



Exploring the benefits of a Hi-EnKF system to forecast an extreme weather event. The 9th October 2018 catastrophic flash flood in Mallorca

D.S. Carrió^{a,b,*}, A. Jansà^c, V. Homar^c, R. Romero^c, T. Rigo^d, C. Ramis^c, A. Hermoso^c, A. Maimó^c

^a School of Earth Sciences, The University of Melbourne, Parkville, Victoria, Australia

^b ARC Centre of Excellence for Climate Extremes, Australia

^c Meteorology Group, Department of Physics, Universitat de les Illes Balears, Palma de Mallorca, Spain

^d Servei Meteorològic de Catalunya. C. Berlin, Barcelona, Spain

ARTICLE INFO

Keywords:

Flash floods
Ensemble data assimilation
EnKF
Radar reflectivity observations
Predictability
State estimation
Nonlinear
Non-Gaussian

ABSTRACT

The afternoon of 9 October 2018 a catastrophic flash flood episode (above 500-y return period) hit the town of Sant Llorenç, in northeastern Mallorca (Spain), leading to 13 casualties and economic losses over 91 M€. Operational models, including HARMONIE-AROME run at the Spanish Meteorological Agency, missed this event and thus no relevant warnings were issued to take preventive actions. Satellite, radar and conventional observations indicate the presence of convective activity along a convergence line, forming during the morning east of Menorca and moving westwards. This convective line is identified as a key factor for the initiation of the moist deep convection responsible for the extreme rain rates associated to this episode. Indeed, the failure of operational numerical models in predicting the flash flood is most likely attributable to their inability to properly simulate the convergence line and the associated convection. To further understand the predictability of this case, we assess multiple different high-resolution ensemble generation strategies, including data assimilation. Results show that among all, only configurations that assimilate reflectivities have the potential to predict significant precipitation amounts over useful lead times in Sant Llorenç basin. In any case, simulated precipitation values are underestimated, and we prove that the predictability of the responsible moist convective systems drops off near the 3–4 h lead times. Numerical predictability challenges of this episode and limitations of the EnKF in these situations are discussed in detail. Overall, the study shows that accurate prediction of this episode is beyond the current prediction capabilities of state-of-the-art numerical models and data assimilation techniques.

1. Introduction

Highly populated coastal regions of the Western Mediterranean are frequently affected by heavy precipitation and flash flood events with important socio-economic impacts (Rivera and Riosalido, 1986; Riosalido et al., 1988; Sénési et al., 1996; Buzzi et al., 1998; Kotroni et al., 1999; Romero et al., 2000; Nuissier et al., 2008; Jansa et al., 2014; del Moral et al., 2020). However, the accurate forecast of such extreme weather events by operational numerical models is still nowadays an important challenge. These weather events are most often linked to deep moist convective systems, which are highly sensitive to complex and highly nonlinear small-scale processes. One of the greatest current

challenges in numerical prediction is the explicit or parameterized representation of physical processes, such as deep convection, and their interactions (e.g., with PBL). Obviously, the limitations in modelling atmospheric processes in detail at current grid resolutions, blinders our path towards reliable early warning systems. Commonly, extreme precipitation events in the Western Mediterranean initiate and mature over maritime regions, which are poorly sampled by routine observational means. This lack of in situ observations translates into poor representations of the atmospheric state, and hence negatively impacts the quality of the forecasts.

In this work, we focus our attention on a recent flash flood that took place in *Sant Llorenç des Cardassar* (Mallorca, see Figs. 1 and 3b) during

* Corresponding author at: School of Geography, Earth and Atmospheric Sciences, ARC Centre of Excellence for Climate Extremes, The University of Melbourne, Melbourne, Australia.

E-mail address: diego.carriocarrio@unimelb.edu.au (D.S. Carrió).

<https://doi.org/10.1016/j.atmosres.2021.105917>

Received 14 June 2021; Received in revised form 15 October 2021; Accepted 4 November 2021

Available online 14 November 2021

0169-8095/© 2021 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

the afternoon of 9 October 2018, killing 13 people and producing a total estimated damage at EUR 90 million (Lorenzo-Lacruz et al., 2019). Registered precipitation in the *Sant Llorenç* area exceeded 220 mm in 24 h and 160 mm in 180 min (Fig. 1). Radar-based estimates over neighbour mountains of the catchment indicate accumulations of 350 mm in 7 h (Lorenzo-Lacruz et al., 2019). Although this was a notable event that affected various basins over eastern Mallorca, operational models run at the Spanish Meteorological Agency (AEMET) failed at reproducing the initiation and interaction of the convective systems linked to the intense rainfalls. Consequently, population could not be warned by civil authorities to take precautionary actions and prevent such human and economic disaster. Short-range forecasts from other national weather services, using different models, techniques and approaches were similar to the ones by AEMET, also exhibiting very low rainfall accumulations over the affected area. Further research is clearly needed to better understand the causes of this kind of forecast misses and identify key aspects to improve its predictability.

The analysis of satellite and conventional observations suggests that an important factor in this event was the westward evolution of a convergence line, which first formed east of Menorca. Along this line, convection with moderate activity was triggered, evolving later into a stationary and vigorous deep system over the affected basin. Difficulties of numerical prediction models to simulate the event are most likely linked to the misrepresentation of this convergence line, which is obviously a mesoscale dynamic feature generated over the sea.

Here we explore the skill of a collection of high horizontal grid resolution (900 m) ensemble generation strategies based on initial conditions perturbations, stochastic parameterizations and also data assimilation (DA) techniques. Since reflectivity observations from a Doppler radar located in southern Mallorca is available, an information source of immense value over data-void maritime regions, we pay special attention to the potential of the DA strategy in generating accurate initial conditions and in consequence, enhancing the short-range prediction skill of this event. Although DA has been widely applied in convective scale phenomena (e.g., Snyder and Zhang, 2003; Dowell et al., 2004; Tong and Xue, 2005; Fujita et al., 2007; Wheatley et al., 2012; Yussouf et al., 2015, among others), it is important to note that in

these studies the weather events took place mainly over flat terrain and over regions densely covered (in time and space) by multiple observation platforms. By contrast, the Western Mediterranean is sparsely observed by in situ measurements because of the predominant maritime surface that defines the region. Also, those studies analysed mostly isolated convective structures, and much less attention has been typically paid to more complex situations such the one presented in this study, with mutual interactions between multiple and *intermittent* convective systems.

In this work, a Monte Carlo approximation to the Kalman filter (Kalman, 1960) data assimilation algorithm, called then ensemble Kalman filter (EnKF; Evensen (1994)) is used. Contrary to alternative methods, such as variational DA schemes (Parrish and Derber, 1992; Courtier et al., 1994), the EnKF explicitly estimates the flow dependent background error covariance in time, providing a more accurate estimation of the observed and unobserved variables through the ensemble correlations (Snyder and Zhang, 2003). The benefits of the EnKF in this type of situations, with poor in situ observations coverage, complex orography and where deep convective systems initiate and develop over the sea, have been barely investigated (e.g., Carrió and Homar, 2016; Carrió et al., 2019). Hence, the application of an EnKF to this flash flood event will contribute to further investigate its potential before these challenging and threatening situations.

The reminder of this paper is organized as follows. Section 2 provides a detailed explanation of the physical mechanisms involved in the evolution of the convective systems that lead to the catastrophic flash flood. This explanation is supported with satellite and conventional observations and by subjective mesoscale analysis. In Section 3, the numerical predictability challenge of the case is illustrated by the assessment of the different ensemble strategies considered. Section 4 introduces the EnKF data assimilation setup and the forecast strategy for the different numerical simulations performed in this study. In Section 5, results are presented and discussed. Conclusions and further final remarks follow in Section 6.

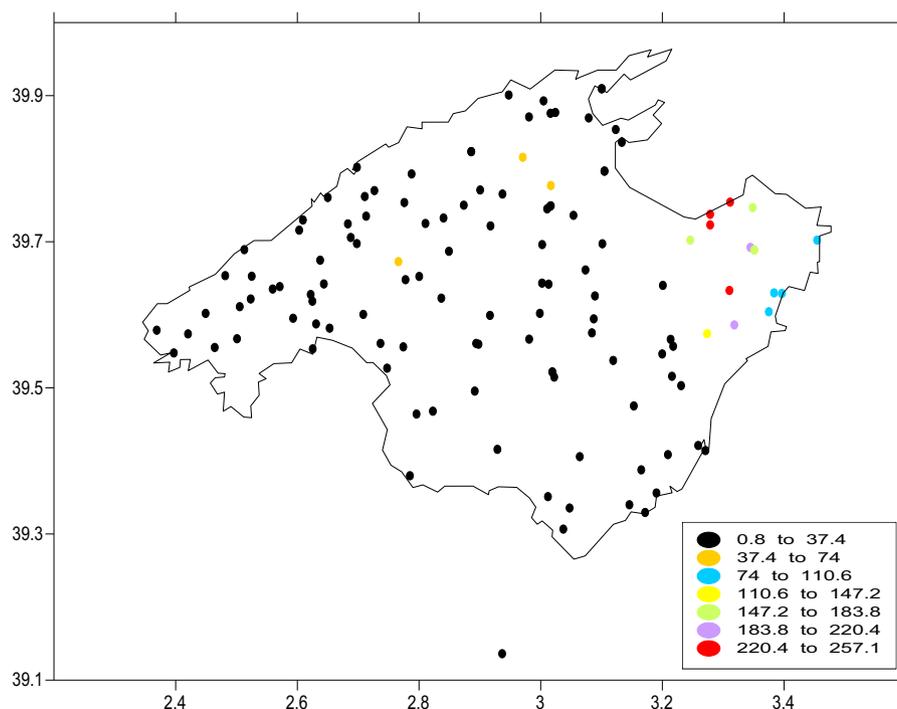


Fig. 1. Accumulated precipitation values (mm) registered in Mallorca by rain gauges in 24 h during the 9 October 2018.

2. Diagnosis of the convergence line

A dynamic feature that affected the initiation and stationarity of the convective system, is the low-levels convergence line. It moved from east to west over the Menorca channel (from Menorca to north-eastern Mallorca), at decreasing speed, until becoming almost stationary when reaching north-east Mallorca.

The SEVIRI-HRV Cloud RGB Meteosat images provide the clearest evidence of the presence, movement, and consequences of the convergence line (Fig. 2). Cloudy features consistent with a convergence line, crossing the island of Menorca around 10 UTC are evident. Alternatively, the 10 min wind data of three wind stations in the island (Airport of Menorca, *Es Mercadal* and *Ciutadella*; see Fig. 3b) also reveal the passage of this convergence line. Although the most common footprint of the passage of a convergence line over a wind station is a veering of the wind direction, it is not the case here. All three stations registered south-easterly winds during most of 9 October, without a clear change in wind direction. Nevertheless, about noon, the passage of a wind speed maximum is observed successively at the three stations (Fig. 4a). Since a maximum of wind speed produces convergence ahead of it (i.e., earlier in time over a meteogram), and divergence behind (i.e., later on), the

presence of convergence associated with the observed wind maximum can be inferred in the zone of maximum increase of wind speed in the meteogram (Fig. 4a). The maxima of the smoothed wind speed series occur at 11:00 UTC at the Airport of Menorca, at 11:20 UTC at *Es Mercadal* and at 11:50 UTC at *Ciutadella* (Fig. 4a). Note that the distance between the Airport and *Ciutadella*, projected zonally, is, approximately, 30 km; this means that the convergence line is moving at about 35 km/h over the island of Menorca. Regarding convergence, the maximum slope in the smoothed wind speed series is found 2–2.5 h before the maximum wind, that is, the convergence line moved about 70 km ahead of the observed maximum surface wind speed. According to this, the line of (maximum) convergence was located around *Ciutadella* at about 09:50/10:05 UTC, and this is consistent with the position inferred from satellite images. Note that, according to satellite images, the convergence line is moving at less than 20 km/h when crossing the Menorca channel, slowing down, as it progressed westward.

The abovementioned findings are reinforced by radar images, which show elongated patches with moderate and high reflectivities consistent with the satellite signals and the conventional data diagnosis (Fig. 16). Thus, we confer high confidence to the presence and diagnosed characteristics of a convergence line moving from Menorca to Mallorca

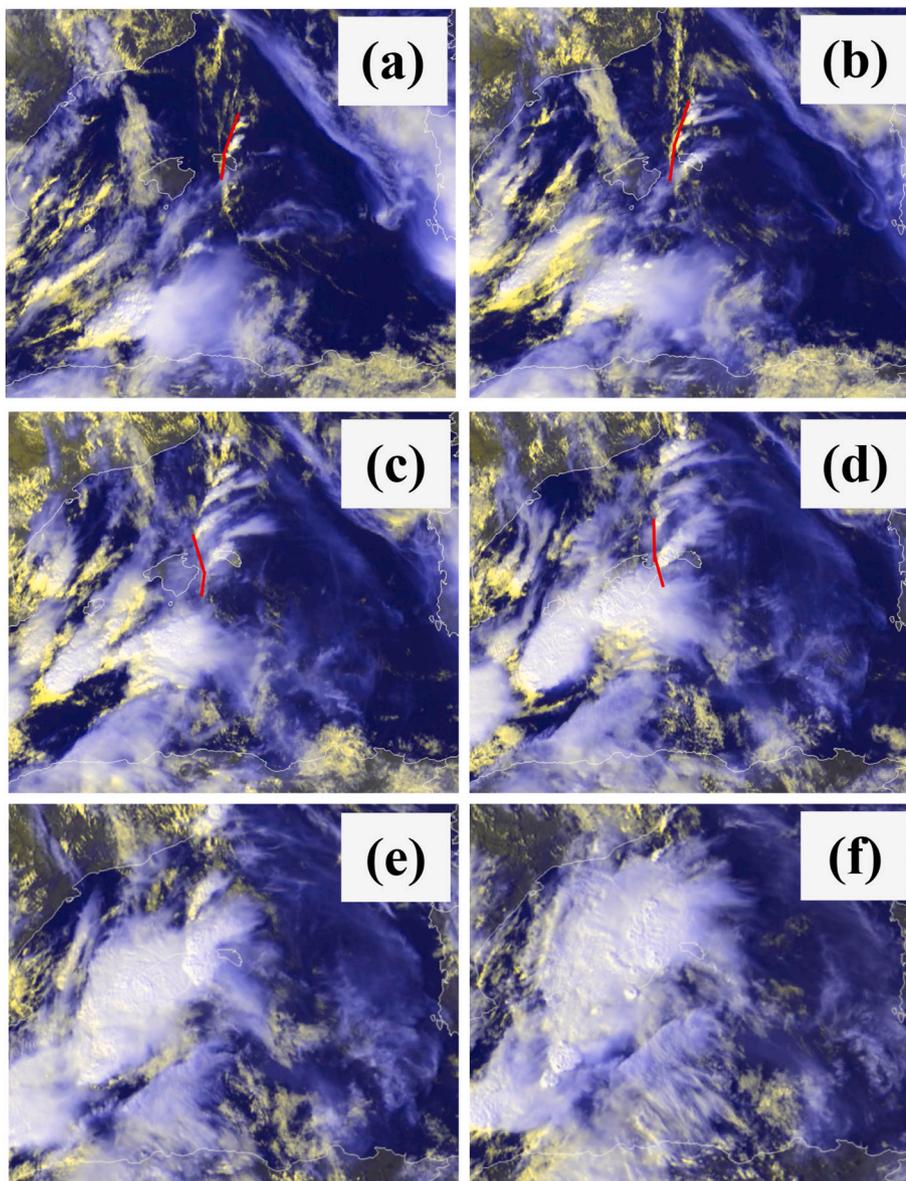


Fig. 2. SEVIRI-HRV Cloud RGB Meteosat images on 9 October 2018 at (a) 09:55 UTC, (b) 10:55 UTC, (c) 11:55 UTC, (d) 12:55 UTC, (e) 13:55 UTC and (f) 14:55 UTC. Red line indicates the convergence line. Yellow tonalities correspond to low-level clouds and blue tonalities are associated to high-level clouds, with pure white indicating convective clouds with deep vertical extent. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

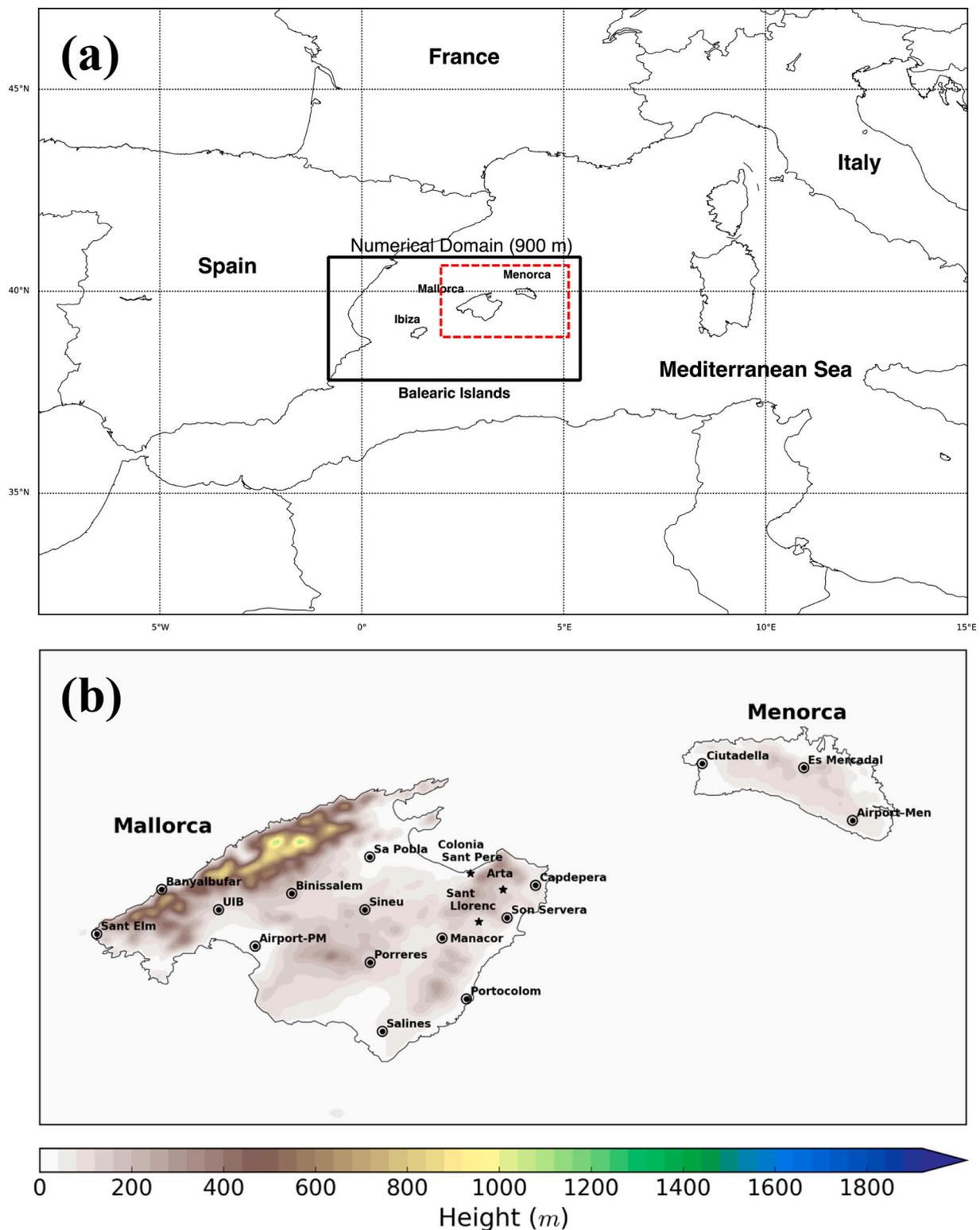


Fig. 3. (a) Numerical WRF domain used to perform all the numerical experiments. Embedded figure in the bottom right panel side shows the three main locations affected by this episode. (b) Zoom in on red rectangle in (a). Islands of Mallorca and Menorca, with the location of the wind stations that are mentioned in the text and in the figures. All these wind stations, except Ciutadella, belong to the Spanish State Meteorological Agency (AEMET). Ciutadella station belongs to Ports de les Illes Balears. The three main locations affected by this episode are also shown here (black stars) for reference. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

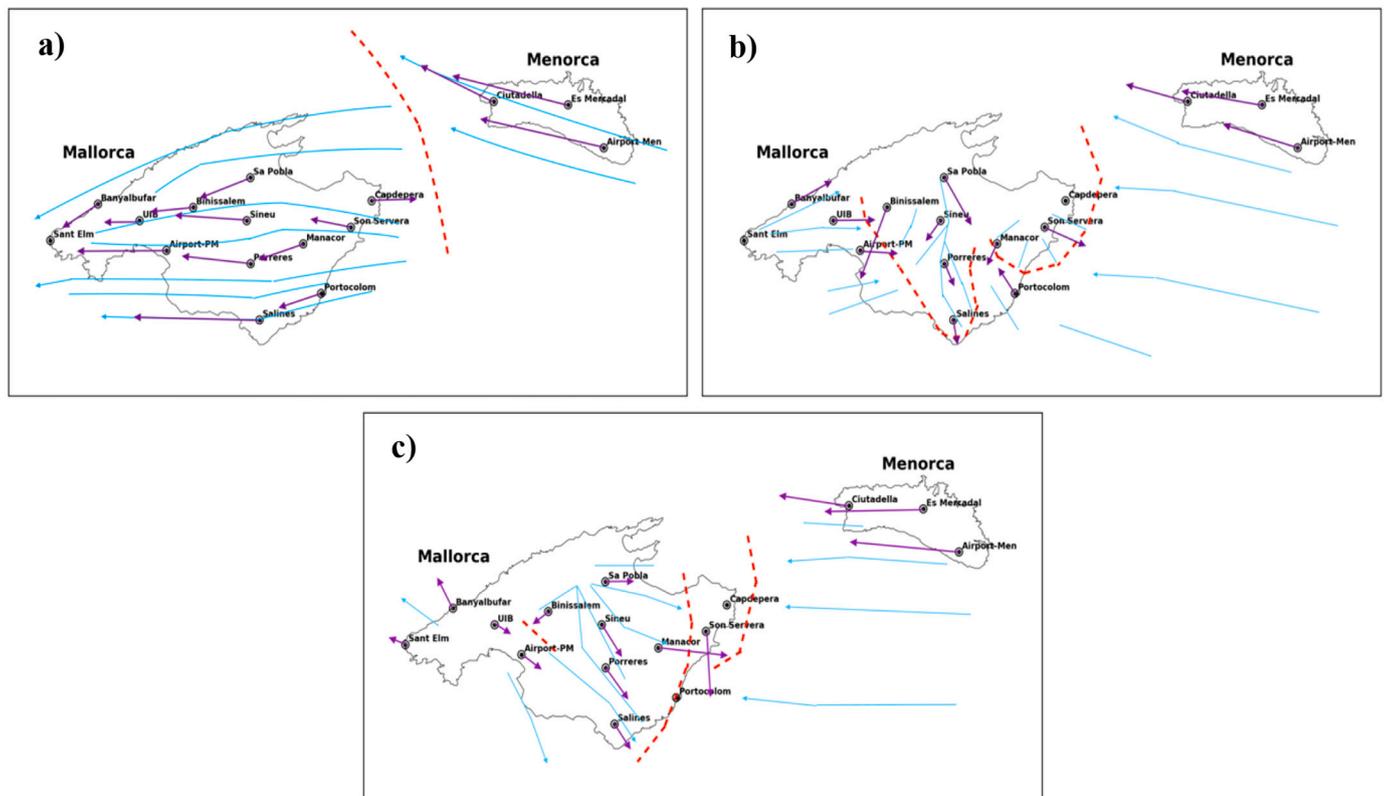


Fig. 6. Handmade wind analyses, on 09 Oct 2018, at 12 UTC (a), at 15 UTC (b) and at 17 UTC (c). Blue lines are inferred streamlines. Red dashed lines indicate possible convergence lines. The convergence line at 12 UTC is drawn as identified through satellite images (Fig. 2c). Wind data come from AEMET and Ports IB stations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

investigate the short predictability of crucial aspects of this episode, we explore the potential of different ensemble strategies operating at sub-kilometric grid resolutions, which allows for some convective scale features of the storm to be resolved. In particular, the entire set of simulations presented in this study will run at 900 m grid resolution. Version 3.9.1.1 of the Advanced Research Weather Research and Forecasting Model (WRF-ARW, Skamarock et al., 2008) is used in this study. A single numerical domain centered over the Balearic Islands with a horizontal grid resolution of 900 m (600×400 grid points) and 35 vertical grid levels from surface to the 50 hPa isobaric level, with higher grid level density at bottom layers than at the upper layers, is adopted for all numerical experiments (Fig. 3a). Initial simulations were performed using 51 vertical levels, however, the extra computation cost (relative to simulations using 35 levels) did not show significant benefits in the performance of the ensembles. For this reason, the simulations use 35 levels to reduce the computational cost. The simulation period extends to 24 h from 00 UTC 9 October 2018 to 00 UTC 10 October 2018. Initial and 3-h updated boundary conditions are generated from the European Center for Medium Range Weather Forecasts Global Ensemble Prediction System (EPS-ECMWF) using an octahedral reduced Gaussian grid (O640) in the horizontal, which translates to an equivalent of 18 km horizontal grid resolution. All the simulations performed in this study were performed using the full EPS-ECMWF ensemble (i.e., 50 ensemble member perturbations). Here, we briefly present the following 4 different ensemble-based strategies used to qualitatively assess the predicting skill of this flash flood event. These 4 EPS strategies are the following: (a) Standard Dynamical Downscaling (SDD), (b) Modified Multiphysics (MMPS), (c) Downscaling + Stochastically Perturbed Physics Tendency (DownSPPT) and (d) Multiphysics + Stochastically Perturbed Parameter (MultiSPP). From the SDD, a 50-member ensemble is generated by a direct downscaling from the EPS-ECMWF and using different WRF physical parameterizations. No cumulus

parameterization is used in any of the simulations performed in this study as most convective activity is expected to be explicitly resolved at 900 m grid resolution. Regarding the MMPS, it uses 10 members out of the 50 available from the EPS-ECMWF. Then, the 10 members exhibiting the largest variability over the WRF numerical domain are selected; see Amengual et al., 2017; Carrió and Homar, 2016; Carrió et al., 2019 for similar applications. The sampling strategy to compose the ensemble is then completed by using additional physical parameterizations to those used in SDD EPS. Using 5 specific parameterization combinations with the 10 initial conditions results in the 50-member ensemble size that allows to easily compare with the other experiments. For the 50-member DownSPPT EPS, uncertainties in the model formulation are introduced by applying the Stochastically Perturbed Physics Tendency technique (SPPT; Buizza et al., 1999; Berner et al., 2015). Namely, total tendencies of temperature, wind components and specific humidity from subgrid processes, except from the microphysics, are multiplicatively perturbed by a random factor. Finally, the MultiSPP EPS aims at sampling both initial condition and model errors. Here, the same 10 ECMWF-EPS members used in the MMPS are also considered. Regarding the multiphysics component, the same schemes used in the DownSPPT experiment are used for this experiment, but now the Stochastically Perturbed Parameter method (SPP; Jankov et al., 2017) is applied to sample model uncertainty at the process level.

Accumulated precipitation among all numerical experiments (Fig. 7) shows extremely low values compared with the corresponding observed values over the eastern part of Mallorca, where the flash flood took place (Fig. 1). This result indicates the lack of skill of these configurations in capturing the essential factors of the observed torrential rainfall and confirms the predictability challenge that this episode entails. These results confirm that the predictability of this event is not sensitive to the numerical model used or the parameterizations chosen. The poor skill of the models with this event is more likely related with a

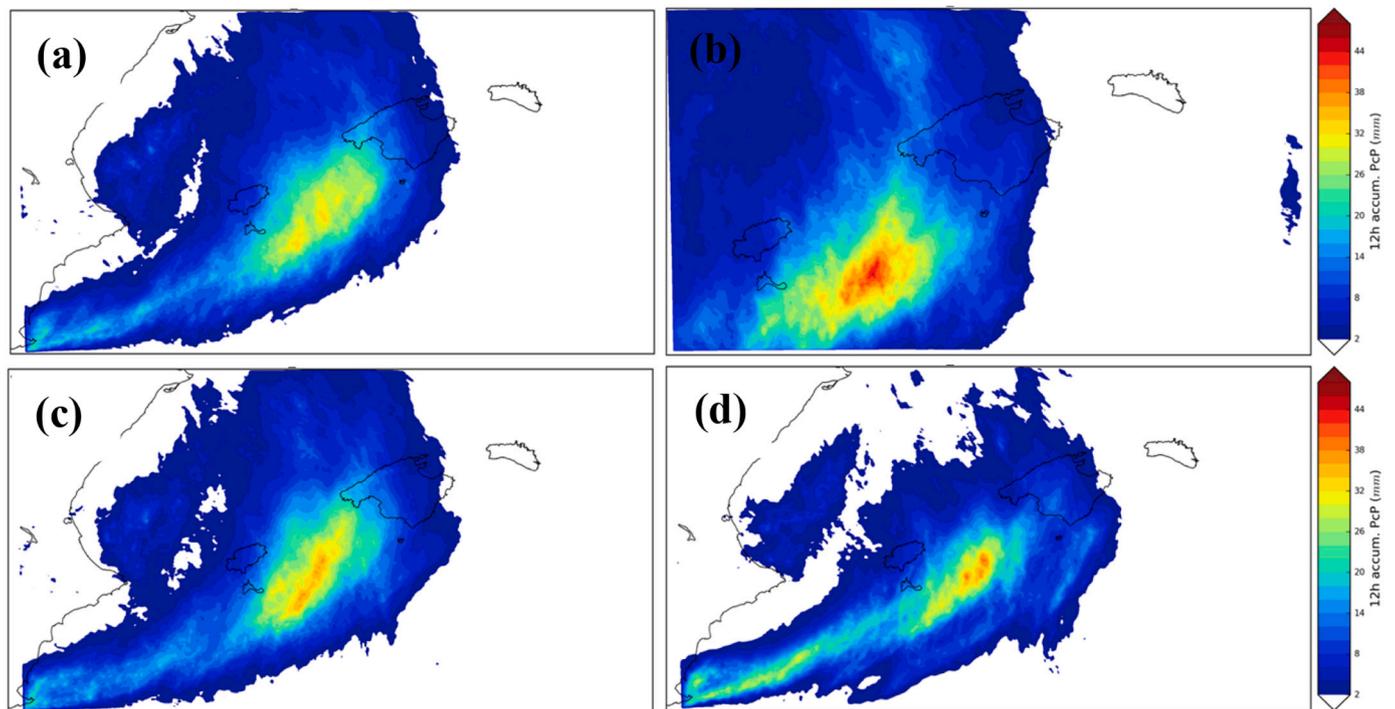


Fig. 7. 12 h Accumulated precipitation (ensemble mean) computed from 06 UTC to 18 UTC 9 October by (a) SDD, (b) MMPS, (c) DownSPPT and (d) MultiSPPT ensemble strategies.

misrepresentation of the initial state of the atmosphere, which is well known as the primary error source in numerical weather prediction, especially when predicting deep convective systems (Wu et al., 2013).

Typically, poor representation of the state of the atmosphere is associated with data-void regions, such as marine bodies. In the case of the Balearic Islands, the initial conditions could suffer from misrepresentation due to the sparsity of in situ observations over the Mediterranean Sea. The only way to improve such estimation of the atmospheric state is to allow the model to be influenced by fresh observations (Leith, 1993). In this context, Data Assimilation (DA; Daley, 1993; Kalnay, 2003; Rabier, 2005) is the statistical procedure of combining observations and the numerical model accounting for the inaccuracies of both, fulfilling certain constraints to finally obtain an optimal estimate (analysis) of the true state of the atmosphere. More generally DA determines the evolving probability density function associated to analysis errors. With the main purpose of improving the initial conditions and, hopefully, the predictability of the *Sant Llorenç* flash flood event, an ensemble-based DA technique is used. Note that the above discussed experiments shown in Fig. 7 were designed to determine the skill of different ensemble strategies which did not use DA.

4. EnKF data assimilation setup and forecast strategy

During the last decades, multiple DA techniques have emerged with the aim of reducing uncertainties in the definition of the initial atmospheric state. Most of these techniques are based on probability theory, which fundamentally apply Bayes' Theorem (Lorenç, 1986) and they differ from each other in the assumptions and approximations used. Currently, some of the most effective DA techniques used in leading operational weather centers are classified in two main categories: Variational and Ensemble DA. On the one hand, variational DA schemes are based on the minimization of a cost function using linear tangent and adjoint models, such as 3D-Var (Parrish and Derber, 1992) and 4D-Var (Courtier et al., 1994). One of the most determinant assumptions in variational DA schemes is the use of static and nearly isotropic and homogenous forecast error covariances, which is far from reality, as

these errors and their covariances may vary substantially with the flow of the day. On the other hand, ensemble-based DA are typically based on Kalman filtering theory (Kalman, 1960; Kalman and Bucy, 1961) and provides the best estimate of the state of the atmosphere, in a least squares sense, by minimizing the analysis error variance estimate. One of the most appealing characteristics of ensemble DA schemes is their ability to provide flow-dependent estimate of the forecast-error covariances. In addition, unlike variational schemes, ensemble methods do not make use of the linear tangent and adjoint models, which are computationally expensive to run and maintain.

In this study we use the parallel version of the ensemble adjustment Kalman Filter scheme (EAKF; Anderson, 2001), from the Manhattan release branch of the Data Assimilation Research Testbed software system (DART; Anderson and Collins, 2007; Anderson et al., 2009). Like in the abovementioned experiments, the EnKF will be run at 900 m horizontal grid resolution using the 50 members from the EPS-ECMWF.

4.1. Assimilated observations

The flash flood event that took place over the region of *Sant Llorenç* was the result of a combination of multi-scale dynamical and thermodynamical processes in both the synoptic and the *meso* scales. To account for successful information in the numerical model at this wide range of scales, two main types of observations were assimilated.

The larger-scale information was mainly contributed by conventional observations obtained from the NOAA's Meteorological Assimilation Data Ingest System (MADIS). This database provides quality-controlled observations distributed globally, which is an important requirement to perform useful DA. MADIS observations assimilated here correspond to temperature, humidity, wind direction, wind speed and pressure variables. These variables are registered by different observation platforms, with specific deployment characteristic for each case (Fig. 8). In this study, radiosondes (vertical profiles), maritime buoys, METARs and aircraft observations are ingested.

The *Sant Llorenç* flash flood was strongly influenced by the interaction of multiple convective storms, originated over the sea, where a very

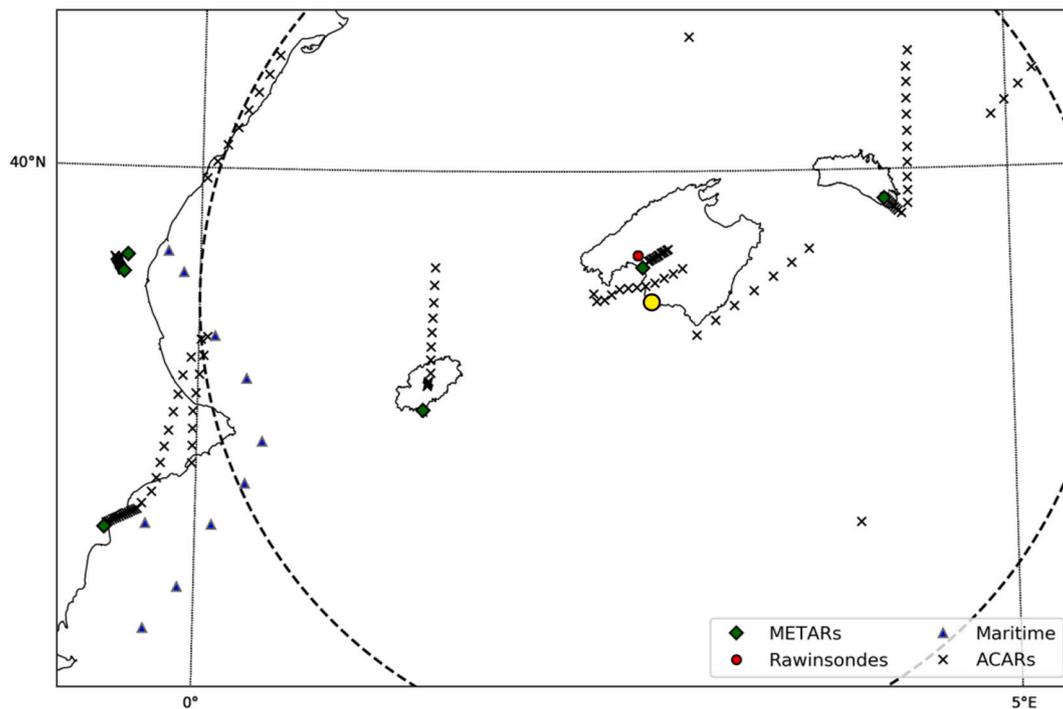


Fig. 8. Spatial distribution of conventional (METARs, maritime buoys, rawinsondes and ACARs) data assimilated by the EnKF scheme between 06 UTC and 18 UTC 9 October 2018. Dotted black circle line denotes the range of the Doppler radar located in the southwestern part of Mallorca (at yellow dot) position. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

limited number of in situ conventional observations are systematically available. To ingest information about the atmospheric state at such small scales, reflectivity observations from a Doppler C-band weather radar located at the southwestern part of Mallorca, specifically at (lat, lon, alt) = (39.38 N, 2.78E, 117 m), were assimilated (Fig. 8). These data contain volumetric radar observations distributed in 8 scan elevations with spatial and temporal resolutions of 1 km and 10 min, respectively. However, the raw data is not routinely quality controlled and radar ground cluttering and false echoes signals must be removed by an ad hoc process before being assimilated. To remove spurious signals associated with all volumes and taking in major consideration the lower elevations fields, which are the most affected by non-meteorological echoes, the following 3 step process based on Lakshmanan and Zhang, 2009, Altube et al., 2015 and Saltikoff et al., 2019 is applied:

- (a) Remove echoes lower than 0 dBZ threshold: these signals are all non-rainfall and despite they can provide information about specific meteorological features, they are not relevant for this study.
- (b) Remove reflectivities below 20 dBZ to improve the data assimilation of the more intense convective zones and.
- (c) Manually identify and clean the rest of non-meteorological echoes.

Finally, due to the high resolution of the data and to avoid signal aliasing and error correlations, which would violate fundamental assumptions in the DA scheme, radar observations are objectively analysed onto a 6 km horizontal regular grid using the Cressman interpolation method (Cressman, 1959; Wheatley et al., 2015; Yussouf et al., 2015). It is noteworthy to mention that initially and considering that the simulations are running using a 900 m grid space resolution, observations were objectively analysed onto 1, 2 and 3 km and then used for additional EnKF experiments. However, they do not result in significant differences in terms of improving the prediction of the flash flood event. For this reason and to reduce computation time, we finally decided to use a 6 km Cressman radius length.

Another valuable source of information from Doppler radars are radial velocities. Unfortunately, these observations were not available or missing for the period this flash flood event last. For this reason, radial velocities could not be assimilated in this study.

4.2. Storm-scale EnKF data assimilation

These two different types of observations (i.e., conventional and radar) are assimilated using two main temporal frequencies: (i) 1-h DA cycles for conventional observations and (ii) 10-min DA cycles for radar observations. For each DA cycle, the EnKF updates,¹ via the assimilation of the abovementioned observations, the following components of the state vector: potential temperature perturbation, geopotential perturbation, perturbation of surface pressure of dry air, horizontal components of the wind, water vapour, cloud, ice, snow, rain and graupel mixing ratios and finally the number concentration of rain and ice species. In addition to these prognostic variables, in this study the diagnostic fields associated with 10-m wind, 2-m temperature, 2-m humidity, surface pressure and 10-cm reflectivity are also updated.

A widely known issue that ensemble filters need to deal with, is related with the *rank deficiency problem*. This occurs because the number of ensemble members, $N \sim 10\text{--}10^2$, is much smaller than the size of the full model state vector $n \sim 10^7\text{--}10^9$. In this situation, where sample covariances are degenerated, multiple directions in the phase space in which the ensemble has no variance exist and thus, the error covariances are not representative of the true error covariances, thus producing poor analysis (Hacker et al., 2007). More specifically, *rank deficiency* introduces two main problems which are known as *filter divergence* (Fitzgerald, 1971) and long-range *spurious correlations* (Hamill et al., 2001).

With the main objective of reducing the negative impact of long-range spurious correlation due to the use of a moderate ensemble size,

¹ In the DA context, the concept “updates” simply refers to the process of combining the previous information from the model and the newly available observations (assimilation).

a covariance localization technique which is based on a distance weighting function that goes to zero at distant regions is used (Houtekamer and Mitchell, 1998; Sobash and Stensrud, 2013). In this study a Gaussian-based localization taper function, with value 1 at 0 distance and value 0 at 2 cutoff half-width distance is used (Gaspari and Cohn, 1999). Here, for conventional observations a cutoff half-width of 510 km and 6.5 km for the horizontal and vertical localizations are used, respectively. In contrast, for radar-derived observations, the covariance localization is set to have a cutoff half-width in the horizontal (vertical) of 9 (3) km (Yussouf et al., 2015).

Another important aspect to consider performing this kind of experiments is the problem of the *filter divergence*, which is associated with the underestimation of the forecast error covariance matrix due to reduction in the ensemble spread after each analysis cycle. This underestimation results with the EnKF ignoring the information from the observations and thus, the analysis solution ends diverging from the true analysis. To deal with this problem, the forecast error covariance matrix is inflated by multiplying the spread of the ensemble by a specific factor, without modifying the ensemble mean. In this study an *adaptive inflation method* based on adding random perturbation from a normal distribution is used for each data assimilation cycle (Anderson and Collins, 2007; Anderson et al., 2009). Here, this normal distribution is defined with a mean of 1.0 and standard deviation of 0.6.

An additional problem that DA filters arise comes when spatially distributed high-density observations are assimilated, causing ensemble underdispersion. In such situations additional spread from a random normal distribution is added. More specifically, this spread is added to horizontal wind, temperature and to the water vapour fields at locations where the observed radar reflectivity exceeds a threshold (25 dBz in this case), using the *additive noise technique* (Dowell and Wicker, 2009). This *additive noise* is applied between the assimilation and the forecast steps of the cycle. The mean of such perturbations is 0 for all the observations considered and the standard deviations are 0.5 ms⁻¹ for horizontal winds and 0.5 K for temperature and dewpoint (Dowell et al., 2011; Yussouf et al., 2013).

Also note that to properly assimilate reflectivity observations from the Doppler radar, here we use the following nonlinear observation operator:

$$Z_e \approx \sum_i c_i \int_0^{\infty} n_i(D) D^6 dD$$

where c_i is the radar calibration coefficient, $n_i(D)$ is the size distribution, and D is the hydrometeor diameter. Note that the size distribution of the i^{th} hydrometeor class is approximated by:

$$n_i(D) = n_{0i} \exp(-\lambda_i D)$$

where n_{0i} is the intercept parameter and λ is the slope parameter of the size distribution, which are related to the mixing ratio of the species q_i following:

$$\lambda_i = \left(\frac{\pi \rho_i n_{0i}}{\rho q_i} \right)^{0.25}$$

with ρ_i being the density of the species and ρ the air density.

An additional quality control method is performed when the difference between the observation and the prior ensemble mean exceeds 3 times the square root of the sum of the prior ensemble variance and the observation error variance. In such situations, the observations are blacklisted (ignored). In this study, observational errors (standard deviations) associated with conventional observations are analogous to Table 3 in Romine et al. (2013) with minor exceptions: METAR altimeter (1.5 hPa), marine altimeter (1.20 hPa) and METAR and marine temperature (1.75 K). Additionally, following similar DA studies (e.g., Yussouf et al., 2015), observation error for radar reflectivity is set to 5

dBz.

4.3. Numerical experiments design

To quantitatively assess the impact of assimilating both in situ conventional and Doppler radar observations on the short-range forecast of the *Sant Llorenç* flash flood, 8 numerical experiments are designed: a control run and 7 sensitivity experiments (Fig. 9).

The control run is designed following two main stages: spin-up and data assimilation. During the first stage, the 50 members initialized with the EPS-ECMWF at 00 UTC 9 October *freely* run for 6 h, allowing for the spin-up of the numerically resolvable scales, up to the nominal horizontal grid resolution of 900 m. Note that the term *free* is used to stress that the numerical model is running without being adjusted by the DA process. After this initial scale-adjustment period, the data assimilation process starts covering a temporal data assimilation window (DAW) of 12 h, from 06 UTC to 18 UTC 9 October. The control run is intended to provide the best estimate of the atmospheric state (i.e., analysis) at the resolved scales every 10 min over the DAW. This analysis is used to initialize the experiments indicated in Fig. 9.

The 7 sensitivity experiments are aimed at assessing the benefits on the prediction of the flash flood from the assimilation of the above observations using different DAWs. This design will allow us to investigate the limit of numerical predictability of the episode and the benefits from assimilating various types of observations. The sensitivity experiments are labelled as EXP_XXDA, where XX refers to the initialization hour of the ensemble forecast, which will run from this time to the end of the day, at 00 UTC 10 October. For example, the EXP_12DA ensemble initiates the free forecast at 12 UTC, using the initial conditions rendered by the corresponding cycle of data assimilation (in this example, after 6 h of continuous data assimilation), and run freely until the end of the day. EXP_13DA is analogous to the previous case, but now the ensemble forecast starts with the analysis obtained at 13 UTC and runs until the end of the day again. A total of 7 experiments are performed starting every hour from 12 UTC to 18 UTC 9 October (Fig. 9). Considering that the maximum hourly accumulated precipitation of this episode was recorded between 17 UTC – 18 UTC, EXP_18DA will simulate the event only partially. This experiment is considered to analyse the potential forecast improvement of the last phase of the flash flood event.

5. Results

5.1. Observation-space diagnostics

To quantitatively assess the performance of the Hi-EnKF system during the 12-h data assimilation window, widely used observation-space diagnostics (Yussouf et al., 2013, 2015; Wheatley et al., 2015; Carrió et al., 2019) are applied. In particular, the following main diagnostics are evaluated: (i) the root-mean squared innovation (*rmsi*), (ii) the *bias* (model-observation), (iii) the total ensemble spread (*TS*; Dowell and Wicker, 2009) and finally (iv) the consistency ratio (*CR*; Dowell et al., 2004) from the background.

The *rmsi* provides a simple quantitative measure of how well the mean of the analysis and forecast ensemble fields fit the observations. This diagnostic is computed before and after each data assimilation cycle to assess the impact of the observations in the respective data assimilation cycles. In the case of the assimilation of reflectivity observations, the *rmsi* was computed every 10 min from 06 UTC to 18 UTC on 9 October 2018. Initially, the *rmsi* for the reflectivity shows high values of 20–22 dBz, but after subsequent data assimilation cycles this error is substantially reduced and stabilized around 5–6 dBz (Fig. 10). The reflectivity bias shows very large values during the initial *spin-up* of the storm-scale ensemble, from 0600 to 0800 UTC, and then decreases significantly with time and remains stable around (–1)–(–5) dBz. It is evident from this diagnostic that the model underpredicts observed reflectivity values throughout the DAW.

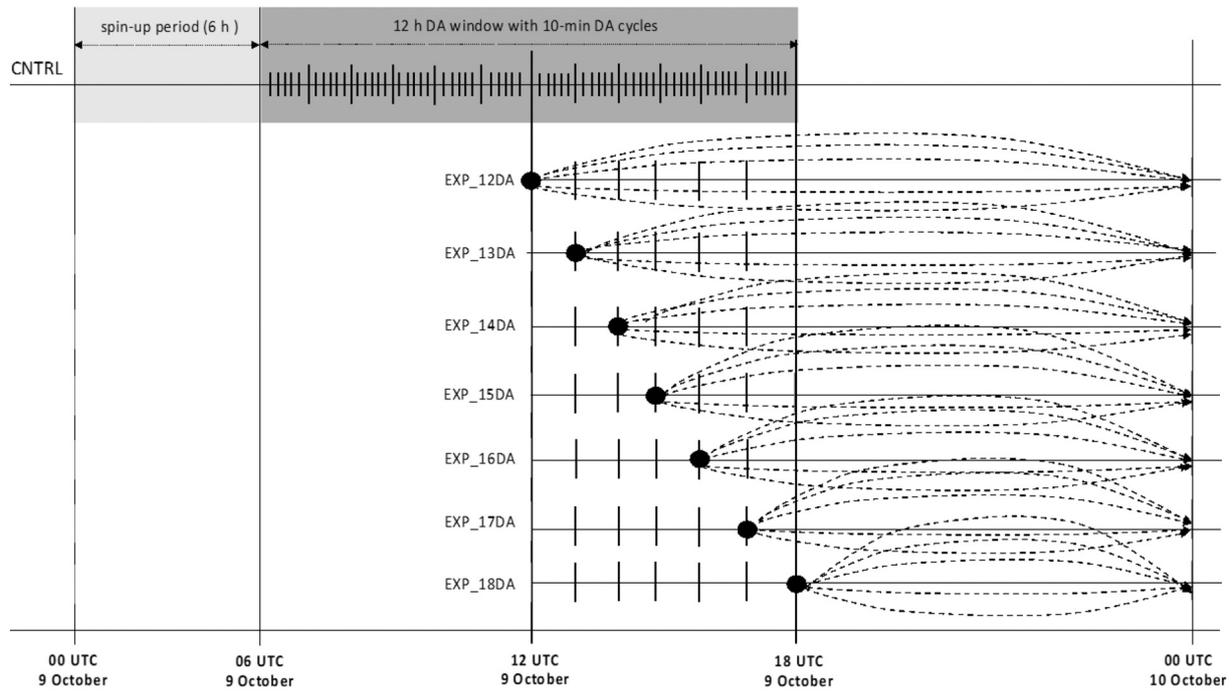


Fig. 9. Configuration scheme used by the different DA experiments performed in this study. Light shaded region represents the first 6 h spin-up period. Dark grey region indicates the 12 h DA window, in which the observations (conventional and radar based) are assimilated. Black dots show the initiation time of the different ensemble forecast simulations initiated with the analysis obtained from the EnKF at that time. Dotted arrows represent the evolution of each ensemble member along the forecast period.

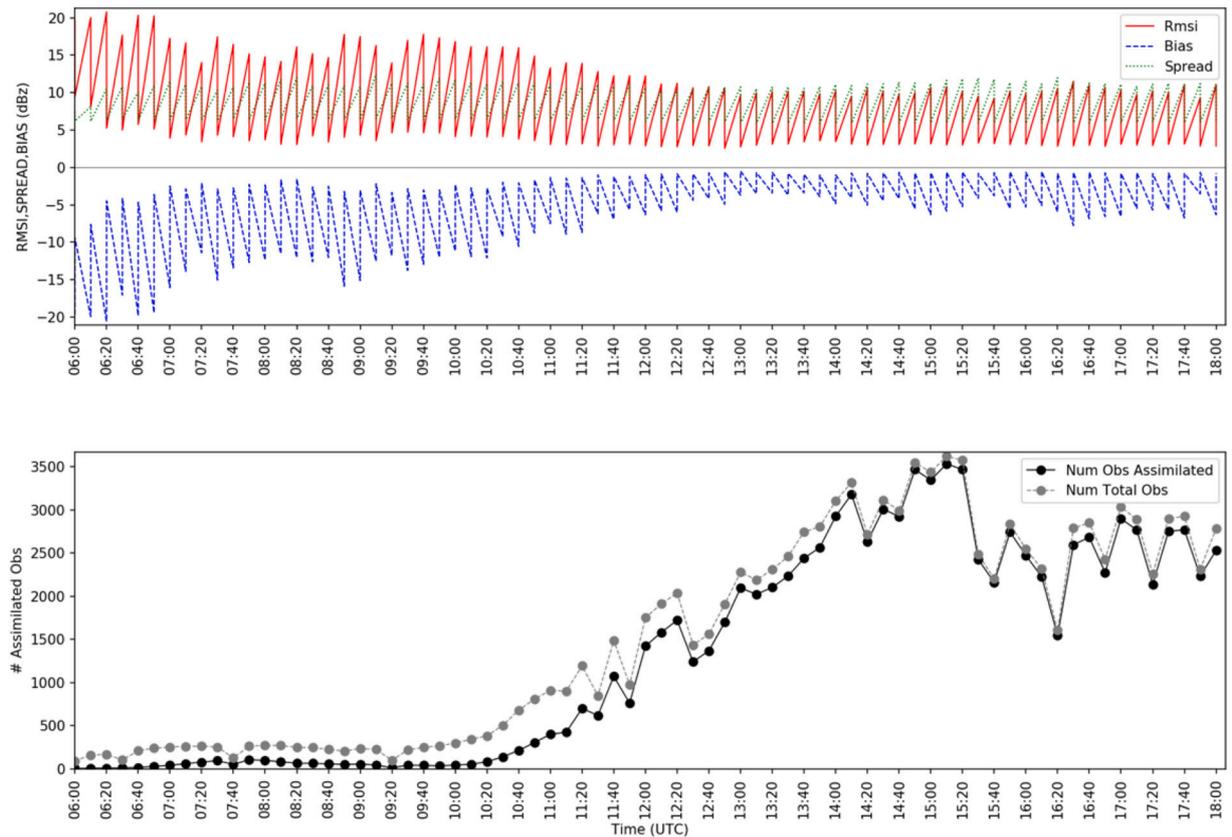


Fig. 10. Observation-space diagnostic statistics for assimilated reflectivity observations from the radar located in Palma during the 12-h-every-10-min storm-scale data assimilation period (upper panel). Red line indicates the rmsi, blue line indicates the bias and the green line the total ensemble Deleted: black standard deviation. Lower panel shows the number of the reflectivity observations assimilated (black line) compared with the total number of reflectivity observations (grey line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The *consistency ratio*, which checks the desirable property that the ensemble spread is representative of the *unknown* true forecast errors (e. g., Dowell et al., 2004; Aksoy et al., 2009; Dowell and Wicker, 2009; Yussouf et al., 2015). This diagnostic tool is based on the main assumption that the observation and forecast errors are independent, so the variance of the *innovation* σ_d^2 , where the *innovation* \mathbf{d} is defined as $\mathbf{d} = \mathbf{y}^o - \overline{H(\mathbf{x}^f)}$, can be easily decomposed as the sum of the observation and forecast error variances: $\sigma_d^2 = \sigma_o^2 + \sigma_f^2$. Then, assuming that the ensemble spread is representative of the forecast error we can replace the forecast error variance by the ensemble error variance: $\sigma_f^2 \rightarrow \sigma_{ens}^2$. Note that both forecast, and ensemble error variances must be in the observation space.

Hence, the consistency ratio checks the consistency of the ensemble spread with the *rmsi* for the assumed amount of observation error as follows:

$$\begin{aligned} \text{consistency ratio} &= \frac{\sigma_o^2 + \sigma_f^2}{\sigma_d^2} = \frac{\sigma_o^2 + \sigma_{ens}^2}{\sigma_d^2} \\ &= \frac{\sigma_o^2 + \left\langle \frac{1}{K-1} \sum_{k=1}^K [H(\mathbf{x}_k^f) - \overline{H(\mathbf{x}^f)}]^2 \right\rangle}{\langle (d - \langle d \rangle)^2 \rangle} = \frac{(TS)^2}{(rmsi)^2} \end{aligned}$$

A consistency ratio ~ 1 shows that the ensemble spread is representative of the unknown true forecast variance for the assumed observation error. In this study, consistency ratio values are small during the first hours, with values near 0.25, indicating lack of ensemble spread (Fig. 11). However, the CR progressively increase until reaching tolerable values around 1 at the end of the DAW. Analogous results are obtained from the diagnostics of other conventional observations assimilated, such as 2 m temperature and dewpoint, surface pressure and 10 m winds.

The *biases* of both the *prior* (before assimilation) and the *posterior* (after assimilation), also called analysis enable to estimate, on average, the benefits of the assimilation through the three-dimensional atmospheric space associated with the numerical domain considered. In general, results show that the posterior *bias* is closer to zero than the *prior* for all observed variables. For the sake of brevity, only temperature, x-component of the wind and dew point temperature at 12:00 UTC 9 October 2020 are shown (Fig. 12). These results clearly reveal the significant impact of the data assimilation process across the entire troposphere. For temperature, the *posterior bias* shows values closer to zero than *prior* values for almost the entire vertical (Fig. 12a). For U wind component, posterior bias values show better scores than the prior values for the entire vertical (Fig. 12b). Regarding dew points, the assimilation also reduces the bias across the vertical (Fig. 12c). Note the remarkable improvement the assimilation scheme is achieving near 400 hPa, reducing the bias from around -40 K to nearly zero.

In general, the abovementioned observation-space diagnostics

suggest that the configuration parameters of the ensemble data assimilation system are sensible and stable. In this favourable situation one is encouraged to proceed with the following steps: (a) inspection of the analysis fields and the increments (posterior-prior) produced by the data assimilation scheme and (b) initialization of short-range forecasts from the analysis fields obtained from the data assimilation scheme.

5.2. Hi-EnKF analysis (model-space diagnostics)

In this section we focus on a quantitatively assessment on the impact of the data assimilation by means of performing diagnostics exclusively using the analysis from the EnKF obtained along the entire DAW (from 06 UTC to 18 UTC 9 October). The effects of assimilating different types of observations can be determined by comparing forecasts fields in model-space. First, we compute the 12 h accumulated precipitation by adding up the partial accumulated precipitation fields obtained in the forecast steps (1 h or 10 min long depending on the experiment) of the assimilation cycle from 06:00 to 18:00 UTC. Results show the great skill of the EnKF in detecting the location and intensity of the maximum accumulated precipitation registered during the flash flood event when reflectivity values are assimilated (Fig. 13a). As far as the authors know, these are the first numerical results in the literature that simulate with high degree of realism the extreme and torrential accumulations registered in this episode.

An additional numerical experiment only assimilating conventional observations (SYN; without reflectivity radar observations; Fig. 13b) reveals that the assimilation of the reflectivity is very effective for this low-predictable event. The spatial and temporal density of such observations play a key role in correctly adjusting the atmospheric state over the radar's area of influence. In particular, the off-diagonal elements of the multivariate covariance matrix are used to update the unobserved variables, such as the wind (both speed and direction), temperature, humidity, or heat fluxes, among others, which proves remarkably beneficial. The increments of various unobserved variables, such as wind speed and direction, temperature and mixing ratios at low levels reveal the effective adjustment of unobserved model variables by means of the assimilation of reflectivity observations (Fig. 14).

In general, large increments rise in regions with significant reflectivity values (Fig. 14a). This indicates the remarkable effect of assimilating reflectivity observations in getting better estimates of the atmospheric state, even in situations dominated by rapid evolution of deep convective structures. Although this first result is unequivocal, a more detailed investigation is needed to better understand the effects of assimilating reflectivity observations by means of an EnKF. In particular, the EnKF skill in reproducing the key elements in the initiation and temporal evolution of the observed convective cells associated with the flash flood is assessed. A simple way of achieving this is to compare the EnKF hourly accumulated precipitation (ensemble mean) with the

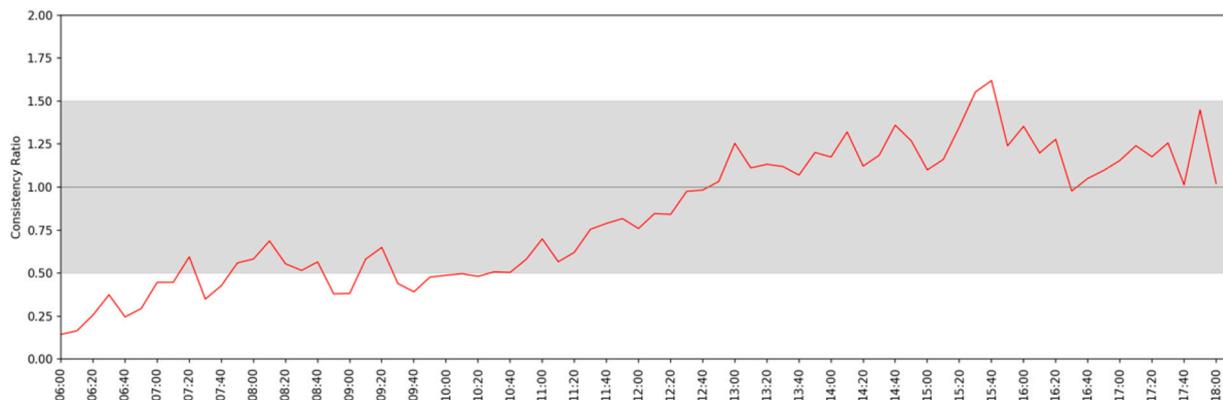


Fig. 11. Consistency ratio (computed from the priors) during the 12-h storm-scale DA period.

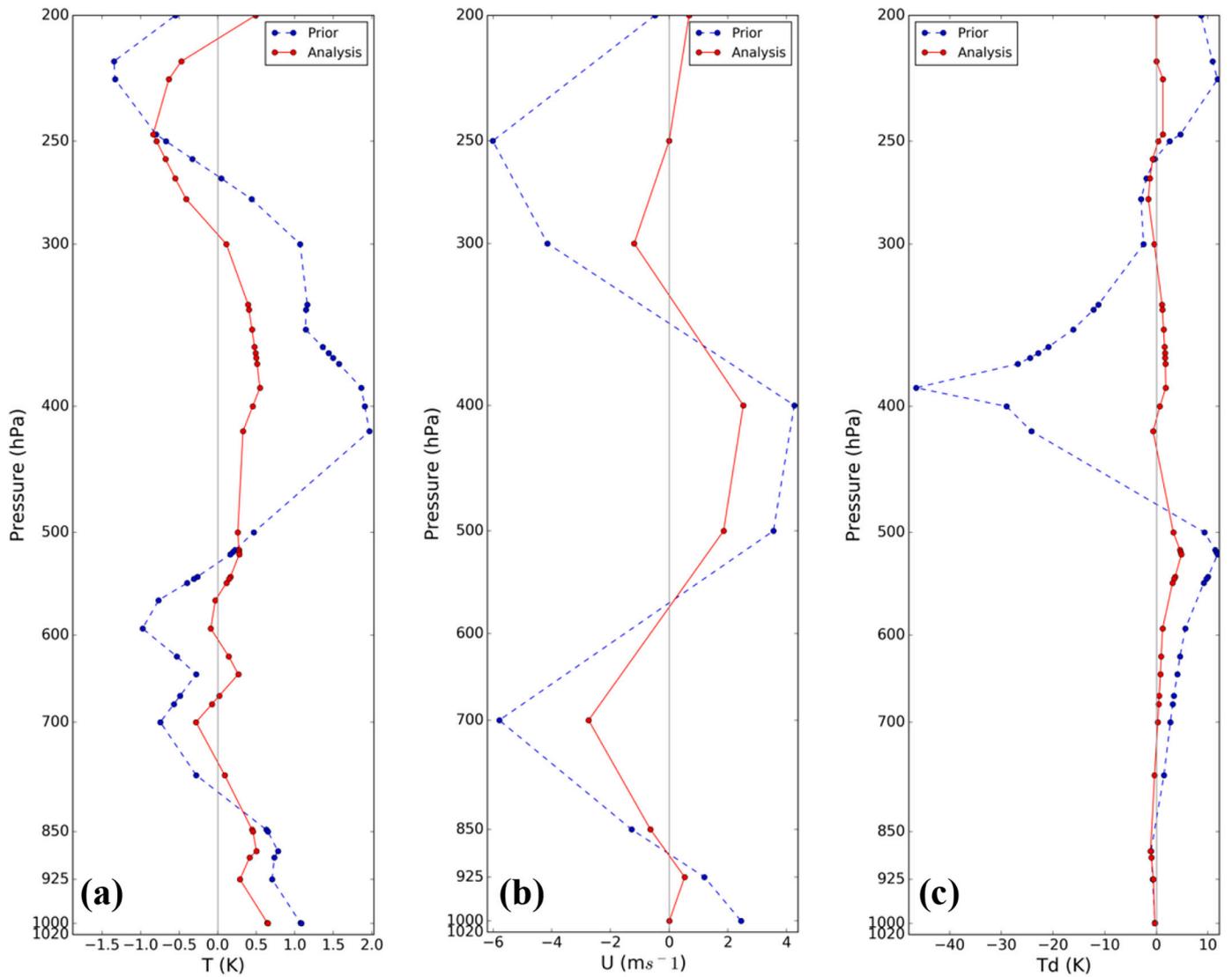


Fig. 12. Vertical bias profile at 12:00 UTC 9 October 2020. Bias before assimilation (prior; blue dashed lines) and after assimilation (analysis or posterior; red dashed lines) are plotted for the variables a) temperature, b) x-component of wind and c) dew point temperature. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

observed reflectivity. After 4 h of assimilation cycle, the EnKF analysis displays the main observed convective structures, showing high amounts of hourly accumulated precipitation at the locations with maximum reflectivity values (Fig. 15). However, during the first hours of the DAW, the EnKF analysis does not successfully capture the initiation and evolution of convection along the convergence line, east of Menorca, diagnosed in Section 2 (indicated in Fig. 15a). The cause for this miss is the rejection by the quality control process -discussed in section 4.2- of the reflectivities associated with the convergence line due to the differences with respect to the prior fields. Indeed, none of the 50 prior members show any reflectivity signal on the observed convergence line. In this situation, it follows that $\mathbf{P}^f \mathbf{H}^T = \mathbf{0}$, which results in a null correction term in the equation for the posterior mean $\bar{\mathbf{x}}^a$ for the EnKF which takes the form:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} (\mathbf{y} - \overline{\mathbf{H}(\bar{\mathbf{x}}^f)}) \quad (1)$$

where $\bar{\mathbf{x}}^f$ is the mean of the prior distribution of possible states, \mathbf{P}^f is the forecast error covariance matrix, \mathbf{R} is the observational error covariance matrix, \mathbf{H} is the Jacobian of the observation operator and \mathbf{y} lists the imperfect observations available. Note that although Eq. 1 provides the

linear form of the EnKF equation, the above statement also holds for the most general (nonlinear) version of the EnKF equation. Thus, in the situation where $\mathbf{P}^f \mathbf{H}^T = \mathbf{0}$ the analysis remains significantly different from observations, with no adequate representation of the convergence line diagnosed in Section 2, which is reflected on the divergence field (Fig. 16).

The causes that could explain the lack of any signal of this convergence line for the unfold of this episode in the EnKF analysis are: (i) the global EPS-ECMWF miss the dynamical structures that contributed to the formation of the convergence line, (ii) large-scale information advected from the boundary conditions could hinder the initiation and development of this convective system in the assimilation cycle, and (iii) lack of accurate representation of subgrid scale processes in the forecast model.

On the contrary, radar-derived radial velocities do not share the above-described weakness of assimilating semi-positive definite and bounded variables, such as the reflectivity observations (Bishop, 2016). Such information could help to identify the observed low-level convergence line and thus it could help to initiate and develop the linear convective system discussed above. Unfortunately, for the *Sant Llorenç* case, radial velocity data corresponding to the whole day was

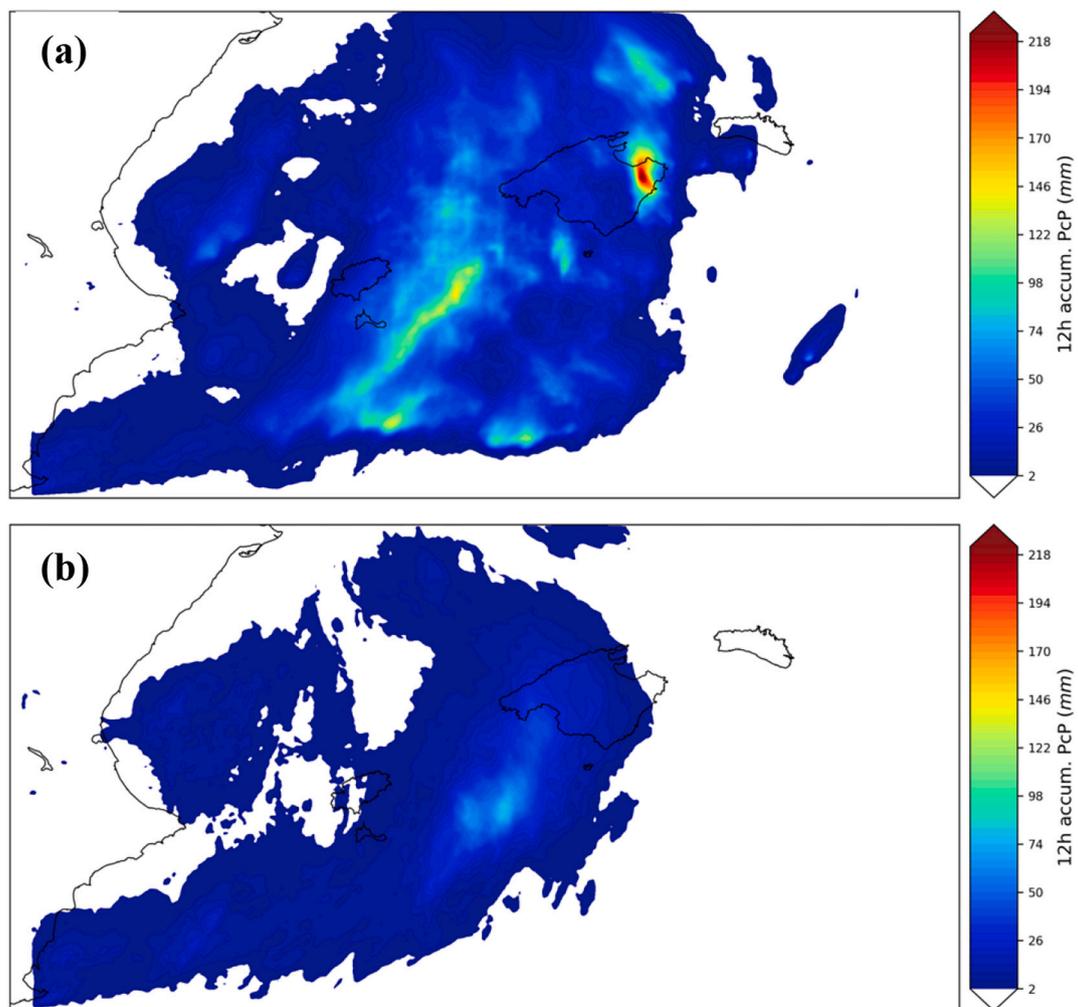


Fig. 13. 12-h accumulated precipitation (ensemble mean) for 06:00–18:00 UTC 9 October corresponding to (a) conventional + reflectivity radar and (b) conventional data assimilation using EnKF.

incomplete or missing, thus no DA of radial velocities was possible.

5.3. Predictability limits using distinct initial conditions from the Hi-EnKF

In the previous section we have shown the great skill of the Hi-EnKF system in generating 12-h accumulated precipitation fields. Note that the above results only refer to the analysis obtained by the DA scheme obtained every 10 min. Each analysis produced throughout the DAW can effectively be used to initialize a free forecast. In this study we are interested to know how much time ahead the flash flood episode could be predicted using the EnKF analysis. Slight improvements on the forecasts lead time play a crucial role for any effective civil protection warning system. Therefore, we investigate the predictability of this event using the EnKF analyses over multiple lead times before 18 UTC, time of maximum 1 h accumulated precipitation registered in *Sant Llorenç* (~18 UTC).

Seven independent ensemble forecasts were launched at 12, 13, 14, 15, 16, 17 and 18 UTC of 9 October, running until 00 UTC 10 October following the schematic diagram of Fig. 9. For each of these experiments, the probability of accumulated precipitation values greater than 20 mm, 50 mm and 100 mm over the entire period of simulation is computed.

Results show that for the 20 mm threshold, there is a significant difference between EXP_12DA and the other experiments. In particular, EXP_12DA does not show relevant signals of accumulated precipitation over the eastern part of Mallorca, and more specifically over the flash

flood area in *Sant Llorenç* (Fig. 17a). However, the rest of experiments show larger probability values over the eastern part of Mallorca. For the sake of brevity, here we only show two of them, EXP_14DA and EXP_16DA (Fig. 17b and Fig. 17c, respectively). The difference between these two experiments is notable, with EXP_14DA showing a maximum of probability located north of *Sant Llorenç* (Fig. 17b), compared with the maximum of EXP_16DA successfully located over the area of the flash flood (Fig. 17c). For the case of the 50 mm threshold, EXP_12DA shows negligible probabilities ($P \leq 0.1$) over the region of *Sant Llorenç* (Fig. 17d), although EXP_14DA and EXP_16DA produce higher probabilities ($P \leq 0.4$) over the same region. Lastly, for the 100 mm threshold, EXP_12DA, EXP_14DA and EXP_16DA show marginal but greater than zero probabilities over the eastern part of Mallorca (Fig. 17g,h,i). As expected, since the initializing analysis is taken further into the DAW (and so closer to 18 UTC), the forecast skill grows. These results indicate that forecasts initiated with EnKF analysis obtained 6 h earlier than the peak of the observed event, do not show potential to issue useful flash flood warning. In contrast, forecasts initialized from EnKF analyses with lead times of 5 h or less increasingly capture part of the damaging characteristics of the flash flood event.

To better understand the low probability values obtained we need to analyse in more depth the results of the above-mentioned experiments. In general, an undesirable behaviour is observed for all the numerical experiments during the first hours of free forecast. Results show that during the first 1–2 h of free forecast, the intensity of the small-scale convective cells generated by the assimilation of reflectivity

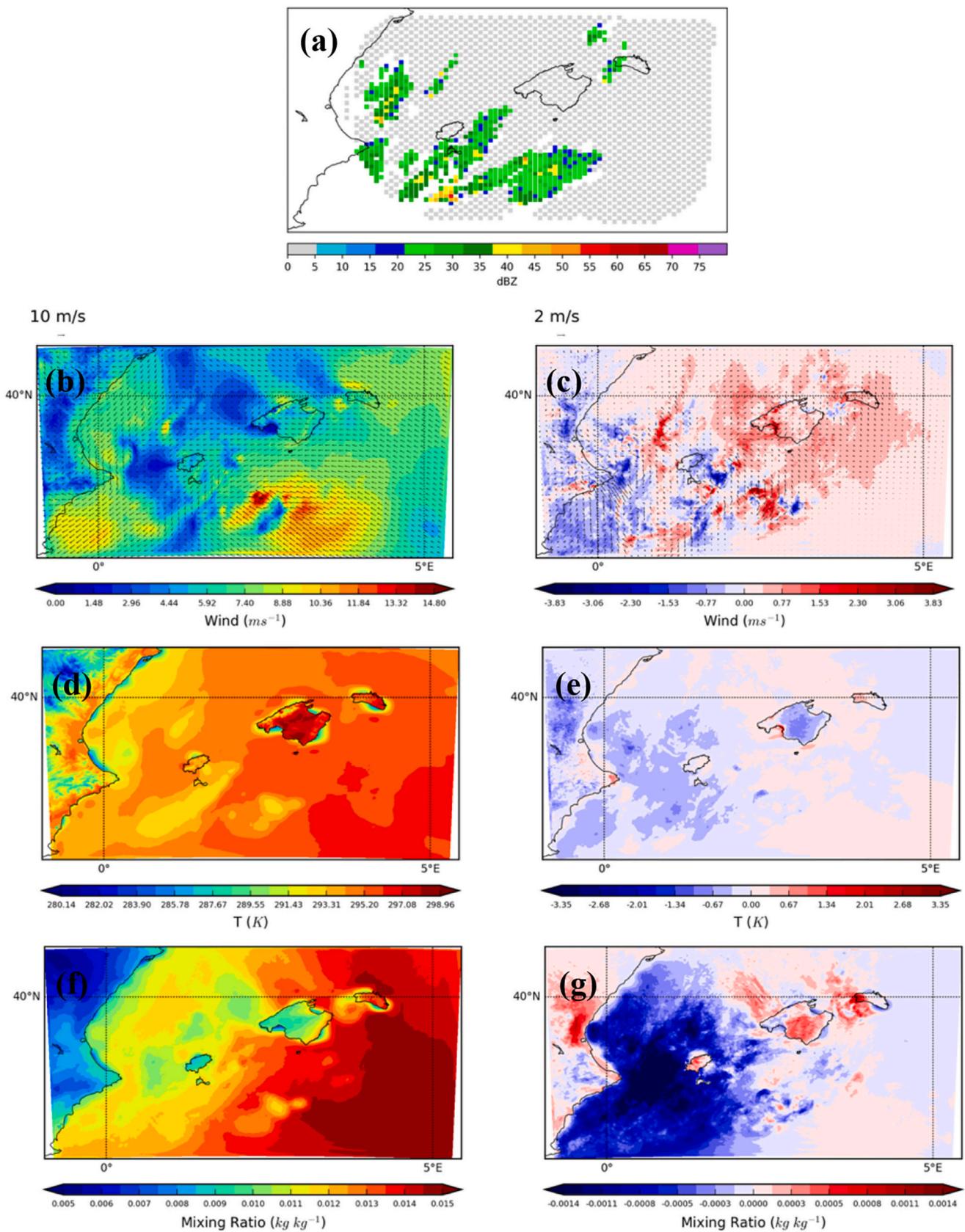


Fig. 14. Impact of the assimilation of reflectivity observations to non-observed variables at 12 UTC 9 October 2018. Panels (b), (d) and (f) show the prior for the wind direction and speed, temperature and mixing ratio at low levels, respectively. Panels (c), (e) and (g) show the increments (posterior-prior) for each of these non-observed variables. Reflectivity observations assimilated at this time is also depicted (a).

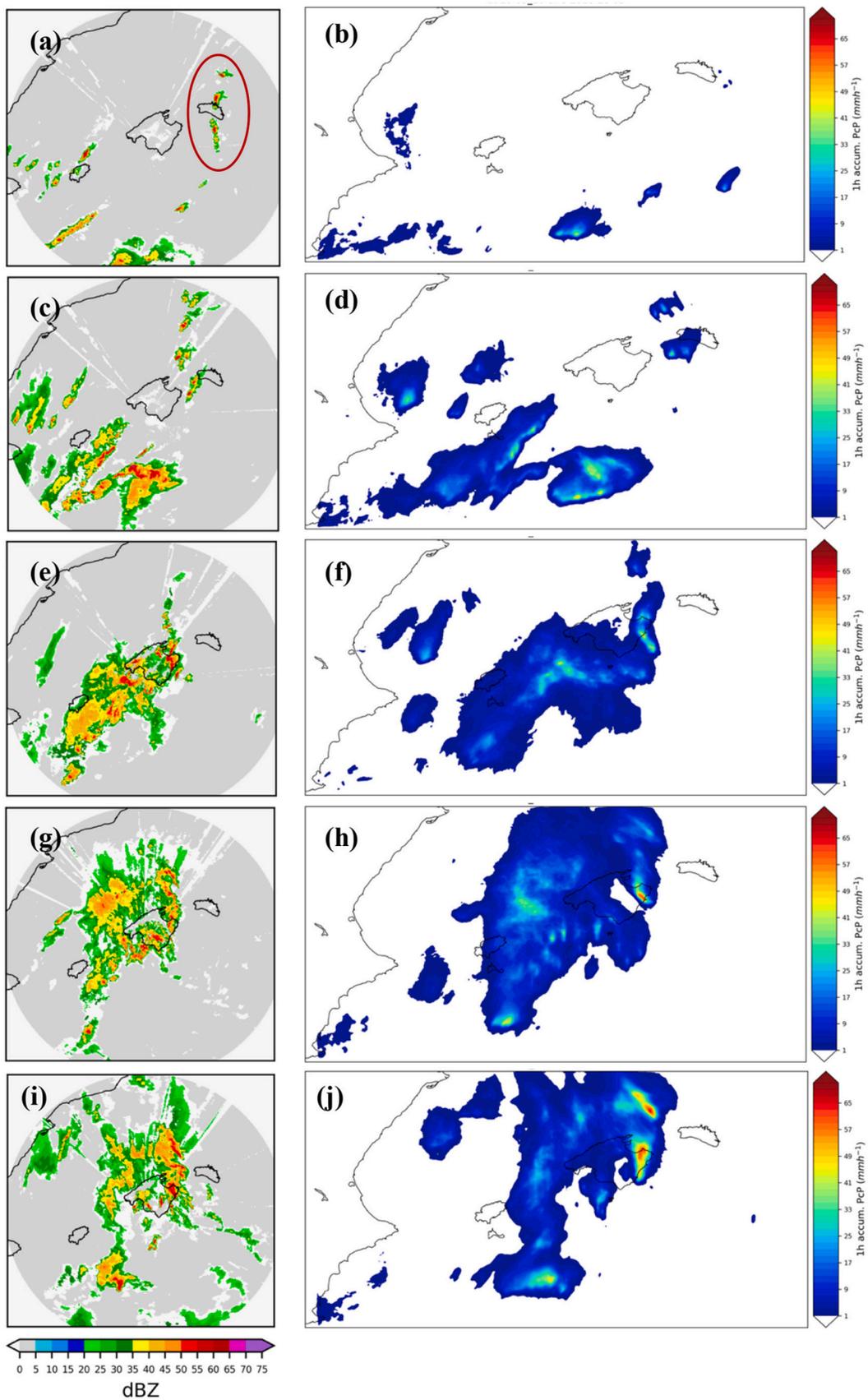


Fig. 15. Radar reflectivity observations at the first elevation angle (left column) and ensemble mean 1 h accumulated precipitation (right column) valid at (a-b) 10:00 UTC, (c-d) 12:00 UTC, (e-f) 14:00 UTC, (g-h) 16:00 UTC and (i-j) 18:00 UTC 9 October 2018. The initial linear convective system discussed in the text is highlighted in red in (a). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

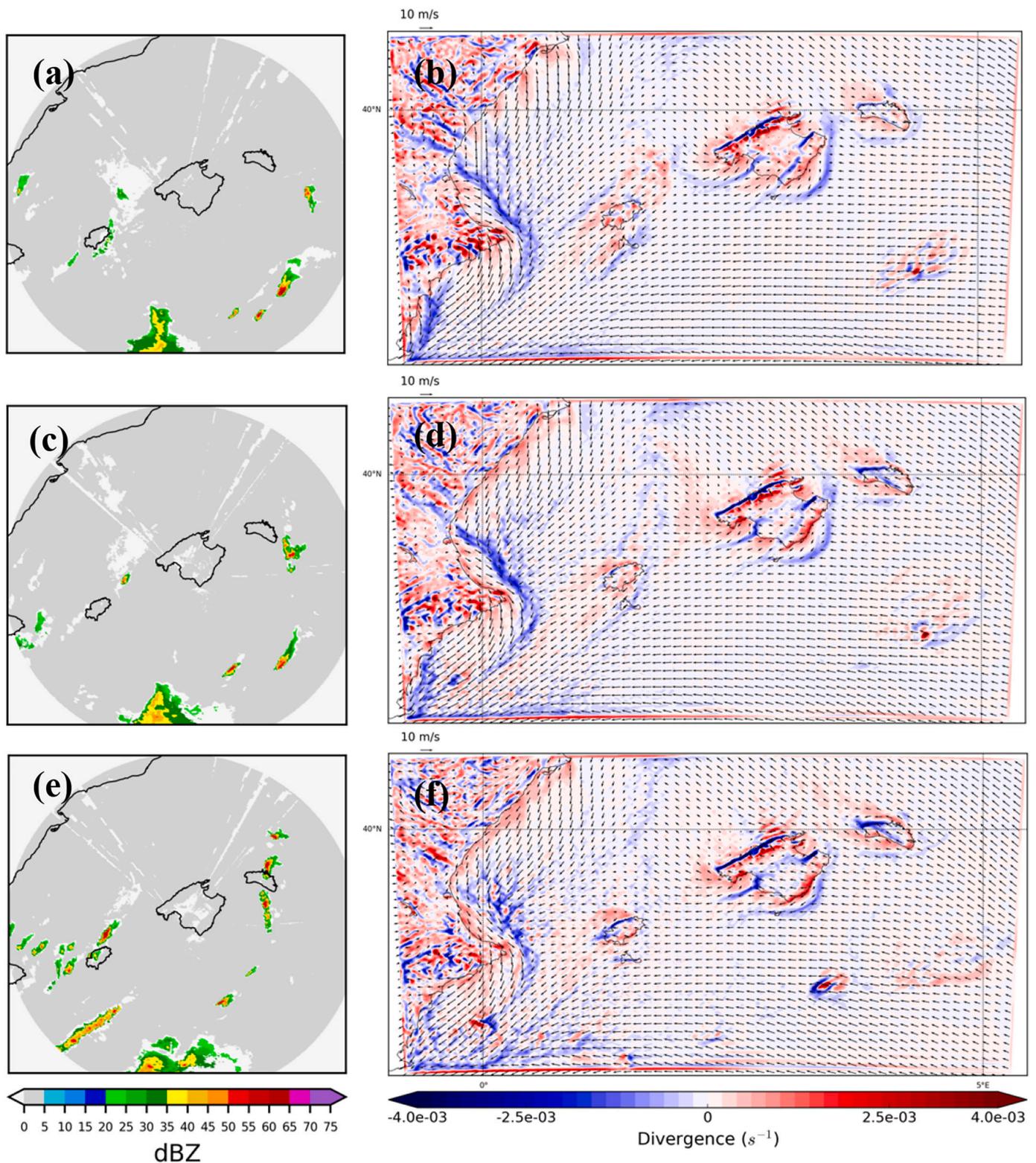


Fig. 16. Reflectivity observations at the first angle of radar elevation (first column) and ensemble mean low-level divergence fields (second column) corresponding to (a-b) 08, (c-d) 09 and (e-f) 10 UTC on 9 October.

observations reduces drastically and some of these cells even disappear (Fig. 18). Admittedly, this behaviour is observed over the whole numerical domain, but in this discussion, we focus on the convective cells generated by the EnKF on the eastern part of Mallorca (black ellipses in Fig. 18). Note that according to radar and satellite observations, these convective cells move westwards, in the opposite direction to the rest of

structures in the domain. However, we observe that during the free forecast, the EnKF convective cells generated on the eastern part of Mallorca do not evolve in the same direction than the ones observed by the radar. Instead, they moved northwards (not shown). These results clearly show that the information introduced by the EnKF is almost forgotten by the numerical model after the first hours of the free forecast.

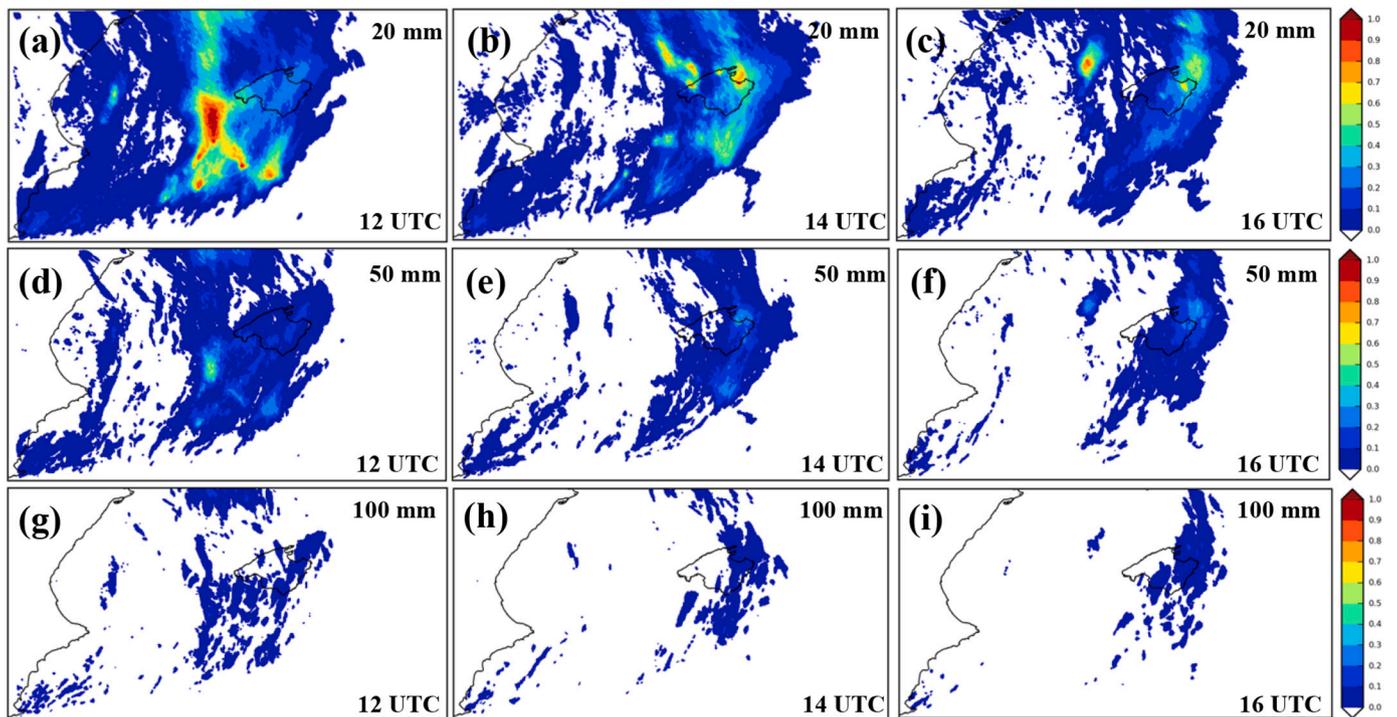


Fig. 17. Probability maps associated with the forecast accumulated precipitation being larger than the following thresholds: (a-c) 1 mm, (d-f) 20 mm and (g-i) 50 mm, specifically, first column (a) 1 mm/12 h, (d) 20 mm/12 h and (g) 50 mm/12 h by EXP_12DA ensemble. Similarly, the second column shows probability values for the following thresholds (b) 1 mm/10 h, (e) 20 mm/10 h and (h) 50 mm/10 h obtained by EXP_14DA ensemble. Finally, the third column corresponds to (c) 1 mm/8 h, (f) 20 mm/8 h and (i) 50 mm/8 h for the EXP_16DA experiment.

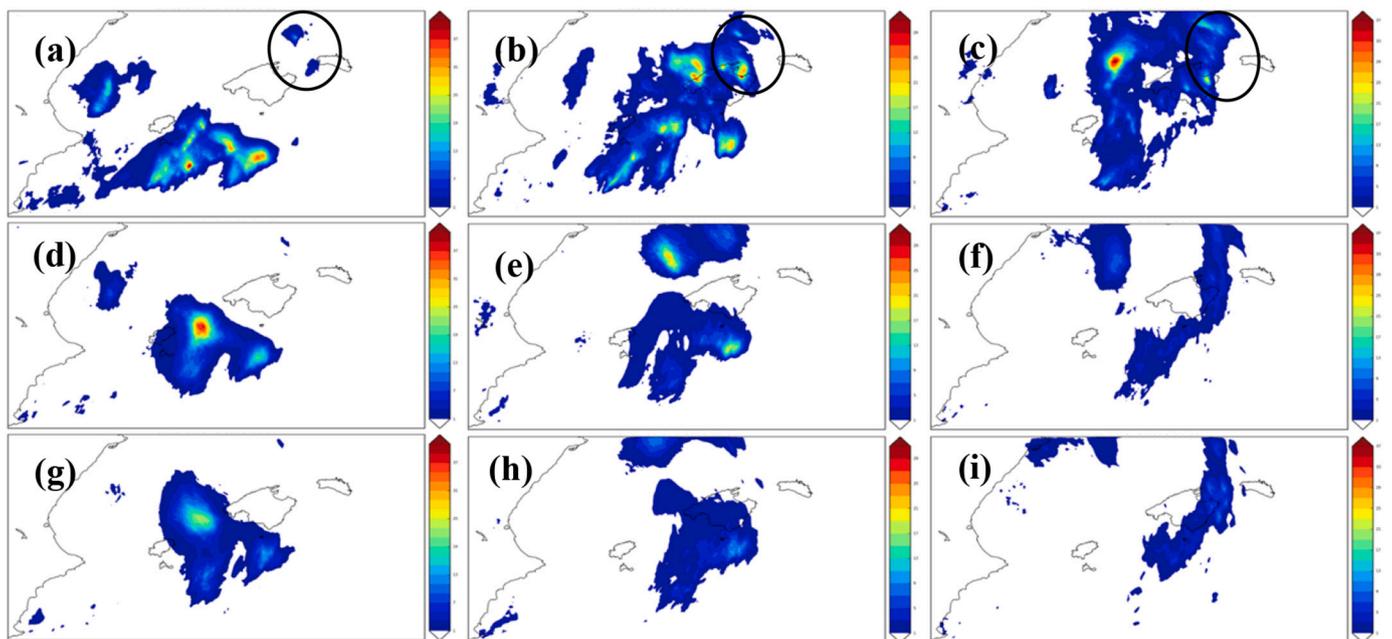


Fig. 18. Hourly forecast accumulated precipitation (ensemble mean) associated to the same experiments than in Fig. 18. Specifically, first column is associated with accumulated precipitation valid at (a) 13 UTC, (d) 14 UTC and (g) 15 UTC of EXP_12DA experiment. The second column show the respective hourly accumulated precipitation associated with (b) 15 UTC, (e) 16 UTC and (h) 17 UTC for the EXP_14DA ensemble. Finally, the third column corresponds to the accumulated precipitation valid at (c) 17 UTC, (f) 18 UTC and (i) 19 UTC for the EXP_16DA experiment. Black ellipses on (a), (b) and (c) highlight the first hour of accumulated precipitation associated with the easternmost convective cells generated by the EnKF for each experiment.

This effect is consistent with the findings of other works (e.g., Bick et al., 2016; Yussouf et al., 2016; Lawson et al., 2018; Bachmann et al., 2019; Carrió et al., 2019). Thus, the predictability of small-scale convective structures introduced by the assimilation of reflectivity radar

observations seems to be constrained to a couple of hours.

Here, we attempt to better understand why convective structures generated by the EnKF show this fast-dissipative effect during the first hours of the free forecasts. For this purpose, we performed an analogous

experiment to the CNTRL one, but without applying DA (hereafter NODA). Then, using the ensemble mean of both experiments, our approach is the following: we identify a small convective cell that evolves into a mature thunderstorm and then some of its thermodynamical features are compared to the ones obtained by convective cells generated by the EnKF in the eastern part of Mallorca, offshore. For the sake of brevity and clarity, here we show only one of these isolated convective systems generated by the EnKF over the eastern part of Mallorca (Fig. 19a). Results from vertical cross-sections show that the convective structure generated by the EnKF are similar in intensity and in vertical depth (Fig. 19c) to the ones observed in small-scale systems that intensify in time (Fig. 19d). Thus, from the point of view of the reflectivity, it seems that the EnKF analysis do not underestimate the convective activity. However, these cells vanish during the first minutes of the free forecast. To better understand this behaviour a cross-section of the equivalent potential temperature (θ_e) and the wind direction/speed is examined. The θ_e cross-section for the EnKF convective system seems to depict a well-defined column of moist and warm air (see black ellipse in Fig. 19e) at the same location where the maximum of reflectivity was shown in Fig. 19c. This vertical distribution of θ_e agrees with the one expected from small-convective systems that potentially grow in time, which typically exhibit an extended tongue of warm and moist air connecting the surface to the upper-levels favouring their vertical instabilities (Fig. 19f). However, a close inspection on the wind field shows

that, in fact, there is not such a clear intrusion of warm air from the surface to the upper levels. In particular, over the location where the maximum of reflectivity was observed, i.e., between the (lat,lon) coordinates (39.82, 3.92) and (39.78, 4.02), the general flux of air across the first 3 km (approx.) is directed downwards, restricting the intensification of such convective system (Fig. 19e). Also, between (lat,lon) coordinates (39.80, 3.97) and (39.78, 4.02), we have a cold column of air with strong downward winds, which would contribute to the dissipation of the convective cell. On the contrary, convective cells that intensify in time typically show a strong plume of upward warm air and moist air, favouring its intensification (Fig. 19f).

These results indicate that the limits in the predictability of this flash flood event are tightly related to the performance of the EnKF, but also to the poor performance of the numerical model showing clear air conditions over a region where rainfall was observed. The above discussed results also underline some fundamental issues associated with current operational DA techniques. These techniques are designed to work accurately under certain assumptions, such as the Gaussianity. In particular, the distribution of the truth given prior information is required to be a multivariate Gaussian distribution and also that the distribution of observations given a particular truth is also Gaussian and independent of the value of the truth. Positive-definite and bounded variables such as aerosol content, cloud, precipitation, or reflectivity are characterized by uncertainty distributions that are skewed and better

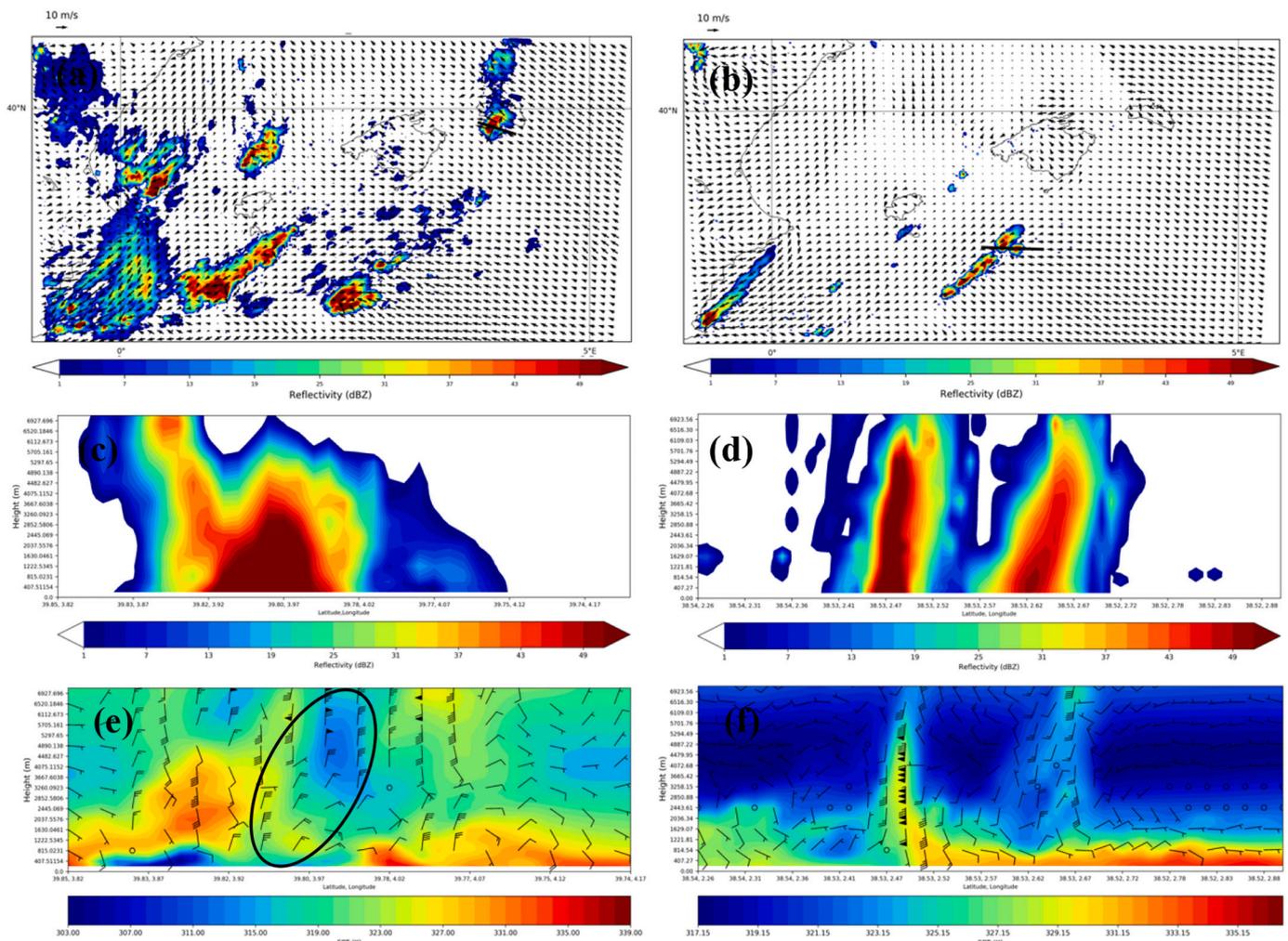


Fig. 19. Low-level reflectivity and wind speed/direction associated to one of the convective cells generated by: (a) EnKF and (b) NODA. Black thick line in (a) and (b) indicates the area where the cross-section is computed. Reflectivity cross-section for the EnKF and NODA ensembles are depicted in (c) and (d), respectively. Cross section of wind and equivalent potential temperature for the EnKF and NODA are also shown in (e) and (f). Black ellipse in (e) highlight the downdraft mentioned in the text.

approximated by gamma and inverse-gamma probability distribution functions than by Gaussian ones (Bishop, 2016). Another strong assumption also considered by these DA schemes is related with linear processes involved in the system. For instance, the relationship between the state and observation space conducted by forward observation operators is assumed to be linear. When DA is used to estimate cloud and precipitation properties, which have non-Gaussian uncertainty distributions and present non-linear relationships, these assumptions return a deficient representation of the Bayesian posterior distribution (Bishop, 2016; Posselt and Bishop, 2018). This study belongs to this case, where reflectivity observations present the same problems but are assimilated using the EnKF, producing inaccurate analyses. This could explain why the wind field of the convective cells performed by the EnKF were not consistent with the equivalent potential temperature field. However, note that the assimilation of radial velocities from the Doppler radar, which can be assumed to have a Gaussian uncertainty distribution, could significantly help to adjust the dynamic variables to the observations and in consequence improve the analysis of the convective cells. Unfortunately, radial velocities were not available during the presented episode.

5.4. Short-range forecast verification

In order to quantitatively assess the impact of the assimilation of conventional and reflectivity observations on the forecast, different categorical and probabilistic verification scores were computed. Verification over the region of this study is very challenging due to the lack of observations. In particular, we are interested in verifying the forecasts over the area affected by the flash flood episode. For this reason, the observations used for the verification are the 10-min rainfall amounts registered by the rain gauges over Mallorca (Fig. 1), which are distributed quasi-uniformly over the island.

We were interested in assessing the potential effect of the 12-h assimilation period (CNTRL experiment) regarding the 24-h accumulated precipitation over Mallorca, and more specifically over the region affected by the flash flood episode. For this purpose, the bias (model-observation) of the 24-h accumulated precipitation in observation space was computed (Fig. 20). In general, results show very low bias values, ranging between -10 mm/24 h and $+10$ mm/24 h over almost the

entire island, indicating a good correspondence between the numerical model and the observations. However, bias values are significantly larger in the region where the flash flood episode took place (black ellipse in Fig. 20): the model significantly underestimates the observed values at *Sant Llorenç*, *Artà* and *Colonia Sant Pere* stations (blue dots within black ellipse) with an approximated bias of -45 mm/24 h and overestimates the observations over *Son Servera* and *Cala Millor* (red points within black ellipse) with bias values slightly greater than 40 mm/24 h.

However, these results only account for the CNTRL experiment. To assess the sensitivity of the accumulated precipitation over the area of *Sant Llorenç* and *Colonia de Sant Pere* using the different EnKF-based initial conditions, we computed hourly series of cumulative precipitation for both regions. The hourly series of cumulative precipitation registered in this region is computed and compared with the values obtained by the CNTRL and the different forecast experiments. To compute the cumulative precipitations for the numerical simulations, an average of six-by-six model grid points centered in *Sant Llorenç* and *Colonia de Sant Pere*, respectively, were used (Fig. 21). Results for *Sant Llorenç* during the period from 07 UTC to 13 UTC show agreement between the EnKF analysis and the observations, with zero cumulative precipitation values (Fig. 21a). From 13 UTC to the end of the DA window at 18 UTC, cumulative precipitation values from the EnKF analysis overestimates the observations. The free forecast using the initial conditions obtained from the EnKF at 18 UTC shows cumulative precipitation values very close to the observations, although slightly overestimated (see dotted blue line in Fig. 21a). For the rest of the free forecast experiments it is observed that during the first one or two hours, the cumulative values are overestimated but after that short period, they start to underestimate the observations. As the initial conditions used by the forecasts departs from 18 UTC, this underestimation increases dramatically, suggesting that forecasts started more than two hours before 18 UTC substantially diverge of the observed state of the atmosphere after the first 4–5 h of free forecast and thus, they should not be trusted to infer information about the future state of the atmosphere. Analogous results were found for the case of *Colonia de Sant Pere* (Fig. 21b).

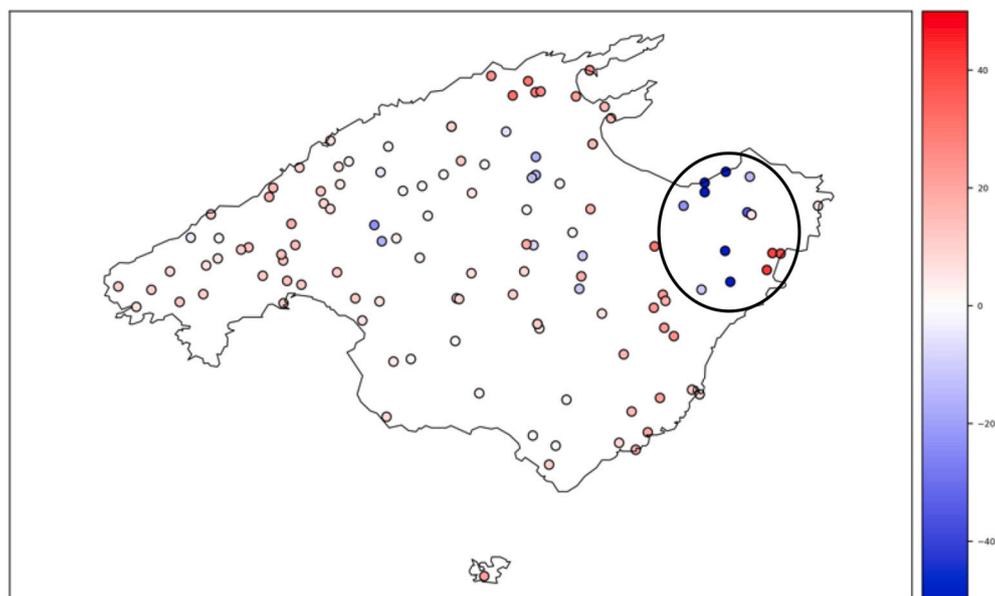


Fig. 20. Model bias (model-observations) for the 24-h accumulated precipitation computed using the rain gauges available over Mallorca. Black ellipse highlights the region most affected by the flash flood episode.

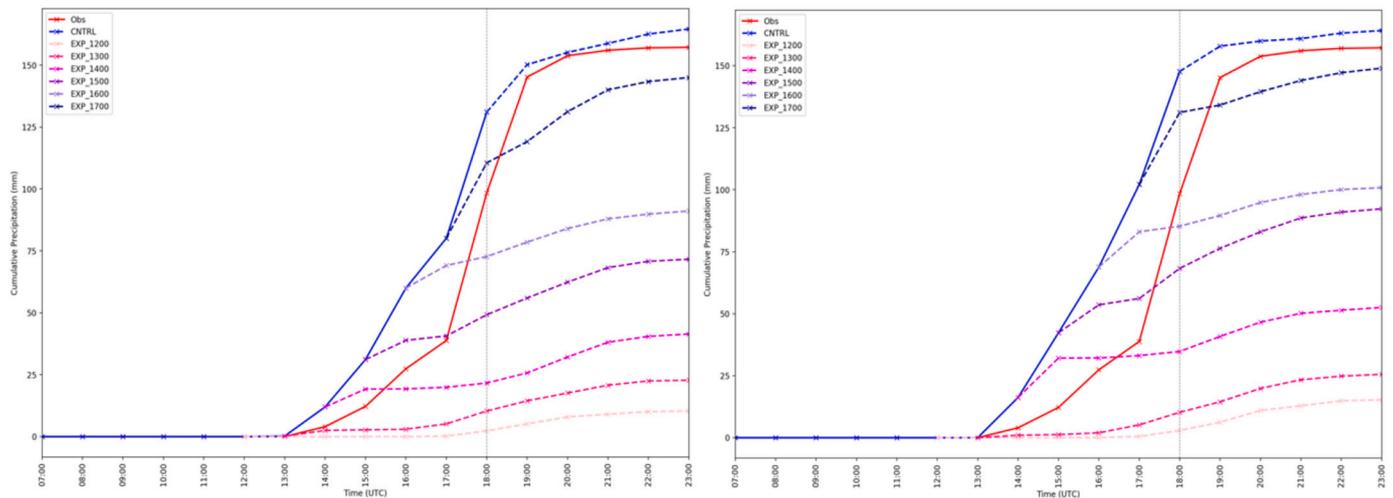


Fig. 21. Evolution of the cumulative precipitation registered by the rain gauges (red solid lines), simulated by the CNTRL forecast experiment (blue solid lines) and computed by the different forecast experiments (coloured dashed lines) at Sant Llorenç (left panel) and Colonia de Sant Pere (right panel). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

6. Conclusions

The Western Mediterranean is known as a geographical region prone to extreme weather events, such as high-intensity precipitation, strong winds or flash floods. These weather events typically affect densely populated coastal regions, leaving serious socio-economic impacts. Unfortunately, the accurate prediction of most of these extreme weather events remains still very challenging due to the small-scale processes and their interactions, among other reasons, involved in their genesis and evolution. In addition, these weather events typically initiate over the sea, which is poorly routinely sampled by standard observations. This fact makes that the representation of the atmospheric state in any numerical model is not sufficiently accurate.

In this study, we focus our attention on a recent flash flood event that took place over the region of *Sant Llorenç* (Mallorca), killing 13 people and producing a total damage exceeding EUR 90 million. Mainly because its small-scale characteristics and the high nonlinear physical processes involved in the genesis and evolution of this weather phenomena, the event was not captured by numerical weather forecast models. Even operational systems failed at predicting this event. For this reason, one of the main objectives of this study was to investigate if more sophisticated techniques, such as the EnKF, using high temporal and spatial observations (e.g., radar), could improve the prediction of this event and therefore provide valuable information to improve warning systems. Hourly in situ conventional observations and 10-min reflectivity radar observations were assimilated.

Results of the EnKF experiments for this extreme event seems to not capture realistically-well the true features of the observed convective systems, especially their dynamics. However, we show that short-range forecasts using analysis from the EnKF provide more valuable information than forecasts where no data assimilation is used. In particular, results reveal that forecasts initiated more than two hours ahead of the peak of registered intensity (~ 18 UTC) dissipate the convective structures obtained by the EnKF very fast, and so they do not provide realistic atmospheric evolutions beyond a couple of hours after the initiation of the numerical model. We hypothesized that the fast-dissipative behaviour of the convective cells introduced by the assimilation of reflectivity observations was due to a fundamental limitation of current DA methods, which are based on linearity and Gaussian assumptions. These assumptions are clearly not satisfied when assimilating nonlinear and non-Gaussian observations such as the radar reflectivity. However, it is important to note that recent research has emerged dealing with this problem. One of the most promising methods is the called GIGG-EnKF,

which assumes non-linearity relationship between state variables and observations and considers non-gaussian distributions, such as gamma and inverse-gamma distributions for semi-positive definite variables such as the reflectivity observations (Bishop, 2016). Although this method has only been applied to very simplistic models (e.g., Lorenz96 model), we are considering applying it to the *Sant Llorenç* flash flood event, in a near future work. Further research should be conducted using different DA schemes, such as the variational 4D-Var, to be able to generalize these results to other DA schemes. However, this comparison is beyond the scope of this study, and it is left for future work.

Also, it is important to note that, in principle, radial velocity observations from Doppler radar could also have a positive impact on better adjusting the dynamics of the assimilated convective systems and reducing their anomalous dissipative behaviour. Unfortunately, these observations are not available for this episode due to technical problems of the Palma radar. Future research will also exploit the potential impacts of these observations in similar cases to the one presented here, where the convective systems are practically generated by the DA process over maritime regions having sparse conventional observations.

Overall, the results presented in this study indicate that even using sophisticated generation techniques such as EnKF, the *Sant Llorenç* flash flood event is still a stimulating predictability challenge for the community.

Acknowledgements

This work is framed within the CGL2017-82868-R [Severe Weather Phenomena in Coastal Regions: Predictability Challenges and Climatic Analysis (COASTEPS)] Spanish project which was partially supported with AEI/FEDER funds. The first author was also supported by the ARC Centre of Excellence for Climate Extremes (CE170100023). Author A. Hermoso received funding from the Spanish Ministerio de Educación, Cultura y Deporte (FPU16/05133). Author A. Maimó received funding from the Spanish Ministry of Science, Innovation and Universities (FPU18/00520). The authors also acknowledge the computer resources at MareNostrum IV and the technical support provided by Barcelona Supercomputing Center (RES-AECT-2020-2-0011 and RES-AECT-2020-3-0001).

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.

References

- Aksoy, A., Dowell, D.C., Snyder, C., 2009. A multicaser comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Mon. Weather Rev.* 137 (6), 1805–1824.
- Altube, Patricia, et al., 2015. Quality control of antenna alignment and receiver calibration using the sun: Adaptation to midrange weather radar observations at low elevation angles. *J. Atmos. Ocean. Technol.* 32 (5), 927–942.
- Amengual, A., Carrió, D.S., 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 (4), 1143–1166.
- Anderson, J.L., 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* 129 (12), 2884–2903.
- Anderson, J.L., Collins, N., 2007. Scalable implementations of ensemble filter algorithms for data assimilation. *J. Atmos. Ocean. Technol.* 24 (8), 1452–1463.
- Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., Avellano, A., 2009. The data assimilation research testbed: a community facility. *Bull. Am. Meteorol. Soc.* 90 (9), 1283–1296.
- Bachmann, K., Keil, C., Weissmann, M., 2019. Impact of radar data assimilation and orography on predictability of deep convection. *Q. J. R. Meteorol. Soc.* 145 (718), 117–130.
- Berner, J., Fossell, K.R., Ha, S.Y., 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.
- Bick, T., Simmer, C., Trömel, S., Wapler, K., Hendricks Franssen, H.J., Stephan, K., Potthast, R., 2016. Assimilation of 3D radar reflectivities with an ensemble Kalman filter on the convective scale. *Q. J. R. Meteorol. Soc.* 142 (696), 1490–1504.
- Bishop, C.H., 2016. The GIGG-EnKF: ensemble Kalman filtering for highly skewed non-negative uncertainty distributions. *Q. J. R. Meteorol. Soc.* 142 (696), 1395–1412.
- Buizza, R., Miller, M., Palmer, T.N., 1999. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Royal Meteorol. Soc.* 125, 2887–2908.
- Buzzi, A., Tartaglione, N., Malguzzi, P., 1998. Numerical simulations of the 1994 Piedmont Flood: role of orography and moist processes. *Mon. Weather Rev.* 126, 2369–2383.
- Carrió, D.S., Homar, V., 2016. Potential of sequential EnKF for the short-range prediction of a maritime severe weather event. *Atmos. Res.* 178, 426–444.
- Carrió, D.S., Homar, V., Wheatley, D.M., 2019. Potential of an EnKF Storm-Scale Data Assimilation System over Sparse Observation Regions with complex Orography. *Atmos. Res.* 216, 186–206.
- Courtier, Philippe, Thépaut, J.N., Hollingsworth, A., 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* 120 (519), 1367–1387.
- Cressman, G.P., 1959. An operational objective analysis system. *Mon. Weather Rev.* 87 (10), 367–374.
- Daley, R., 1993. *Atmospheric Data Analysis*, vol. No. 2. Cambridge university press.
- Dowell, D.C., Wicker, L.J., 2009. Additive noise for storm-scale ensemble data assimilation. *J. Atmos. Ocean. Technol.* 26 (5), 911–927.
- Dowell, D.C., Zhang, F., Wicker, L.J., Snyder, C., Crook, N.A., 2004. Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble Kalman filter experiments. *Mon. Weather Rev.* 132 (8), 1982–2005.
- Dowell, D.C., Wicker, L.J., Snyder, C., 2011. Ensemble Kalman filter assimilation of radar observations of the 8 May 2003 Oklahoma City supercell: Influences of reflectivity observations on storm-scale analyses. *Mon. Weather Rev.* 139 (1), 272–294.
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* Oceans 99 (C5), 10143–10162.
- Fitzgerald, R., 1971. Divergence of the Kalman filter. *IEEE Trans. Autom. Control* 16 (6), 736–747.
- Fujita, R., Stensrud, D.J., Dowell, D.C., 2007. Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties. *Mon. Weather Rev.* 135 (5), 1846–1868.
- Gaspari, G., Cohn, S.E., 1999. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.* 125 (554), 723–757.
- Hacker, J.P., Anderson, J.L., Pagowski, M., 2007. Improved vertical covariance estimates for ensemble-filter assimilation of near-surface observations. *Mon. Weather Rev.* 135 (3), 1021–1036.
- Hamill, T.M., Whitaker, J.S., Snyder, C., 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.* 129 (11), 2776–2790.
- Houtekamer, P.L., Mitchell, H.L., 1998. Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.* 126 (3), 796–811.
- Jankov, I., Berner, J., Beck, J., Jiang, H., Olson, J.B., Grell, G., Smirnova, T.G., Benjamin, S.G., Brown, J.M., Jankov, I., Berner, J., Beck, J., Jiang, H., Olson, J.B., Grell, G., Smirnova, T.G., Benjamin, S.G., Brown, J.M., 2017. A performance comparison between multiphysics and stochastic approaches within a north American RAP ensemble. *Mon. Weather Rev.* 145, 1161–1179.
- Jansa, A., Alpert, P., Arbogast, P., Buzzi, A., Ivančan-Picek, B., Kotroni, V., Llasat, M.C., Ramis, C., Richard, E., Romero, R., Speranza, A., 2014. MEDEX: a general overview. *Nat. Hazards Earth Syst. Sci.* 14, 1965–1984. <https://doi.org/10.5194/nhess-14-1965-2014>.
- Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. *J. Fluids Eng.* 82 (1).
- Kalman, R.E., Bucy, R.S., 1961. *New Results in Linear Filtering and Prediction Theory*. Kalnay, E., 2003. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge university press.
- Kotroni, V., Lagouvardos, K., Kallos, G., Ziakopoulos, D., 1999. Severe flooding over central and southern Greece associated with pre-cold frontal orographic lifting. *Q. J. R. Meteorol. Soc.* 125, 967–991.
- Lakshmanan, V., Zhang, J., 2009. Censoring biological echoes in weather radar images. In: 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery, vol. 5. IEEE, pp. 491–495. August.
- Lawson, J.R., Kain, J.S., Yussouf, N., Dowell, D.C., Wheatley, D.M., Knopfmeier, K.H., Jones, T.A., 2018. Advancing from convection-allowing NWP to Warn-on-Forecast: evidence of progress. *Weather Forecast.* 33 (2), 599–607.
- Leith, C.E., 1993. Numerical models of weather and climate. *Plasma Phys. Control. Fus.* 35 (8), 919.
- Lorenz, A.C., 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* 112 (474), 1177–1194.
- Lorenzo-Lacruz, J., Amengual, A., García, C., Morán-Tejeda, E., Homar, V., Maimó-Far, A., 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 (11), 2597–2617.
- del Moral, A., del Carmen Llasat, M., Rigo, T., 2020. Connecting flash flood events with radar-derived convective storm characteristics on the northwestern Mediterranean coast: knowing the present for better future scenarios adaptation. *Atmos. Res.* 238, 104863.
- 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. *Q. J. R. Meteorol. Soc.* 134, 111–130.
- Parrish, D.F., Derber, J.C., 1992. The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Weather Rev.* 120 (8), 1747–1763.
- Posselt, D.J., Bishop, C.H., 2018. Nonlinear data assimilation for clouds and precipitation using a gamma inverse-gamma ensemble filter. *Q. J. R. Meteorol. Soc.* 144 (716), 2331–2349.
- Rabier, F., 2005. Overview of global data assimilation developments in numerical weather-prediction centres. *Q. J. R. Meteorol. Soc.* 131 (613), 3215–3233.
- Riosalido, R., Rivera, A., Martin, F., 1988. Development of a mesoscale convective system in the Spanish Mediterranean Area. In: Proc. 7th Meteosat Scientific Users' Meeting, Madrid, 27–30 September 1988, EUMETSAT, Darmstadt, pp. 375–378.
- Rivera, A., Riosalido, R., 1986. Mediterranean convective systems as viewed by Meteosat, a case study. In: Proc. 6th Meteosat Scientific Users Meeting, Amsterdam, 25–27 November 1986, EUMETSAT, Darmstadt, pp. 101–104.
- Romero, R., Doswell, C.A., Ramis, C., 2000. Mesoscale numerical study of two cases of longlived quasi-stationary convective system over eastern Spain. *Mon. Weather Rev.* 128, 3731–3751.
- Romine, G.S., Schwartz, C.S., Snyder, C., Anderson, J.L., Weisman, M.L., 2013. Model bias in a continuously cycled assimilation system and its influence on convection-permitting forecasts. *Mon. Weather Rev.* 141 (4), 1263–1284.
- Saltikoff, Elena, et al., 2019. OPERA the radar project. *Atmosphere* 10 (6), 320.
- Sénési, S., Bougeault, P., Chêze, J.L., Cosentino, P., Thepenier, R.M., 1996. The Vaison-La-Romaine flash flood: mesoscale analysis and predictability issues. *Weather Forecast.* 11, 417–442.
- Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D., Barker, D., Duda, M., Powers, J., 2008. A Description of the Advanced Research WRF Version 3, NCAR Technical Note. Mesoscale and Microscale Meteorology Division. National Center for Atmospheric Research, Boulder, Colorado, USA.
- Snyder, C., Zhang, F., 2003. Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Weather Rev.* 131 (8), 1663–1677.
- Sobash, R.A., Stensrud, D.J., 2013. The impact of covariance localization for radar data on EnKF analyses of a developing MCS: observing system simulation experiments. *Mon. Weather Rev.* 141 (11), 3691–3709.
- Tong, M., Xue, M., 2005. Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS experiments. *Mon. Weather Rev.* 133 (7), 1789–1807.
- Wheatley, D.M., Stensrud, D.J., Dowell, D.C., Yussouf, N., 2012. Application of a WRF mesoscale data assimilation system to springtime severe weather events 2007–09. *Mon. Weather Rev.* 140 (5), 1539–1557.
- Wheatley, D.M., Knopfmeier, K.H., Jones, T.A., Creager, G.J., 2015. Storm-scale data assimilation and ensemble forecasting with the NSSL Experimental Warn-on-Forecast System. Part I: Radar data experiments. *Weather Forecast.* 30 (6), 1795–1817.
- Wu, X., Zhang, S., Liu, Z., Rosati, A., Delworth, T.L., 2013. A study of impact of the geographic dependence of observing system on parameter estimation with an intermediate coupled model. *Clim. Dyn.* 40 (7–8), 1789–1798.
- Yussouf, N., Gao, J., Stensrud, D.J., Ge, G., 2013. The impact of mesoscale environmental uncertainty on the prediction of a tornadic supercell storm using ensemble data assimilation approach. *Adv. Meteorol.* 2013.
- Yussouf, N., Dowell, D.C., Wicker, L.J., Knopfmeier, K.H., Wheatley, D.M., 2015. Storm-scale data assimilation and ensemble forecasts for the 27 April 2011 severe weather outbreak in Alabama. *Mon. Weather Rev.* 143 (8), 3044–3066.
- Yussouf, N., Kain, J.S., Clark, A.J., 2016. Short-term probabilistic forecasts of the 31 May 2013 Oklahoma tornado and flash flood event using a continuous-update-cycle storm-scale ensemble system. *Weather Forecast.* 31 (3), 957–983.