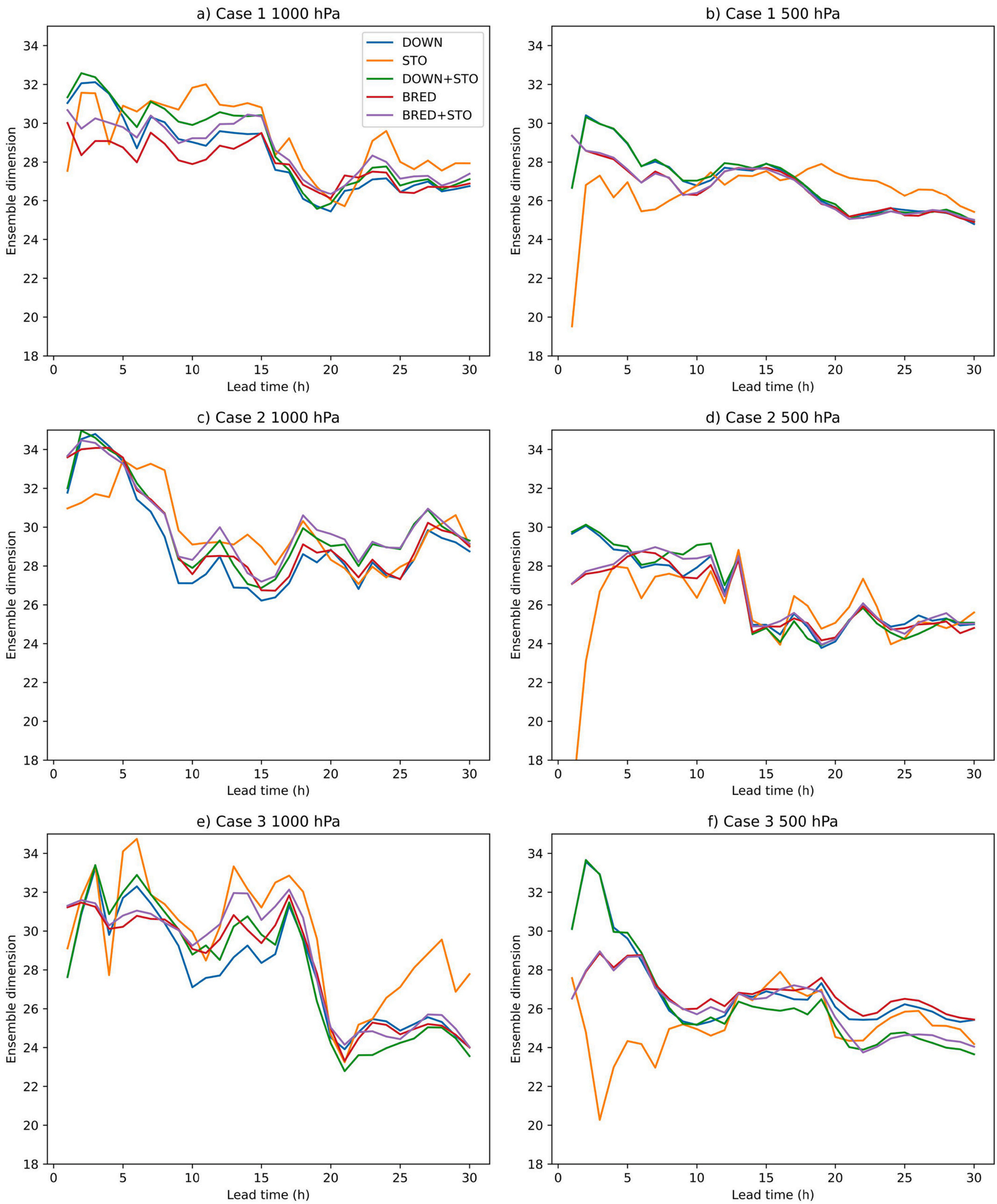


**Fig. 12.** Ensemble dimension of moist total energy perturbations as a function of lead time for each ensemble experiment at 1000 hPa (left column) and 500 hPa (right column) for the three different selected case studies (rows).

diversity. Indeed, ensembles including only stochastic perturbations to subgrid parameterizations have generally larger ensemble dimension. The reason for this behavior is related to the way in which perturbations sampling each error source are produced. IC perturbations focus on structures with a certain spatial correlation, such as fronts or low pressure systems, especially in DOWN, while stochastic perturbations at short lead times display patterns with low spatial correlation, which are controlled by the parameters of the SPPT scheme, and grow over areas of intense convection. This random character of the perturbation, which is intensified by the microphysics processes, generates larger linear independence among ensemble members than IC/LBC perturbations. Consequently, given the lower amplitude and earlier saturation of the stochastic perturbations, when both perturbations are combined, the impact of IC/LBC dominates the ensemble dimension. Therefore, value of ensemble dimension is reduced compared to the ensemble built exclusively on stochastic perturbations.

A drop in ensemble dimension can be observed for all cases, especially at low levels. The fact that this behavior is common for events characterized by different atmospheric settings suggests that this reduction is produced by a collapse of perturbations towards dominant growing modes. Indeed, at short lead times, when ensemble dimension is generally higher, the presence of small-scale perturbations is more frequent due to the initial development of convective activity after the spin-up period, increasing linear independence among the whole set of perturbations. By contrast, at longer lead times, perturbations acquire a larger spatial correlation, reducing their linear independence, and thus their dimension.

### 4.5. Amplitude and localization

An additional important aspect of the perturbations is related to their amplitude and scale. The quantification of these magnitudes by means of $\log\rho$ and $\omega^2$ reveals the different nature of each kind of ensemble perturbation (Fig. 13. STO has considerably lower amplitude and higher localization than the other experiments including IC/LBC variations. Indeed, these perturbations, especially the part affecting microphysics processes are highly localized perturbations with low amplitude at early forecast times. They significantly grow over particular areas in which deep moist convection develops (Thompson et al., 2021), acquiring larger spatial correlation (lower $\omega^2$) at low levels during the first part of the forecast, especially for Case 1 and 2 (Fig. 13). This evolution consisting in a rapid growth of small-scale perturbations and saturation at low amplitudes is consistent with the conceptual scheme of Toth and Kalnay (1997), and quantified in various studies (e.g., Judt et al., 2016; Selz and Craig, 2015). However, at upper levels, the localization of perturbations increases with lead time. The effect of PBL perturbations is smaller and the effect of microphysics perturabations is noticeable when deep convective systems are sufficiently developed, generating localized perturbations aloft. A similar localization evolution holds for Case 3 at low levels. This can be related to the characteristics of the convective systems for this event, which are much more scattered than for the other two case studies, which limits the spatial correlation of the perturbations. By contrast, large-scale perturbations inherited from the ECMWF-EPS in DOWN and DOWN+STO, have high initial amplitude and low localization, as they are focused over larger-scale structures, such as the low pressure system in Case 1 or the position of the upper-level trough, lee cyclone and the contrast between air masses in Case 2. Regarding tailored bred perturbations, their progression is similar to that of DOWN perturbations, except for the spin-up period, in which perturbations adapt to model dynamics. Indeed, a downside of the tailored bred technique is that the exponential modification of bred vectors eliminates the finite fluctuation character of the perturbation. That is, the perturbation applied is not directly obtained from a model run as with the traditional breeding method. However, as shown by Hermoso et al. (2020), despite this issue, orthogonalization and scale modification substantially increase ensemble performance, overcoming the typical collapse of bred vectors. Although there are appreciable differences among the three investigated cases, the $\log\rho - \omega^2$ characterization shows more consistent behavior across the different selected cases than ensemble dimension, indicating that these features are more generalizable. For low-levels, the amplitude of perturbations including IC/LBC sampling grows and their scale decreases during the first part of the forecast and increases afterwards. The STO perturbations increase their amplitude, but also their initial scale. At upper levels, all ensembles follow a similar pattern but STO lies on a different area of the $\log\rho - \omega^2$ space.

## 5. Conclusions

The potential and structural characteristics of multiple ensemble generation strategies to forecast severe weather episodes, especially heavy precipitation events, has been analyzed. In particular, the characteristics of the ensemble resulting from the application of IC/LBC perturbations, either downscaling from a global model or generated by means of tailored bred vectors following Hermoso et al. (2020) have been examined. In addition, the effect of perturbations on subgrid parameterizations by means of a combination of SPPT and a method used to perturb some relevant microphysics parameters described in Hermoso et al. (2021b) has also been investigated. The study focuses on the characteristics of each type of perturbation separately or in combination for three heavy precipitation episodes that affected the western Mediterranean basin in the last decade.

Initial and boundary condition perturbations are found to provide substantial diversity after a spin-up period, despite the initial lack of variability at the small scales, which exerts a significant influence on the genesis and evolution of convective systems. These perturbations concentrate over larger systems, such as low pressure systems or fronts, but after a few simulation hours, perturbations become more intense and localized, as revealed by the $\log\rho - \omega^2$ analysis.

The introduction of perturbations covering multiple scales by means of tailored bred vectors generally increases ensemble diversity, especially at low levels. However, this behavior is case dependent in such a way that downscaled perturbations yield more diversity when larger-scale dynamics dominates, such as the case characterized by a deep cyclone. Bred perturbations undergo a sharp variation in amplitude and localization at short lead times. Since the applied orthogonalization and exponential transformation push perturbations away from finite fluctuations, the part of the perturbation that does not project onto dynamical growing modes decays. Despite this drawback, the method generates more diverse and skillful forecasts and for the events analyzed in this study, the ensemble diversity produced by the tailored bred vectors technique is generally higher, especially at low levels and for Case 2 and 3, wherein smaller-scale characteristics are more relevant. It should also be noted that in the experiment design used herein, DOWN and BRED share LBC perturbations, reducing differences between these experiments for long lead times.

The greatest contrasts between ensemble strategies emerge when perturbations are applied only to subgrid parameterizations. Although these perturbations generate substantial diversity at low levels, measured as the linear independence among perturbations, which is larger than that attained with IC/LBC perturbations, their amplitude is still small. Indeed, for short lead times, model perturbations exhibit lower spatial correlation and intensity than their IC counterparts, and grow over areas with intense convective developments, exceeding initial condition perturbations at specific locations. However, the verification of precipitation fields reveals that only introducing perturbations to physical parameterizations results in lower spatial spread in the rainfall fields, leading to overconfident forecasts. Furthermore, rainfall accumulations are generally too low when only model perturbations are used. Therefore, it is not advisable to design an ensemble prediction system based on these perturbations only. However, these downsides are lessened when scattered convection is predominant, indicating the

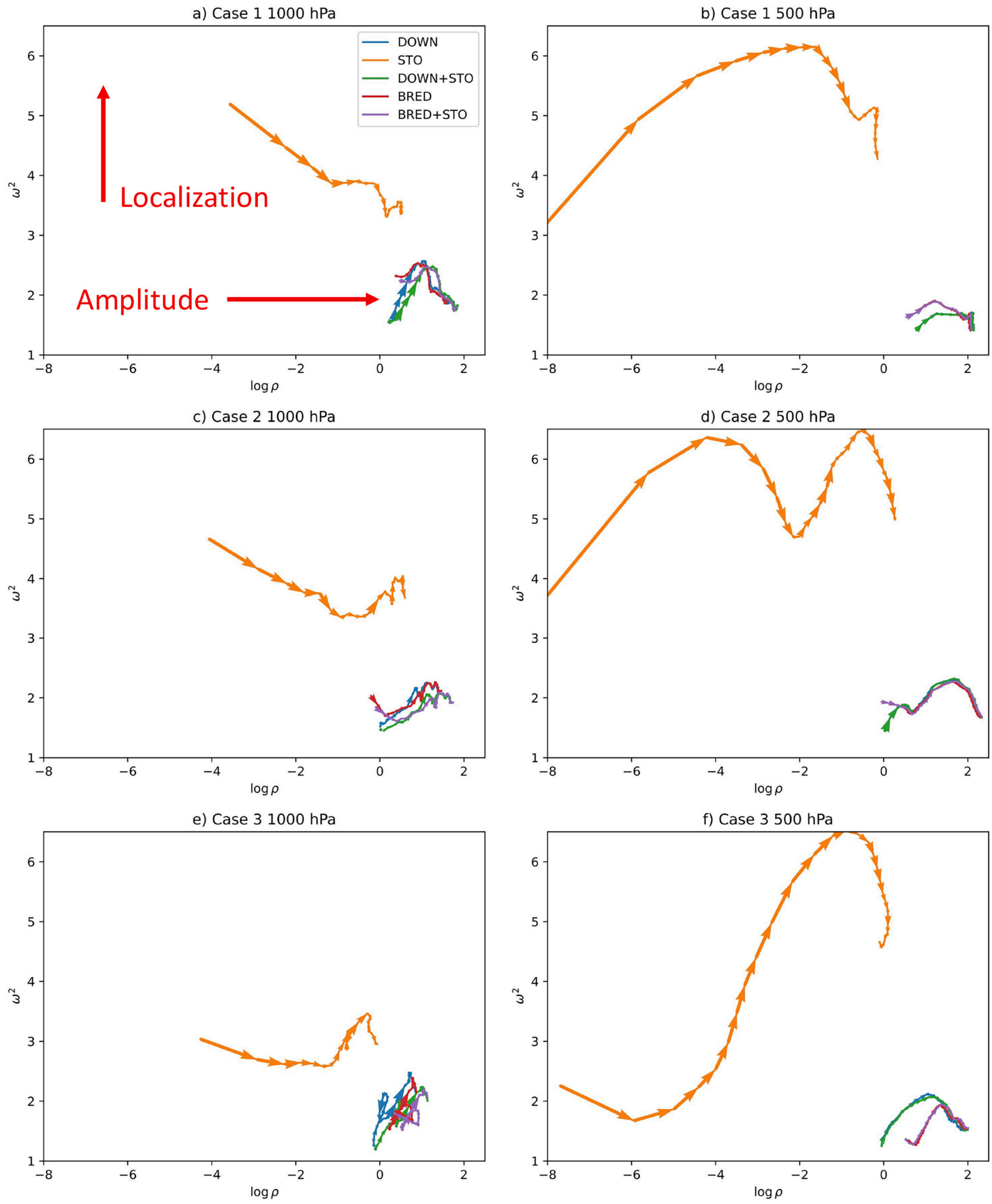


**Fig. 13.** Evolution of mean $\log\rho - \omega^2$ at 1000 hPa (left column) and 500 hPa (right column) for the different case studies considered. Arrows in each trajectory represent the temporal evolution of $\log\rho - \omega^2$ and red arrows in panel a) indicate the characteristic displayed in each axis. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

potential to generate growing perturbations in this situation.

The combination of IC/LBC and model error sampling increases ensemble dimension and perturbation intensity in most cases for both DOWN and BRED, particularly at low levels. Although the combination of generation techniques consistently produces more diverse perturbations, the effects of sampling multiple uncertainty sources under the nonlinear dynamics, which dominates the convective-scale regime, may not necessarily be additive. In the three investigated case studies diversity produced by IC/LBC perturbations is generally increased when model error sampling is introduced, although at certain times combining model perturbations negatively impacts ensemble diversity.

While the number of case studies considered is too low to draw statistical conclusions, the characteristics of the different error sampling strategies investigated in this study provide some guidelines for ensemble design. Indeed, the dependency of the benefits of each method on the particular atmospheric setting that leads to the heavy precipitation event recommends use of a flexible environment in which multiple ensemble strategies could be used depending on the specificities of each event. Nevertheless, extending this study to a large sample of events including a thorough verification would enable the extraction of statistically significant findings.

The ensemble generation strategies considered, especially stochastic parameterizations, include parameters that can be tuned to optimize performance. In this sense, idealized experiments to study multiple combinations of the adjustable parameters could reveal insightful details regarding the specific influence of each process.

Additionally, data assimilation techniques, which have not been considered in this study, should be included in future analyses, in particular in combination with stochastic methods. Considering the large maritime bodies present in the Mediterranean region, data assimilation including satellite and/or radar information can produce substantial improvements in the forecast of convective systems, which are often initiated over the sea. In this framework, the characteristics of multiple combinations of data assimilation and stochastic sampling methods can provide relevant information for the enhancement of convective-scale ensemble prediction systems in the Mediterranean basin.

## CRediT authorship contribution statement

**Alejandro Hermoso :** Conceptualization, Methodology, Software, Investigation, Writing – original draft, Visualization. **Victor Homar :** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Romualdo Romero :** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgments

This work was sponsored by: Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación/TRAMPAS (PID2020-113036RB-I00 / AEI / 10.13039/501100011033) and COASTEPS (CGL2017-82868-R). The authors thankfully acknowledge the computer resources granted at MareNostrum 4 and the technical support provided by Barcelona Supercomputing center (RES-AECT-2019-3-0012) as well as for the use of ECMWF's computing and archive facilities in this research. Two anonymous reviewers are acknowledged for their insightful comments, which helped to improve the quality of the paper.

## References

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., 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. Am. Meteorol. 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. ECMWF. https://doi.org/10.21957/tx1epjd2p.
Bretherton, C.S., Widmann, M., Dymnikov, V., Wallace, J., Bladé, I., 1999. The effective number of spatial degrees of freedom of a time-varying field. J. Climate 12, 1990–2009. https://doi.org/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2.
Bryan, G., Wyngaard, J., Fritsch, J., 2003. Resolution requirements for the simulation of deep moist convection. Mon. Weather Rev. 131, 2394–2416. https://doi.org/10.1175/1520-0493(2003)131<2394:RRFTSO>2.0.CO;2.
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.
Buzzi, A., Tibaldi, S., 1978. Cyclogenesis in the lee of the Alps: a case study. Quart. J. Royal Meteor. Soc. 104, 271–287. https://doi.org/10.1002/qj.49710444004.
Buzzi, A., Davolio, S., Fantini, M., 2020. Cyclogenesis in the lee of the Alps: a review of theories. Bull. Atmos. Sci. Technol. 1, 433–457. https://doi.org/10.1007/s42865-020-00021-6.
Capecchi, V., Antonini, A., Benedetti, R., Fibbi, L., Melani, S., Rovai, L., Ricchi, A., Cerrai, D., 2021. Assimilating X- and S-band radar data for a heavy precipitation event in Italy. Water 13, 1727. https://doi.org/10.3390/w13131727.
Caumont, O., Mandement, M., Bouttier, F., Eeckman, J., Brossier, C., Lovat, A., Nuissier, O., Laurantin, O., 2021. The heavy precipitation event of 14-15 October 2018 in the Aude catchment: a meteorological study based on operational numerical weather prediction systems and standard and personal observations. Nat. Hazards Earth Syst. Sci. 21, 1135–1157. https://doi.org/10.5194/nhess-21-1135-2021.
Centre for Research on the Epidemiology of Disasters (CRED), 2021. 2020 the Non-covid Year in Disasters: Global Trends and Perspectives. URL. https://dial.uclouvain.be/pr/boreal/object/boreal:245181.
Davolio, S., Miglietta, M., Diomede, T., Marsigli, C., Montani, A., 2013. A flood episode in northern Italy: Multi-model and single-model mesoscale meteorological ensembles for hydrological predictions. Hydrol. Earth Syst. Sci. 17, 2107–2120. https://doi.org/10.5194/hess-17-2107-2013.
Davolio, S., Ferra, S., Laviola, S., Miglietta, M., Levizzani, V., 2020. Heavy precipitation over Italy from the mediterranean storm "Vaia" in October 2018: Assessing the role of an atmospheric river. Mon. Weather Rev. 148, 3571–3588. https://doi.org/10.1175/MWR-D-20-0021.1.
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, Am. Meteor. Soc., Chicago. 14 pp. URL: https://ams.confex.com/ams/27WAF23NWP/webprogram/Paper273421.html.
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.
European Centre for Medium-Range Forecasts (ECMWF), 2020. Part V: Ensemble prediction system. In: IFS Documentayion - Cy47r1. ECMWF, p. 23. URL. https://www.ecmwf.int/node/19749.
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.
Federico, S., Claudia Torcasio, R., Avolio, E., Caumont, O., Montopoli, M., Baldini, L., Vulpiani, G., Dietrich, S., 2019. Improvement of RAMS precipitation forecast at the short-range through lightning data assimilation. Nat. Hazards Earth Syst. Sci. 19, 1839–1864. https://doi.org/10.5194/nhess-19-1839-2019.
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.

Flack, D.L.A., Clark, P.A., Halliwell, C.E., Roberts, N.M., Gray, S.L., Plant, R.S., Lean, H. W., 2021. A physically based stochastic boundary layer perturbation scheme. Part ii: perturbation growth within a superensemble frame-work. J. Atmos. Sci. 78, 747–761. https://doi.org/10.1175/JAS-D-19-0291.1.

Flora, M., Potvin, C., Wicker, L., 2018. Practical predictability of supercells: Exploring ensemble forecast sensitivity to initial condition spread. Mon. Weather Rev. 146, 2361–2379. https://doi.org/10.1175/MWR-D-17-0374.1.

Frogner, I.L., 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-MuÃoz, 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.

Gebhardt, C., Theis, S., Paulat, M., Ben Bouallègue, Z., 2011. Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. Atmos. Res. 100, 168–177. https://doi.org/10.1016/j.atmosres.2010.12.008.

Hacker, J., Ha, S., Snyder, C., Berner, J., Eckel, F., Kuchera, E., Pocernich, M., Rugg, S., Schramm, J., Wang, X., 2011. 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.

Haiden, T., Janousek, M., Bidlot, J.R., Ferranti, L., Prates, F., Vitart, F., Bauer, P., Richardson, D., 2016. Evaluation of ECMWF Forecasts, Including the 2016 Resolution Upgrade. Technical Memorandum. ECMWF. https://doi.org/10.21957/vltarm3ql.

Han, J., Hong, S., 2018. Precipitation forecast experiments using the Weather Research and forecasting (WRF) model at gray-zone resolutions. Wea. Forecasting 33, 1605–1616. https://doi.org/10.1175/WAF-D-18-0026.1.

Hermoso, A., Homar, V., Greybush, S., Stensrud, D., 2020. Tailored ensemble prediction systems: application of seamless scale bred vectors. J. Meteor. Soc. Jpn. 98, 1029–1050. https://doi.org/10.2151/jmsj.2020-053.

Hermoso, A., Homar, V., Amengual, A., 2021a. 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. Hydrometeorol. 22, 1117–1138. https://doi.org/10.1175/JHM-D-20-0182.1.

Hermoso, A., Homar, V., Plant, R., 2021b. Potential of stochastic methods for improving convection-permitting ensemble forecasts of extreme events over the western mediterranean. Atmos. Res. 257, 105571 https://doi.org/10.1016/j.atmosres.2021.105571.

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.

Huffman, G., Stocker, E., Bolvin, D., Nelkin, E., Tan, J., 2019. GPM IMERG early precipitation L3 half hourly 0.1 degree x 0.1 degree V06. In: Goddard Earth Sciences Data and Information Services Center (GES DISC). URL: https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGHHE_06/summary?keywords=rainfall.

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.

Insua-Costa, D., Lemus-Cánovas, M., Miguez-Macho, G., Llasat, M., 2021. Climatology and ranking of hazardous precipitation events in the western Mediterranean area. Atmos. Res. 155, 105521 https://doi.org/10.1016/j.atmosres.2021.105521.

Jankov, I., Berner, J., Beck, J., Jiang, H., Olson, J., Grell, G., Smirnova, T., Benjamin, S., Brown, J., 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.

Johnson, A., Wang, X., 2020. Interactions between physics diversity and multiscale initial condition perturbations for storm-scale ensemble forecasting. Mon. Weather Rev. 148, 3549–3655. https://doi.org/10.1175/MWR-D-20-0112.1.

Judt, F., Chen, S., Berner, J., 2016. Predictability of tropical cyclone intensity: scale-dependent forecast error growth in high-resolution stochastic kinetic-energy backscatter ensembles. Quart. J. Royal Meteor. Soc. 142, 43–57. https://doi.org/10.1002/qj.2626.

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.

Lagouvardos, K., Kotroni, V., Defer, E., Bousquet, O., 2013. Study of a heavy precipitation event over southern France, in the frame of HYMEX project: Observational analysis and model results using assimilation of lightning. Atmos. Res. 134, 45–55. https://doi.org/10.1016/j.atmosres.2013.07.003.

Lambert, D., Argence, S., 2008. Preliminary study of an intense rainfall episode in Corsica, 14 September 2006. Adv. Geosci. 16, 125–129. https://doi.org/10.5194/adgeo-16-125-2008.

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. 147, 1364–1381. 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.

Llasat, M.C., Puigcerver, M., 1994. Meteorological factors associated with floods in the North-Eastern part of the Iberian Peninsula. Nat. Hazards 9, 81–93. https://doi.org/10.1007/978-94-011-0976-5_5.

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.

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.

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.

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.

Markowski, P., 2020. What is the intrinsic predictability of tornadic supercell thunderstorms? Mon. Weather Rev. 148, 3157–3180. https://doi.org/10.1175/MWR-D-20-0076.1.

Martín, M., Santos-Muñoz, D., Morata, A., Luna, M.Y., Valero, F., 2006. An objectively selected case study of a heavy rain event in the Mediterranean Basin: a diagnosis using numerical simulation. Atmos. Res. 81, 187–205. https://doi.org/10.1016/j.atmosres.2005.12.002.

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

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.

Michaelides, S., Karacostas, J., Sánchez, J.L., Retalis, A., Pytharoulis, I., Homar, V., Romero, R., Zanis, P., Giannakopoulos, C., Bühl, J., Ansmann, A., 2018. Reviews and perspectives of high impact atmospheric processes in the Mediterranean. Atmos. Res. 208, 4–44. https://doi.org/10.1016/j.atmosres.2017.11.022.

Mittermaier, M., Roberts, N., Thompson, S., 2013. A long-term assessment of precipitation forecast skill using the Fractions Skill score. Meteorol. Appl. 20, 176–186. https://doi.org/10.1002/met.296.

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.

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.

Navarro, A., García-Ortega, E., Merino, A., Sánchez, J., Kummerow, C., Tapiador, F., 2019. Assessment of IMERG precipitation estimates over Europe. Remote Sens. 11, 2470. https://doi.org/10.3390/rs11212470.

Nuissier, O., Ducrocq, D., 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.

Patil, D., Hunt, B., Kalnay, E., Yorke, J., Ott, E., 2001. Local low dimensionality of atmospheric dynamics. Phys. Rev. Lett. 86, 5878–5881. https://doi.org/10.1103/PhysRevLett.86.5878.

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.

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, 11. Environmental Science and Technology Library. Springer, Dordrecht, pp. 133–151. https://doi.org/10.1007/978-94-017-3354-0_13.

Raynaud, L., Bouttier, F., 2017. The impact of horizontal resolution and ensemble size for convective-scale probabilistic forecasts. Quart. J. Royal Meteor. Soc. 143, 3037–3047. https://doi.org/10.1002/qj.3159.

Romero, R., Ramis, C., Homar, V., 2015. On the severe convective storm of 29 October 2013 in the Balearic Islands: observational and numerical study. Quart. J. Royal Meteor. Soc. 689, 1208–1222. https://doi.org/10.1002/qj.2429.

Romine, G., Schwartz, 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.

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.

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.

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.

Sun, Y., Zhang, F., 2016. 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.

Tapiador, F., Navarro, A., García-Ortega, E., Merino, A., Sánchez, J., Marcos, C., Kummerow, C., 2020. The contribution of rain gauges in the calibration of the IMERG product: results from the first validation over Spain. J. Hydrometeorol. 21, 161–182. https://doi.org/10.1175/JHM-D-19-0116.1.

Thompson, G., Berner, J., Frediani, M., Otkin, J., Griffin, S., 2021. A stochastic parameter perturbation method to represent uncertainty in a microphysics scheme. Mon. Weather Rev. 149, 1481–1497. https://doi.org/10.1175/MWR-D-20-0077.1.

Torcasio, R., Federico, S., Puca, S., Vulpiani, G., Comellas-Prat, A., Dietrich, S., 2020. Application of lightning data assimilation for the 10 October 2018 case study over Sardinia. Atmosphere 11, 541. https://doi.org/10.3390/atmos11050541.

Toth, Z., Kalnay, E., 1993. Ensemble forecasting at NMC: the generation of perturbations. Bull. Am. Meteor. Soc. 74, 2317–2330 doi:110.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2.

Toth, Z., Kalnay, E., 1997. Ensemble forecasting at NCEP and the breeding method. Mon. Weather Rev. 125, 3297–3319. https://doi.org/10.1175/1520-0493(1997)125<3297:EFANAT>2.0.CO;2.

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.

Zhang, F., Bei, N., Rotunno, R., Snyder, C., Epifanio, C., 2007. Mesoscale predictability of moist baroclinic waves: Convection-permitting experiments and multistage error growth dynamics. J. Atmos. Sci. 64, 3579–3594. https://doi.org/10.1175/JAS4028.1.

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.