Atmospheric Research 274 (2022) 106215

Contents lists available at ScienceDirect

Atmospheric Research

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

WRF hourly evaluation for extreme precipitation events

Andrés Merino [a,*], Eduardo García-Ortega [a], Andrés Navarro [b], José Luis Sánchez [a], Francisco J. Tapiador [b]

[a] Universidad de León (ULE), Atmospheric Physics Group (GFA), Environmental Institute, 24071 León, Spain
[b] University of Castilla-La Mancha (UCLM), Institute of Environmental Sciences (ICAM), Faculty of Environmental Sciences and Biochemistry, Earth and Space Sciences (ESS) Group, Avda. Carlos III s/n, 45071 Toledo, Spain

ARTICLE INFO

Keywords:
WRF model
Physics parameterizations
Precipitation extremes
Hourly precipitation

ABSTRACT

Precipitation is one of the most relevant fields in atmospheric modeling because of its environmental, social and economic implications. However, precipitation validation from weather model outputs presents substantial challenges, such as measurement uncertainties, use of gridded datasets vs. direct observations, and the selection of statistical goodness-of-fit measures. The main difficulty of working with precipitation is that it can be spatially irregular, especially in extreme events. High temporal aggregation smooths the field and reduces verification uncertainty. For this reason, validations are usually focused on a daily scale. However, many extreme events occur on shorter periods, for which a sub-daily precipitation assessment is required. In this paper, hourly precipitation verification of the Weather Research and Forecasting (WRF) model is explored for 45 extreme precipitation events (EPEs) recorded in northeastern Spain. For this, stations with recorded EPEs were classified according to the hourly distribution of precipitation. WRF simulations were established considering three microphysics and two planetary boundary layer (PBL) parameterizations. Finally, several statistical goodness-of-fit measures and spatial and temporal precipitation distributions were used for evaluating WRF performance. The results showed that microphysics were more important than PBL parameterizations. Goddard and Thompson together with Mellor-Yamada-Nakanishi and Nino PBL gave better results for most of the analyzed characteristics. However, an optimal combination of parameterizations was not obtained for all EPEs, because event characteristics had important effects on model performance.

1. Introduction

Precipitation forecasting is one of the most relevant fields in atmospheric modeling, owing to its environmental, social and economic implications. Numerical weather prediction (NWP) models are currently the basic tool on which precipitation forecasts are based (Sun et al., 2014). Accordingly, the Weather Research and Forecasting (WRF) model (Skamarock and Klemp, 2008) has a large worldwide community, with widespread use in research (García-Ortega et al., 2017; Karki et al., 2018; Li et al., 2021; Yang et al., 2021) because it is open-source code in the public domain and has a set of realistic and continuously evolving physical parameterization schemes.

Accumulated precipitation analysis under different temporal aggregates is of great interest for climate studies. However, when dealing with meteorological risks, intense precipitation deserves specific attention. Rainfall is the main driver of water resource availability but its concentration in time produces extreme precipitation events (EPEs) and severe negative effects on society, the environment and economy (Jonkman, 2005). In the Mediterranean Sea area, precipitation has strong spatiotemporal variability, resulting in a high frequency of EPEs (Llasat et al., 2013; Tramblay and Somot, 2018). This characteristic, combined with the substantial vulnerability of this densely populated area, make it a "hotspot" because of powerful impacts related to climate change (Giorgi, 2006). This justifies the interest in EPE studies.

Given the above, there have been several studies in recent years on the characteristics of EPEs, their spatiotemporal distributions (Grazzini et al., 2020), trends (Merino et al., 2016), and relationships to large-scale patterns (Vicente-Serrano et al., 2009; Mastrantonas et al., 2021). In addition, several studies have evaluated numerical model sensitivity for extreme precipitation event forecasts in the Mediterranean area (Romero et al., 2015; Trapero et al., 2013; Vich and Romero, 2010) and more specifically in the Ebro Valley (García-Ortega et al.,

* Corresponding author.
E-mail address: andres.merino@unileon.es (A. Merino).

https://doi.org/10.1016/j.atmosres.2022.106215
Received 27 August 2021; Received in revised form 16 February 2022; Accepted 19 April 2022
Available online 23 April 2022

Table 1
Ensemble model: microphysics and PBL parameterizations.

	Microphysics	PBL
Combination 1	Goddard	MYNN
Combination 2	Goddard	MYJ
Combination 3	Thompson	MYNN
Combination 4	Thompson	MYJ
Combination 5	Morrison	MYNN
Combination 6	Morrison	MYJ

2007, 2017). However, accurate forecasting of EPEs remains problematic and can even go undetected by operational forecast systems (Carrió et al., 2022). The reasons are known: strong spatiotemporal variability of precipitation during EPEs, highly localized EPEs triggered by local-scale factors, and a lack of full understanding of related physical processes. EPEs are generally associated with short-term convective precipitation, but mesoscale convective systems or persistent orographic precipitation can also produce EPEs of variable duration (Merino et al., 2021).

Accurate EPE forecasting remains a basic tool to minimize the associated risk, although such forecasts remain uncertain (Xie et al., 2015; Stegall and Kunkel, 2019). Several studies using climate (Sun and Liang, 2020) and weather models (García-Ortega et al., 2017; Comin et al., 2018) have investigated model sensitivity to varying physics representations and underlying mechanisms, because the most important uncertainty lies in representing physical processes (Allen and Ingram, 2002). Karki et al. (2018) evaluated several microphysics parameterization schemes using WRF simulations of EPEs over the Central Himalayas, showing the Thompson scheme to have the best agreement with observations. Yang et al. (2021) evaluated the performance a WRF multi-physics ensemble for EPEs in China. The results suggest that selection of the cumulus scheme was more important than microphysics schemes for the precipitation forecast in the case of low-resolution simulations. Another source of uncertainty influences precipitation verification because of the intrinsic nature of observations, e.g., errors in measurement systems, available station density, and verification methods.

To minimize uncertainties, rainfall has usually been studied on a daily scale (Douluri and Chakraborty, 2021). However, EPEs can have a shorter duration, especially convective events, necessitating sub-daily assessment. Model evaluation at sub-daily scale constitutes a major challenge because EPEs are characterized by strong spatiotemporal variability in the precipitation field. Therefore, it is reasonable to expect

Table 2
Average intensity, duration, and altitude of events by cluster.

Cluster	Max hourly intensity (mm)	Max daily intensity (mm)	Duration (hours)	Altitude (m)
1	12.1	68	14.5	1013
2	28.3	91.3	10.9	932
3	13.1	83.5	16	1150
4	24.3	69.2	8.7	865
5	9.6	69.3	17.3	1019

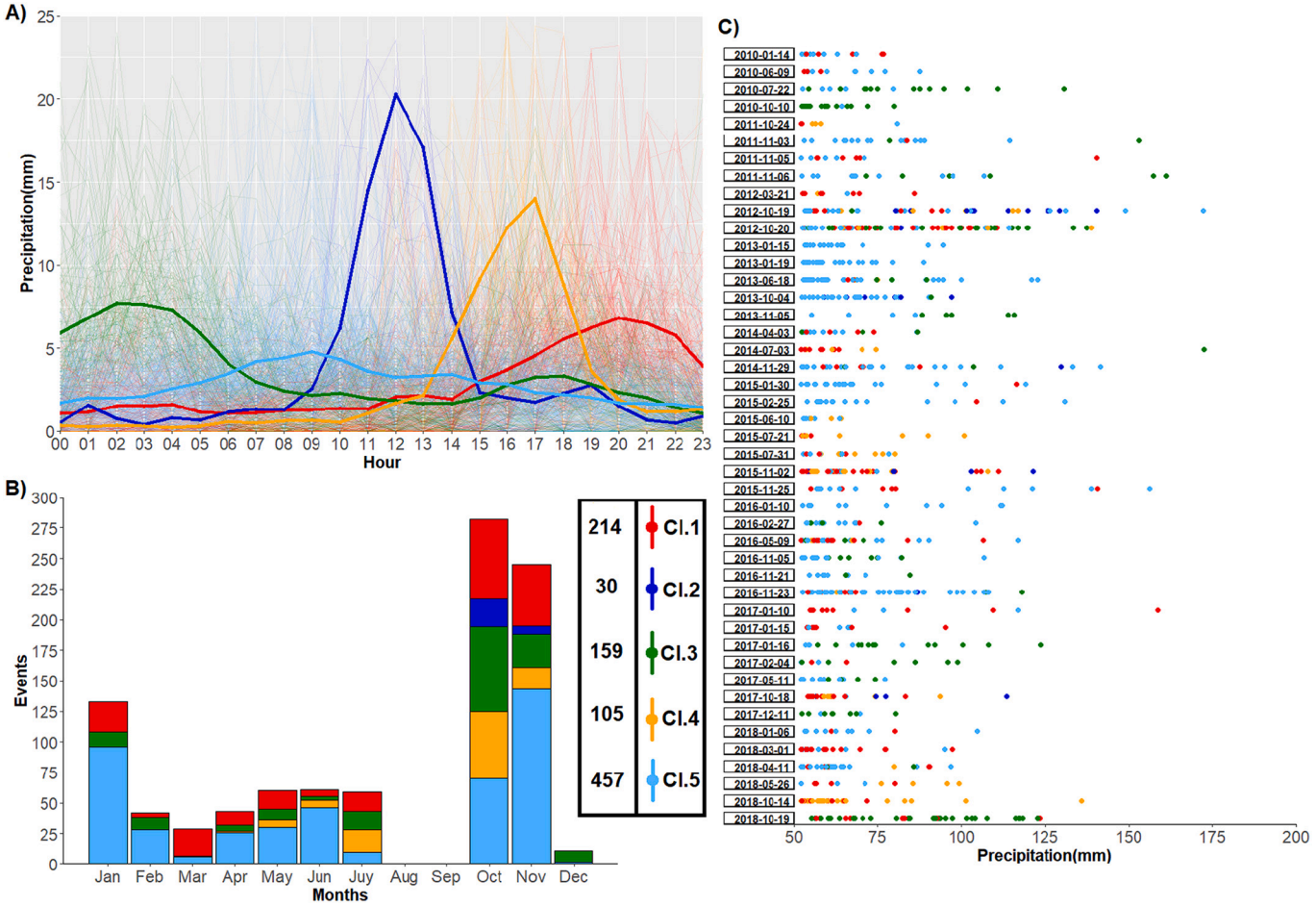

Fig. 1. Cluster classification of EPEs: red Cl.1; blue CL.2, green Cl.3; orange Cl.4; light blue Cl.5. A) Hourly precipitation (mm h^{-1}) distribution, thick line shows mean cluster; B) monthly distribution; C) EPE distributions by study days and 24-h accumulated precipitation by cluster (colour dots). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

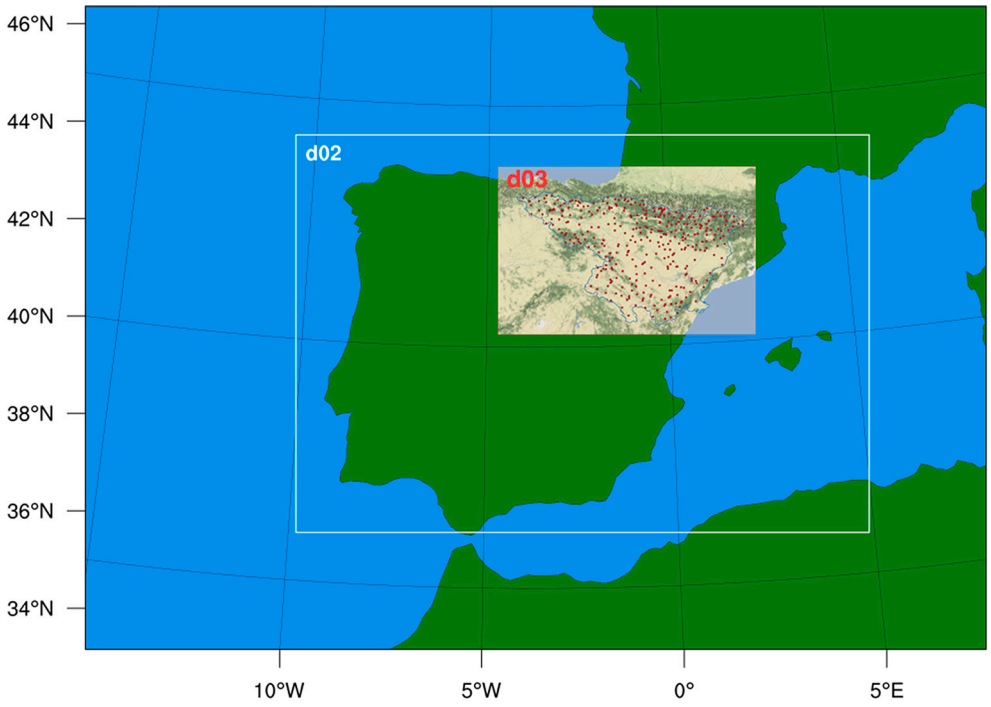

Fig. 2. WRF domains. Topographic map of Ebro River basin and rain gauge network (red dots). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

results substantially worse than for longer temporal aggregates. Recently, Kong et al. (2022) examined the sensitivity of hourly precipitation to cumulus parameterization and radiation schemes using a tropical-belt version of WRF. They found greater sensitivity for precipitation frequency than for amount. Frequencies of greater than moderate rate were well reproduced, whereas frequencies of more than light drizzle and heavy rates had large deviations.

Precipitation from grid models and point observations can be compared using two approaches. The first interpolates rain gauge observations to the resolution of the grid model and the second compares point-to-grid. Both options can lead to uncertainties in the verification results. In the first case, grid quality strongly depends on the distribution and spatial density of the stations (Merino et al., 2021). In many cases, the available observational data do not allow construction of the intended grid resolution without considerably increasing uncertainties. In the second case, spatial uniformity of the pixel must be assumed. The higher the grid resolution and more spatially uniform the field, the less inaccuracy. Considering the results of Merino et al. (2021) and Hofstra et al. (2008), who questioned the applicability of grid datasets for validation studies at grid-cell level (particularly on shorter time scales), and considering the high spatial resolutions using in mesoscale models, it seems a better option to use the second approach to evaluation.

The objective of the present work was to appraise the WRF model for its hourly precipitation performance for EPEs. First, a WRF ensemble was designed with three microphysics and two planetary boundary layer (PBL) schemes for each day on which EPEs were recorded. Subsequently, for studying the EPEs based on precipitation type, they were classified according to the hourly precipitation distribution. Finally, WRF hourly precipitation by cluster was investigated using scatter plots, density plots, and statistical goodness-of-fit measures.

This paper is organized as follows. Databases and study area descriptions are in Section 2. Section 3 describes the WRF model and methods for precipitation validation. Results are summarized in Section 4, and a discussion and conclusions are found in Section 5.

2. Data and study area

The Ebro Basin in northeastern Spain has characteristics making the area suitable for precipitation studies: a dense rain gauge network with 10-min data maintained and calibrated by the Ebro Hydrographic Confederation; strong climatic variability and rainfall gradients; a pronounced annual precipitation cycle with large regional differences; frequent EPEs in all seasons. For this reason, there have been studies with distinct approaches in recent years: deep convection and hailstorm analysis (García-Ortega et al., 2012; Merino et al., 2013; Marcos et al., 2021); hailstorm variability (García-Ortega et al., 2014); precipitation measurement by remote sensing (Merino et al., 2014; Navarro et al., 2020); physical scheme validation (García-Ortega et al., 2017); identification of precipitation extremes (Merino et al., 2018); gridded precipitation dataset analysis (Merino et al., 2021).

In our study, a set of 367 rain gauge (tipping-bucket type) stations was used. The stations are distributed throughout the basin, with a greater density in the highlands (Fig. 2 and Fig. 1A in Merino et al., 2021). Hourly and daily precipitation data were retrieved from the stations between 2010 and 2018. Quality control (QC) was done using the R package *reddprec* (Serrano-Notivoli et al., 2017). See Merino et al. (2021) for more details about the QC results.

3. Methods

First, EPEs were selected from precipitation datasets. The EPE definition followed the methodology of Merino et al. (2016), i.e., days on which the total precipitation amount exceeded the station-specific 99th percentile of daily precipitation on "wet" days (precipitation $\geq$1 mm). Days on which 10 or more stations met the criteria were thereby selected, giving 45 cases with 965 stations exceeding the 99th percentile. The station number criterion was used to avoid too localized events that could not be properly captured by the model resolution.

Later, hourly precipitation data were retrieved from the 965 stations. The objective was to analyze precipitation on a sub-daily scale, because EPEs based on daily precipitation include intense precipitation events over a few hours and very persistent rainfall events that produce heavy

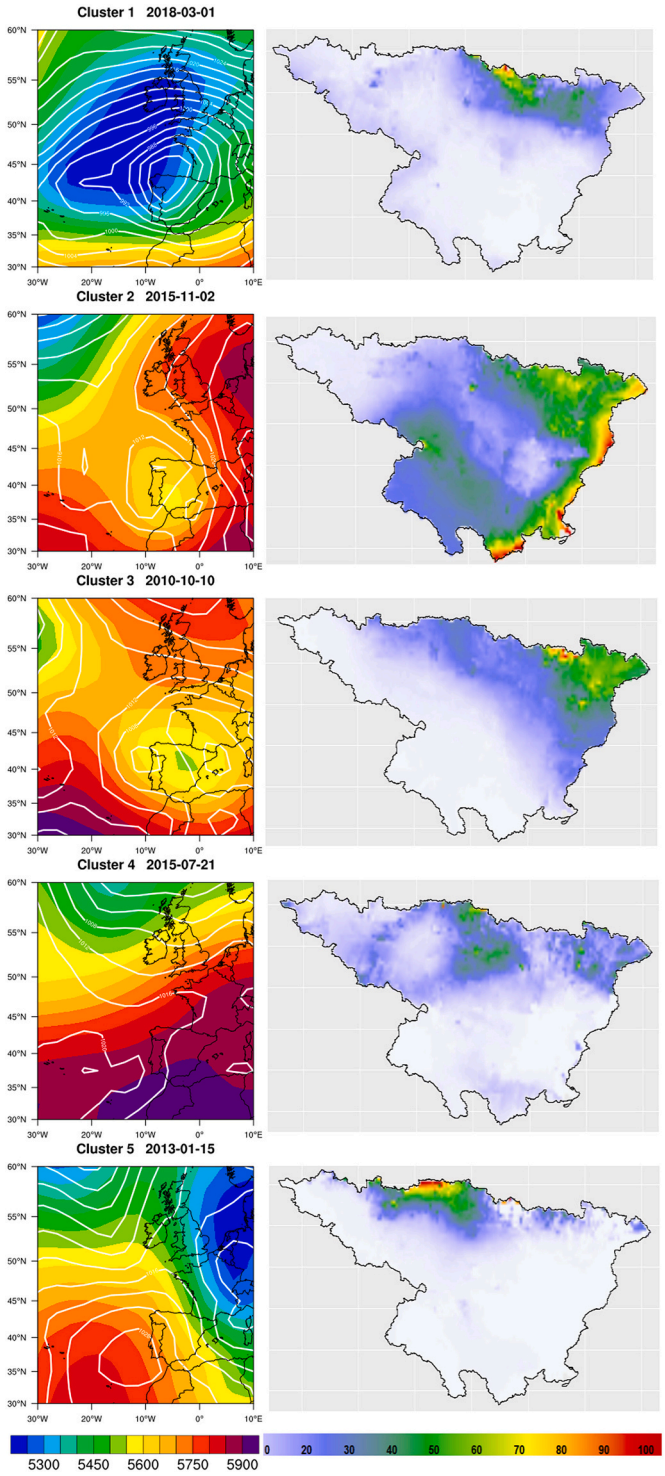

Fig. 3. Example study cases for each cluster. Left: Synoptic environment, colored areas represent geopotential height at 500 hPa and contours sea-level pressure. Right: 24-h accumulated precipitation (mm).

accumulated precipitation over many hours. Both types of event can pose considerable hydrologic hazards but the nature of the precipitation can be very different. It is therefore interesting to analyze model behavior vs. rainfall type. Rainfall type was defined by considering the intensity and duration of the precipitation and its spatial and seasonal distribution. For this, the 965 stations were classified using non-hierarchical k-means clustering (Hartigan and Wong, 1979) and hourly precipitation. The objective was to classify stations with similar rainfall evolution throughout the day. Euclidean distance was chosen for classifying groups of data according to their similarity. One of the sources of subjectivity in this method is the requirement that the number of final conglomerations (k) is determined ahead of time. The selection of k can be done objectively by computing the minimum decrease of intragroup distances. Nevertheless, the decision regarding the number of groups is not completely objective because there is some subjectivity based on researchers' experience (Gong and Richman, 1995). In the present study, the total within-group sums of squared distances were computed for k = 2, 3, 4, … 15, showing a minimum decrease for k = 5, considered the optimal cluster number.

Numerical simulations of the 45 study cases were run using the non-hydrostatic Advanced Research WRF model version 4.1 (Skamarock and Klemp, 2008). Initial and boundary conditions were from the National Oceanic and Atmospheric Administration / National Centers for Environmental Prediction Global Forecasting Model reanalysis with 0.25° horizontal grid spacing, providing time-varying lateral boundary conditions at 6-h intervals. Following a two-way nesting strategy, three domains with horizontal resolutions 27, 9 and 3 km were developed (Fig. 2). Each selected case was a 24-h precipitation event, but a 30-h simulation was done for the first 6 h as a spin-up period. A total of 50 vertical sigma levels was defined with progressive spacing, which was denser at low levels for better representation of the PBL and convective initiation.

Parameterizations are essential for representing the effects of subgrid processes in the atmosphere, such as turbulence, convection, and microphysics. Transport phenomena and thermodynamic processes alter the population of cloud hydrometeors that is mainly governed by microphysics and PBL schemes. In particular, precipitation is very sensitive to the choice of microphysics in NWP models (Tapiador et al., 2019a). According to previous results from the study area (García-Ortega et al., 2017), three microphysics schemes were selected (Table 1): (I) The Goddard Cumulus Ensemble one-moment bulk microphysical scheme (Tao et al., 2016). It includes cold rain processes (Rutledge and Hobbs, 1984; McCumber et al., 1991) and a cloud ice-to-snow conversion process. (II) The new Thompson scheme (Thomson et al., 2004, 2008) with a double-moment scheme for cloud ice and rain. In this, graupel is represented by a generalized gamma function (Jankov, 2011). (III) The Morrison 6-class double moment scheme (Morrison et al., 2009), which includes mixing ratios and predicted number concentrations for liquid and solid phases. Different rates of rain evaporation in stratiform and convective regions were considered, as well as predicted rain size distributions.

The poor-accuracy representation of lower-tropospheric thermodynamic and kinematic structures is a source of uncertainty in mesoscale models (Jankov et al., 2005; Stensrud, 2007; Cohen et al., 2015), particularly regarding severe weather events. The PBL plays a fundamental role in precipitation-forming processes (Moya-Álvarez et al., 2020). Turbulent eddies and mixing in the PBL occur on spatiotemporal scales smaller than grid scale and must be represented by parameterization schemes. Two local formulation schemes for the PBL were considered in our study (Table 1): (I) Mellor-Yamada-Janjic (MYJ; Mellor and Yamada, 1982; Janjic, 1994), which uses a 1.5-order (level 2.5) turbulence closure model. It determines local vertical mixing and evaluates the turbulent kinetic energy diffusion outside the mixed layer, improving the representation of entrainment. (II) The Mellor-Yamada Nakanishi and Nino Level 2.5 scheme (MYNN) that uses liquid water potential temperature and total water content. Buoyancy processes are considered in diagnosing the pressure covariance terms. This scheme deals with the effect of stability on the turbulent length scale.

In addition, based on the results of other studies, the following were selected: Dudhia (1989) for shortwave radiation; the Rapid Radiative Transfer Model (Mlawer et al., 1997) for longwave radiation; Eta surface layer described by Janjic (1994) and Noah Land Surface Model (Chen and Dudhia, 2001); Kain-Fritsch cumulus scheme (Kain, 2004) with explicit resolution of cumulus in the inner domain.

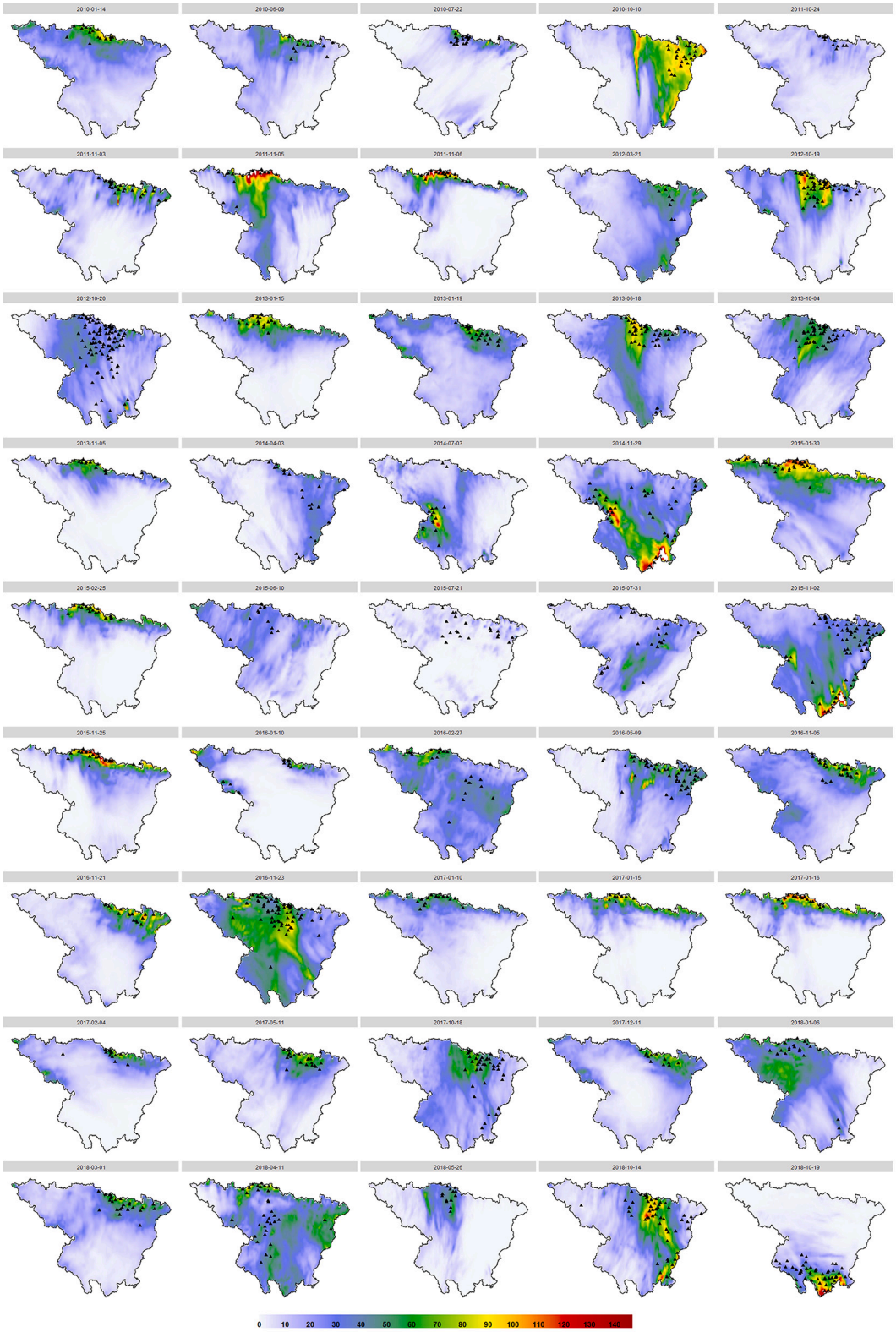

Fig. 4. Spatial distribution of 24-h accumulated precipitation (mm) mean ensemble for each study case. Black-filled triangles represent stations exceeding the threshold of EPE.

The rain gauges are very unevenly distributed in our study area, so the grid building for WRF verification can generate additional uncertainties (Merino et al., 2021). Thus, the verification was performed using WRF grid-to-point comparisons. Although this method may also lead to uncertainties, considering pixels to be spatially uniform and use of the WRF model with high spatial resolution minimizes the impact.

The hourly precipitation by cluster was evaluated using density plots and statistical goodness-of-fit measures. First, the modified Kling-Gupta efficiency (KGE; Gupta et al., 2009, Kling et al., 2012) was chosen. This index compares observed and modeled precipitation, decomposing total

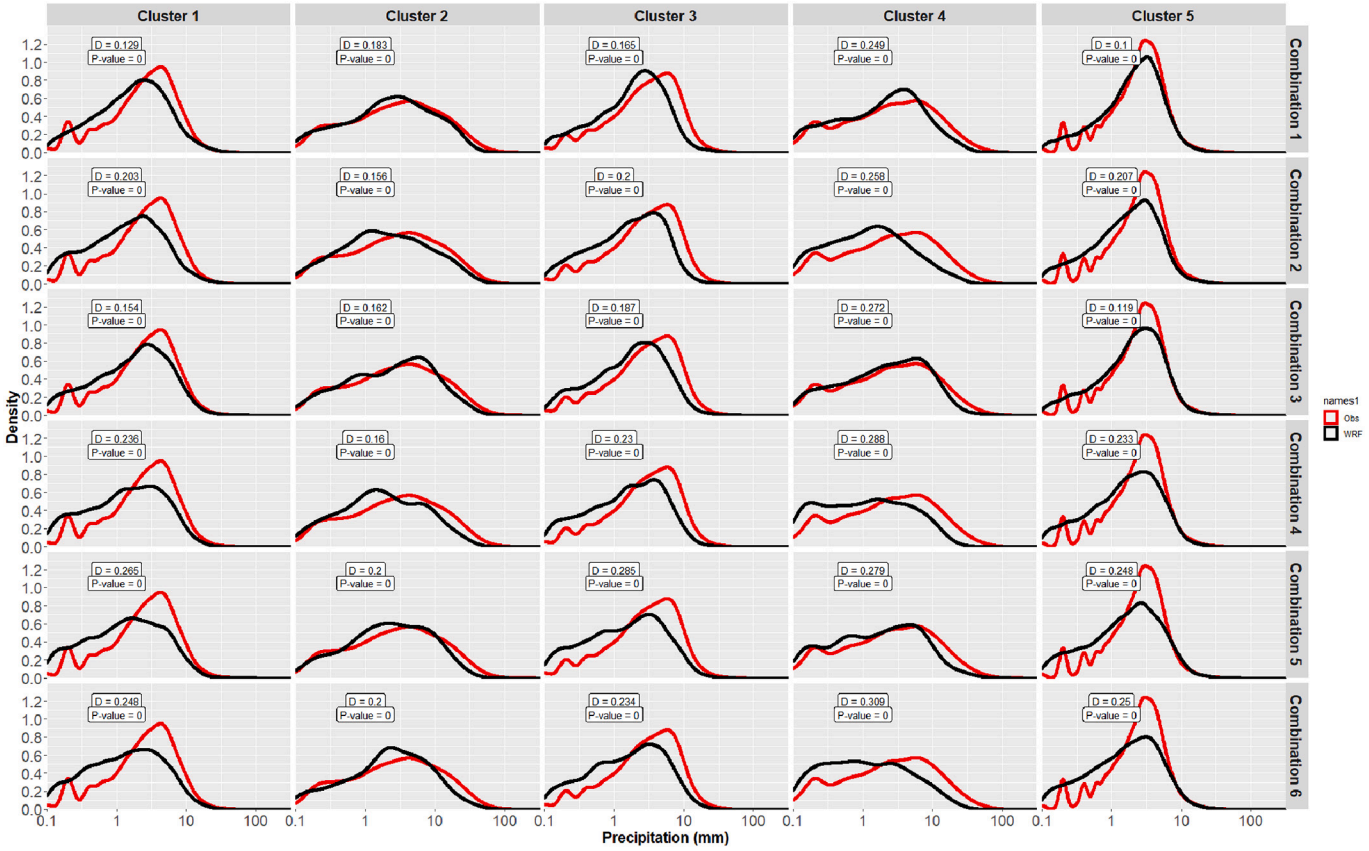

Fig. 5. Density plots for observed hourly precipitation distribution (red line) vs. WRF precipitation (black line). Rows show different WRF combinations and columns show clusters. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

performance into three components: linear correlation (r), bias ratio (β), and variability ratio (γ). r is the Pearson product-moment correlation coefficient, β measures total precipitation compared to ground-based observations, indicating the average tendency of the grid precipitation to underestimate ($\beta < 1$) or overestimate ($\beta > 1$), and γ measures the relative dispersion between the gridded and ground-based measurements. The optimal value for KGE and all its components is one. This index has been widely used to evaluate the performance for precipitation (e.g., Somos-Valenzuela and Manquehual-Cheuque, 2020; Merino et al., 2021).

Classic performance measures based on pairwise comparison between model and observation have several drawbacks for hourly precipitation. Precipitation is a discontinuous variable with strong spatiotemporal variability. For this reason, models are usually validated using daily accumulated precipitation. In such a case, hourly precipitation data, the temporal precipitation distribution throughout the event, and the distance of maximum precipitation between observed and model have been evaluated using density plots, which is a smoothed version of the histogram. Visual comparison of the distribution is complemented by adding the D statistic from the Kolmogorov-Smirnov (K-S) test as a measure of deviation between the two distributions. That test measures the maximum difference between observed and modeled cumulative distributions. The distance statistic of the K-S test and its *p*-value were computed. The null hypothesis assumes that both samples come from a population with the same distribution. When that hypothesis can be significantly rejected, there is a significant difference between the distribution of the two samples.

4. Results

WRF precipitation was evaluated considering its hourly distribution throughout an event. The strong climate variability of the study area facilitates extreme events spanning a large spatiotemporal range. Isolated and intense convective episodes producing extreme rainfall over limited areas are frequent. Convective events are more common in spring and summer but are also possible in winter. On the other hand, persistent precipitation events, associated with strong advection and orographic lift, produce large accumulations over many hours. These events are more frequent in winter and tend to cover extensive areas near mountains. (Merino et al., 2016; Merino et al., 2018).

First, the 965 stations (events) exceeding the 99th percentile were classified according to observed hourly precipitation. Five clusters were obtained, with different precipitation distributions across the events (Fig. 1). The mean hourly precipitation by cluster (Fig. 1A) shows that clusters 2 and 4 have heavy precipitation concentrated over a few hours, around 12 UTC for cluster 2 and 17 UTC for cluster 4, whereas observed precipitation is scarce away from this peak. This behavior is typical of convective precipitation, which has small spatiotemporal scales. Thus, events belonging to these clusters have been recorded only in summer and fall (Fig. 1B). Clusters 1 and 3 represent substantial precipitation over several hours (Fig. 1A). The precipitation is concentrated in the final hours of the event for cluster 1 and in the first hours for cluster 3. These are events midway between short and intense convective events and prolonged stratiform events. Finally, in cluster 5, precipitation occurs throughout the event, albeit with weak intensity, typical of sustained stratiform events. These findings are consistent with the average intensity, duration, and altitude of events by cluster (Table 2). Events of clusters 2 and 4 had the greatest maximum intensity of hourly precipitation and shortest duration. In addition, the stations affected by extreme events have a lower average altitude, because convective precipitation can appear throughout the basin. Cluster 5 events had the weakest hourly intensity and maximum duration, characteristics of very

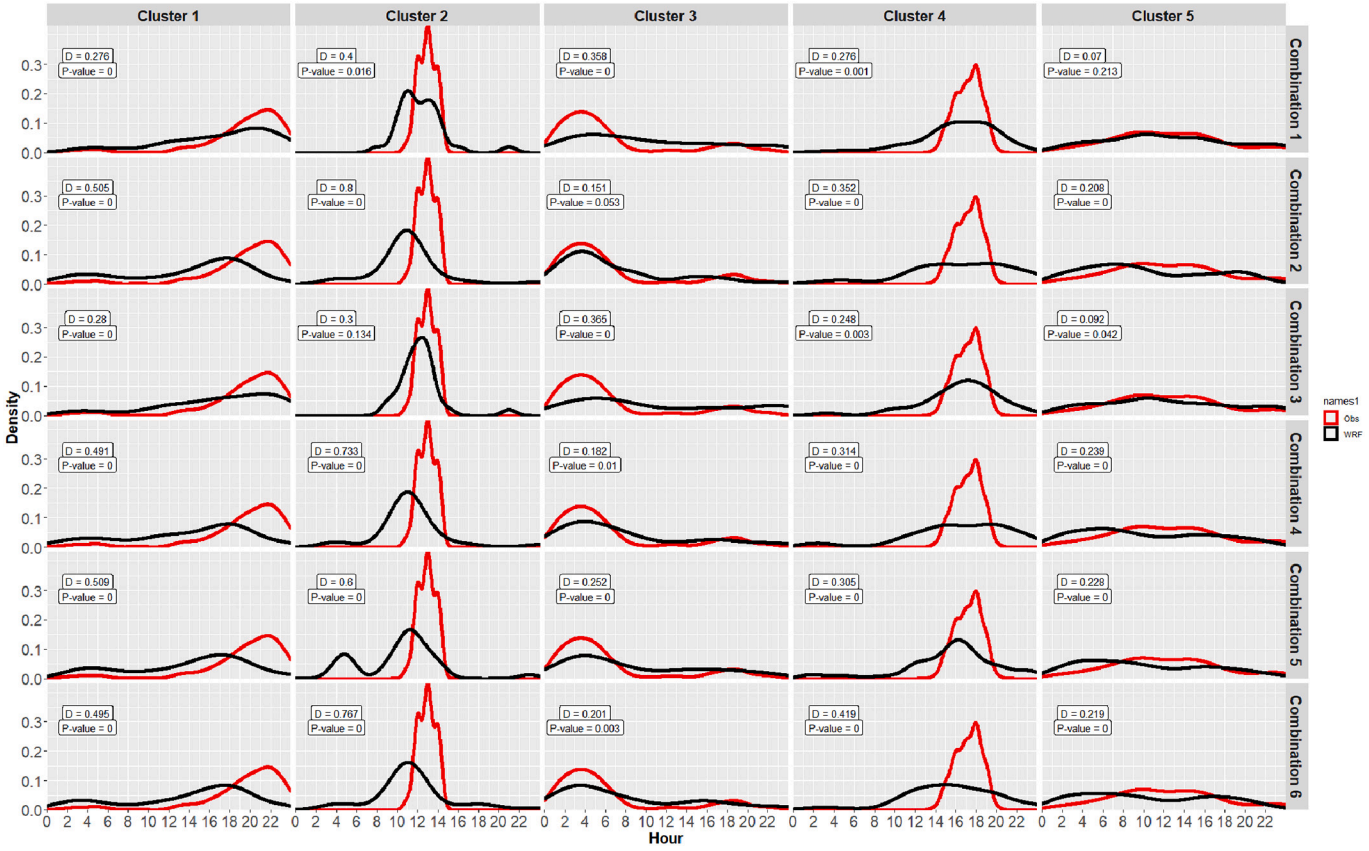

Fig. 6. Density plots for temporal distribution (UTC) of observed hourly precipitation (red line) vs. WRF precipitation (black line). Rows show different WRF combinations and columns show clusters. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

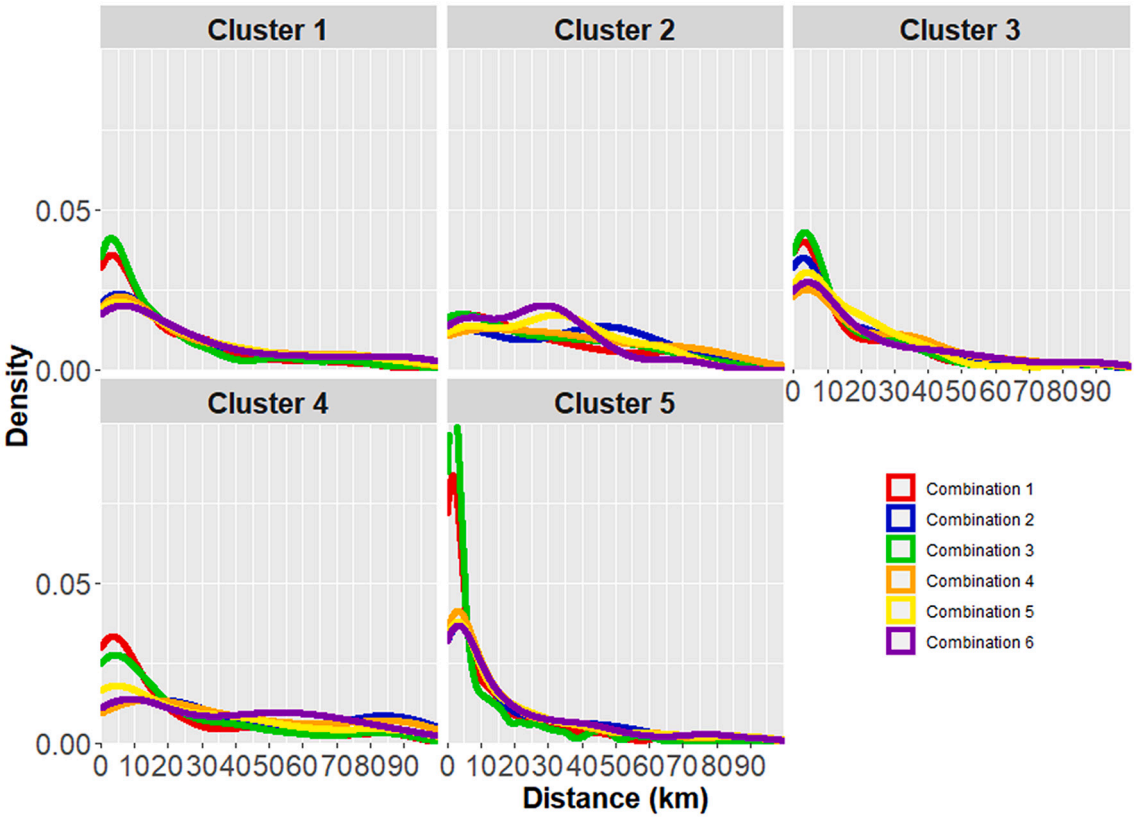

Fig. 7. Density plots for distance distributions (km) between observed and WRF precipitation, by clusters and WRF combination (colors).

Table 3
Statistical goodness-of-fit measures of observed hourly precipitation vs. WRF precipitation by different WRF combinations and clusters.

	Combinations	1	2	3	4	5	6
Cluster 1	KGE	0.223	−0.084	0.168	−0.094	−0.143	−0.139
	r	0.283	0.067	0.25	0.068	0.076	0.047
	β	0.758	0.583	0.718	0.53	0.519	0.515
	γ	1.178	1.361	1.225	1.326	1.471	1.393
Cluster 2	KGE	0.375	0.167	0.391	0.091	0.213	0.09
	r	0.415	0.235	0.446	0.198	0.299	0.153
	β	0.819	0.677	0.755	0.573	0.646	0.667
	γ	0.875	1.064	0.936	1.026	0.946	1.001
Cluster 3	KGE	0.112	0.025	0.055	−0.007	−0.052	−0.026
	r	0.21	0.156	0.181	0.147	0.168	0.119
	β	0.625	0.553	0.584	0.53	0.483	0.547
	γ	1.154	1.195	1.223	1.255	1.385	1.267
Cluster 4	KGE	0.137	−0.048	0.235	−0.093	0.041	−0.092
	r	0.186	0.104	0.274	0.082	0.136	0.078
	β	0.718	0.457	0.775	0.406	0.589	0.417
	γ	0.949	1.031	0.911	0.988	1.064	1.05
Cluster 5	KGE	0.273	0.095	0.233	0.075	0.078	0.044
	r	0.286	0.179	0.254	0.168	0.203	0.164
	β	0.866	0.696	0.853	0.672	0.691	0.684
	γ	1.024	1.232	1.102	1.237	1.345	1.339

persistent orographic precipitation, whereas the events of clusters 1 and 3 had moderate intensity and duration.

The monthly distribution of the events (Fig. 1B) shows a higher frequency in October and November. The influence of the Mediterranean Sea, which during these months reaches a high sea surface temperature, together with the arrival of baroclinic waves from the polar jet, favors extreme precipitation events. The small number of events in summer is due to convective events affecting limited areas. When stratiform precipitation is observed (cluster 5), there are more stations affected by extreme precipitation in each event (Fig. 1C).

For a deeper analysis of precipitation characteristics in each cluster, we analyzed the synoptic environment of the day with the highest frequency of stations assigned to each cluster (Fig. 3). The synoptic situation of these paradigmatic days allows further insight into the atmospheric mechanisms producing EPEs. The day selected for cluster 1 shows a dynamic low pressure located northwest of the Iberian Peninsula, favoring continuous rainfall on the southern face of the Pyrenees. The synoptic environment of cluster 2 indicates a cutoff low at mid-levels, favoring substantial moisture advection from the Mediterranean with very intense precipitation at the eastern end of the Ebro Basin. This situation favors the formation of mesoscale convective systems. The situation of cluster 3 is similar to that of cluster 1 but with the low pressure centered on the Iberian Peninsula, favoring precipitation in the northeastern Ebro Basin, owing to the interaction of moist flow with the Pyrenees. The selected day of cluster 4 is paradigmatic in terms of triggering convection in the middle Ebro Valley, with very irregular intense rainfall (García-Ortega et al., 2012). Finally, the cluster 5 day shows powerful northwest advection with persistent precipitation confined to the highlands of the northwestern Pyrenees. The synoptic analysis of these paradigmatic days confirms the rainfall regime in each cluster suggested by the previous explorations.

The 45 selected study cases encompassing the 965 stations with extreme precipitation were simulated using the WRF model. Six parameterization combinations were used to validate model behavior for extreme precipitation events (Table 1). The mean ensemble accumulated precipitation for each event (Fig. 4) shows a great variety of situations: from events with generalized precipitation throughout the study area (2016-11-23; 2016-02-27) to those affecting very small areas (2016-01-10; 2015-07-21). The latter have a diverse nature, including advective events affecting basin headwaters (2017-02-04; 2015-02-25) and localized convective events (2015-07-21; 2018-10-19). In general, the heaviest mean accumulations match the areas where extreme precipitation was recorded (black filled triangles), indicating that total accumulated precipitation was well simulated by the model. These results have been shown in previous studies (García-Ortega et al., 2017). However, we now examine the hourly structure of precipitation in these events.

The validation of WRF vs. observation hourly precipitation from point to point showed poor results, as expected. In Table 3, the dispersion between WRF and observed precipitation is clear, as can also be seen by the remoteness from the value of one for KGE and its parameters. This highlights the difficulty of simulating the hourly precipitation distribution in extreme events. Even so, important differences between the parameterization combinations and various clusters are evident. Better performances of the KGE and parameters can be seen for cluster 5 and surprisingly for cluster 2. The events classified in cluster 5 are essentially stratiform with greater spatial and temporal continuity, so a better model performance seems reasonable. However, cluster 2 presents a behavior opposite to cluster 5, with precipitation highly concentrated in time. Observing the temporal distribution of cluster 2, the events occur during autumn and the hours of greatest surface warming, when convection tends to be initiated. These events are therefore related to organized convection induced by cutoff lows in the western Mediterranean. Thus, these types of precipitating structures, despite being also convective, are better represented by the model than is isolated convection (cluster 4).

The choice of parameterizations also had a major impact on model performance. Combinations 1 and 3 yielded a better KGE for all clusters (Goddard and Thompson with MYNN PBL), whereas Morrison (combinations 5 and 6) gave the poorest results. In all cases there was model underestimation of hourly precipitation ($\beta < 1$), with combination 1 coming closest to the value of one. The variability of model precipitation remained similar to that observed ($\gamma \approx 1$).

Such poor results may be attributable to verifications based on grid-to-point comparisons. This method can produce a biased view in the case of strong spatiotemporal variability and/or discontinuous fields. The hourly precipitation verification for extreme events did not provide satisfactory results because of its strong spatiotemporal irregularity. To avoid the drawbacks that traditional indices have in verifying precipitation, alternative techniques using density plots were explored, based on three precipitation characteristics in EPEs, namely, intensity, temporal and spatial distribution.

Fig. 5 shows the density distribution of observed hourly precipitation (red) and WRF precipitation (black), together with the K-S distance and *p*-value by cluster and WRF combination. It should be emphasized that in all cases, the observed and predicted distributions have significant differences (*p*-value = 0), largely due to the tendency for WRF to underestimate. This can be seen in the rightward shift of the red curve

relative to the black one. However, the K-S distance provides valuable information for comparing WRF combinations and clusters. The largest K-S distances are associated with the stations of cluster 4, and the smallest with cluster 5. This agrees with the precipitation nature of each cluster, i.e., convective in the summer and autumn in cluster 4 and stratiform in cluster 5. However, the combination of parameterizations used in WRF has a substantial influence on the results. Thus, the best score for all clusters was obtained with combination 1, except for cluster 2, which was with combination 2. Both combinations share Goddard microphysics, obtaining performances superior to others. Regarding the PBL, MYNN achieved slightly better results than MYJ, except for cluster 2, characterized by very intense precipitation events.

Fig. 6 shows the hourly precipitation distribution by cluster and WRF combination. For EPEs, the models may not be able to accurately represent the temporal distribution of precipitation, but short delays or advances in precipitation modeling do not make the model useless. The first point is a clear smoothing effect of the modeled temporal precipitation distribution (black line). That is, the model tends to spread the precipitation over a longer time than observed. From the comparison of temporal distributions using the K-S test, different behaviors can be seen across clusters and WRF combinations. In three cases, the distributions cannot be considered different at the 0.05 significance level: cluster 5 for combination 1; cluster 3 for combination 2; cluster 2 for combination 3.

In the case of cluster 5, it is plausible that the distances between model and observed temporal distributions are smaller, because the precipitation is distributed throughout the day. For combination 1, the coincidence between distributions is especially pronounced. The strong coincidence between the distributions of cluster 2 for combination 3 is surprising, even though the precipitation is intense and short-lived. This once again suggests that the precipitation of this cluster, although convective, is associated with organized mesoscale structures. However, clusters 1 and 4, characterized by convective precipitation associated with diurnal heating, show the greatest temporal smoothing. Regarding the WRF configurations, there is no single optimal solution, because there is a dependence on precipitation behavior. Thompson microphysics performed better for convective precipitation (clusters 2 and 4) and Goddard microphysics for stratiform precipitation (clusters 3 and 5). MYNN PBL was slightly better than MYJ PBL, except for orographic precipitation (cluster 3).

Finally, we computed the distance between the maximum observed precipitation in each event and that in the WRF field. As in the temporal analysis, small spatial deviations in the WRF precipitation field does not invalidate model use. However, the spatiotemporal deviations of WRF precipitation penalize the model greatly when point-to-point comparisons are used. Fig. 7 shows distance distributions by cluster and WRF combination. For cluster 5, the highest frequency is for distances <10 km, with very few cases in which the distance between observed and model precipitation exceeding 20 km. However, in cluster 2, the distance distribution shows larger values, i.e., there is greater spatial deviation between the modeled and observed precipitation. This is unsurprising, because in stratiform rainfall (clusters 5 and 3) there is less spatiotemporal dispersion than in convective precipitation (clusters 2 and 4). Thus, rain type has an important effect on the model temporal deviations of precipitation as well as microphysics and PBL settings. Two WRF combinations achieved better performance in the spatial location of precipitation. Goddard and Thompson and MYNN PBL (Combination 1 and 3) exhibited shorter distances more frequently than the other combinations. Thompson was slightly better than Goddard, except for cluster 4. For orographic precipitation (cluster 3), the differences between combinations were smaller, but increased for convective precipitation (cluster 4).

5. Discussion and conclusions

The choice of precipitation as a measure of model performance has been questioned in the scientific literature (Tapiador et al., 2019b). The reasons derive from the drawbacks of this variable compared to other model fields, specifically, spatiotemporal discontinuity and strong variability, precipitation dependence on model parameterization, and increased propagation of error, generating large uncertainties in the reference data. However, precipitation is one of the most important fields in weather modeling, given its major implications for the environment, economic activity, energy resources, and human health.

NWP models are currently the basic tool on which precipitation forecasts are based (Sun et al., 2014). Although spatial and temporal model resolution has substantially improved in recent years, the models still struggle to accurately forecast local weather systems (such as the intensity and shape of a storm) and orographic processes (Kaufmann et al., 2003; Richard et al., 2003). This is precisely why many EPEs are not well modeled. WRF precipitation has usually been assessed on a daily scale (Douluri and Chakraborty, 2021), owing to the difficulty that models have with sub-daily precipitation. However, in terms of hydrologic risk, the same precipitation amount occurring over a very short period or several hours has very different implications. Consequently, there is a need for model validation on a sub-daily scale and for evaluation of their performance for hydrologic risk assessment.

Our study area has strong climatic variability and rainfall gradients, causing EPEs to be frequent and have diverse temporal durations and spatial extents, related to the nature of the precipitation (Merino et al., 2018). Therefore, these events were classified based on the temporal distribution of precipitation throughout an event. This information, together with the EPE monthly and spatial distribution, allowed us to infer precipitation characteristics. In summer, EPEs are convective and affect small areas. In autumn, the influence of the Mediterranean Sea favors mesoscale convective events affecting larger areas. In winter, EPEs are caused by persistent precipitation events associated with strong advection and orographic lift. In the present study, we selected 965 stations where extreme precipitation was recorded, distributed across 45 case studies.

In recent years, parameterization schemes have received great attention because they are one of the most influential factors in model performance. When model resolution increases to smaller than a 10-km grid spacing, convection processes should not be parameterized; instead, they should be explicitly resolved (García-Ortega et al., 2012; Suhas and Zhang, 2015). Moreover, there are some discrepancies between microphysics and PBL parameterization performances, because results depend on the study area, event, and method used (Duzenli et al., 2021; Karki et al., 2018). Therefore, comparison between studies is complicated.

Although the performance of multi-physics ensembles for convective precipitation events were already analyzed by García-Ortega et al. (2017), in this work we evaluated the hourly distribution of precipitation throughout an event. We verified the model by WRF grid-to-point comparisons, ruling out the use of grid precipitation because of the strong uncertainty for sub-daily precipitation.

The validation of WRF hourly precipitation vs. observation from point to point showed poor results. However, the parametrizations were relevant, with the Goddard and Thompson microphysics and MYNN PBL performing better. On the contrary, Morrison performed poorly. The events best captured by WRF were the stratiform (cluster 5) and mesoscale convective related to Mediterranean cutoff lows (cluster 2).

To avoid the strong penalty on skill scores based on point-to-point comparisons, density plots were explored, evaluating three EPE characteristics: precipitation amount distribution, temporal precipitation distribution, and the distribution of distance between maximum WRF precipitation and maximum observed. In all cases, microphysics were more important than the PBL. Goddard and Thompson and MYNN PBL gave better results for most of the analyzed characteristics. However, an optimal combination of parameterizations was not obtained for all events. Goddard was slightly superior to Thompson for the distribution of precipitation amount, temporal distribution of stratiform precipitation (clusters 3 and 5), and spatial distribution of convective precipitation (cluster 4). MYNN PBL performed better than MYF, except for the

distribution of precipitation amount in convective situations (cluster 2) and temporal distribution of orographic precipitation (cluster 3). Besides the aforementioned parameterizations, EPE characteristics influenced model performance. EPEs derived from prolonged rainfall (cluster 5) showed better performance than EPEs very localized in space and time (cluster 4). These results are consistent with those from previous studies. However, direct comparisons are not desirable because of variable methods and study areas.

The results show that models still have shortcomings in representing the temporal and spatial distributions of EPE precipitation. The development of computing capabilities have facilitated the increase of high-resolution numerical models and use of more realistic physics schemes. However, further efforts are still needed to improve both models and observations. Consequently, achieving accurate EPE forecasts will be essential for input to hydrologic models in order to reduce associated hydrologic risks.

Code and data availability

Model data are freely available from UCAR (https://www2.mmm.ucar.edu/wrf/users/download/get_sources.html). Precipitation data can be downloaded from http://www.saihebro.com/saihebro/. Scripts and code are available on request.

Author contribution statement

A.M. led the research, performed the analysis, prepared the Figures, and drafted the manuscript. All coauthors contributed to manuscript writing. F.J.T.

and E.G.O. led the project and provided the funding.

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.

Acknowledgements

Funding came from projects LE240P18 (Consejería de Educación, Junta de Castilla y León) and CGL2016-78702-C2-1-R, PID2019-108470RB-C22 and PID2019-108470RB-C21 (Ministerio de Economía y Competitividad). The authors acknowledge the Automatic Hydrological Information System (SAIH-CHEBRO) for the rain gauge data.

References

Allen, M.R., Ingram, W.J., 2002. Constraints on future changes in climate and the hydrologic cycle. Nature. 419, 224–232. https://doi.org/10.1038/nature01092.

Carrió, D.S., Jansà, A., Homar, V., Romero, R., Rigo, T., Ramis, C., Hermoso, A., Maimó, A., 2022. Exploring the benefits of a Hi-EnKF system to forecast an extreme weather event. In: The 9th October 2018 catastrophic flash flood in Mallorca. Atmospheric Research, 265, art. no. 105917. https://doi.org/10.1016/j.atmosres.2021.105917.

Chen, F., Dudhia, J., 2001. Coupling and advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: model implementation and sensitivity. Mon. Weather Rev. 129 (4), 569–585. https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2.

Cohen, A.E., Cavallo, S.M., Conigilio, M.C., Brooks, H.E., 2015. A review of planetary boundary layer parameterization schemes and their sensitivity in simulating Southeasterns U.S. cold season severe weather environments. Wea. Forecast. 30, 591–612. https://doi.org/10.1175/WAF-D-14-00105.1.

Comin, A.N., Schumacher, V., Justino, F., 2018. Impact of different microphysical parameterizations on extreme snowfall events in the Southern Andes. Weather Clim. Extrem. 21, 65–75. https://doi.org/10.1016/j.wace.2018.07.001.

Douluri, D.L., Chakraborty, A., 2021. Assessment of WRF-ARW model parameterization schemes for extreme heavy precipitation events associated with atmospheric rivers over West Coast of India. Atmos. Res. 249, 105330 https://doi.org/10.1016/j.atmosres.2020.105330.

Dudhia, J., 1989. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. J. Atmos. Sci. 46 (20), 3077–3107. https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.

Duzenli, E., Yucel, I., Pilatin, H., Yilmaz, M.T., 2021. Evaluating the performance of a WRF initial and physics ensemble over Eastern Black Sea and Mediterranean regions in Turkey. Atmos. Res. 248, 105184 https://doi.org/10.1016/j.atmosres.2020.105184.

García-Ortega, E., Fita, L., Romero, R., López, L., Ramis, C., Sánchez, J.L., 2007. Numerical simulation and sensitivity study of a severe hailstorm in Northeast Spain. Atmos. Res. 83, 225–241. https://doi.org/10.1016/j.atmosres.2005.08.004.

García-Ortega, E., Merino, A., López, L., Sánchez, J.L., 2012. Role of mesoscale factors at the onset of deep convection on hailstorm days and their relation to the synoptic patterns. Atmos. Res. 114-115, 91–106. https://doi.org/10.1016/j.atmosres.2012.05.017.

García-Ortega, E., Hermida, L., Hierro, R., Merino, A., Gascón, E., Fernández-González, S., Sánchez, J.L., López, L., 2014. Anomalies, trends and variability in atmospheric fields related to hailstorms in North-Eastern Spain. Int. J. Climatol. 34 (11), 3251–3263. https://doi.org/10.1002/joc.3910.

García-Ortega, E., Lorenzana, J., Merino, A., Fernández-González, S., López, L., Sánchez, J.L., 2017. Performance of multi-physics ensembles in convective precipitation events over northeastern Spain. Atmos. Res. 190, 55–67. https://doi.org/10.1016/j.atmosres.2017.02.009.

Giorgi, F., 2006. Climate change hot-spots. Geophys. Res. Lett. 33 https://doi.org/10.1029/2006GL025734.

Gong, X., Richman, M.B., 1995. On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. J. Clim. 8, 897–931. https://doi.org/10.1175/1520-0442(1995)008<0897:OTAOCA>2.0.CO;2.

Grazzini, F., Craig, G.C., Keil, C., Antolini, G., Pavan, V., 2020. Extreme precipitation events over northern Italy. Part I: a systematic classification with machine-learning techniques. Q. J. R. Meteorol. Soc. 146, 69–85. https://doi.org/10.1002/qj.3635.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol. 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Hartigan, J.A., Wong, M.A., 1979. A K-means clustering algorithm. Appl. Stat. 28, 100–108. https://doi.org/10.2307/2346830.

Hofstra, N., Haylock, M., New, M., Jones, P., Frei, C., 2008. Comparison of six methods for the interpolation of daily, European climate data. J. Geophys. Res. 113 (21), D21110. https://doi.org/10.1029/2008JD010100.

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

Jankov, I., 2011. An evaluation of five ARW-WRF microphysics schemes using synthetic GOES imagery for an atmospheric river event affecting the California Coast. J. Hydrometeorol. 12, 618–633. https://doi.org/10.1175/2010JHM1282.1.

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

Jonkman, S.N., 2005. Global perspectives on loss of human life caused by floods. Nat. Hazards 34, 151–175. https://doi.org/10.1007/s11069-004-8891-3.

Kain, J., 2004. The Kain-Fritsch convective parameterization: an update. J. Appl. Meteorol. 43 (1), 170–181. https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2.

Karki, R., Hasson, S.U., Gerlitz, L., Talchabhadel, R., Schenk, E., Schickhoff, U., Scholten, T., Böhner, J., 2018. WRF-based simulation of an extreme precipitation event over the Central Himalayas: Atmospheric mechanisms and their representation by microphysics parameterization schemes. Atmos. Res. 214, 21–35. https://doi.org/10.1016/j.atmosres.2018.07.016.

Kaufmann, P., Schubiger, F., Binder, P., 2003. Precipitation forecasting by a mesoscale numerical weather prediction (NWP) model: eight years of experience. Hydrol. Earth Syst. Sci. 7, 812–883. https://doi.org/10.5194/hess-7-812-2003.

Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. J. Hydrodyn. 424, 264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011.

Kong, X., Wang, A., Bi, X., Sun, B., Wei, J., 2022. The Hourly Precipitation Frequencies in the Tropical-Belt Version of WRF: Sensitivity to Cumulus Parameterization and Radiation Schemes. J. Clim. 35 (1), 285–304. https://doi.org/10.1175/JCLI-D-20-0854.1.

Li, S.H., Sifton, V., Lundgren, J., McClellan, C., Gibbons, M., 2021. Extreme wind climate of the Arabian Peninsula characterized by using WRF simulation. Weather Clim. Extrem. 33, 100330 https://doi.org/10.1016/j.wace.2021.100330.

Llasat, M.C., Llasat-Botija, M., Petrucci, O., Pasqua, A.A., Rosselló, J., Vinet, F., Boissier, L., 2013. Towards a database on societal impact of Mediterranean floods within the framework of the HYMEX project. Nat. Hazards Earth Syst. Sci. 13, 1337–1350. https://doi.org/10.5194/nhess-13-1337-2013.

Marcos, J.L., Sánchez, J.L., Merino, A., Melcón, P., Mérida, G., García-Ortega, E., 2021. Spatial and temporal variability of hail falls and estimation of maximum diameter from meteorological variables. Atmos. Res. 247, 105142 https://doi.org/10.1016/j.atmosres.2020.105142.

Mastrantonas, N., Herrera-Lormendez, P., Magnusson, L., Pappenberger, F., Matschullat, J., 2021. Extreme precipitation events in the Mediterranean: Spatiotemporal characteristics and connection to large-scale atmospheric flow patterns. Int. J. Climatol. 41 (4), 2710–2728. https://doi.org/10.1002/joc.6985.

McCumber, M., Tao, W.-K., Simpson, J., Penc, R., Soong, S.-T., 1991. Compar- ison of ice-phase microphysical parameterization schemes using numeri- cal simulations of tropical convection. J. Appl. Meteorol. 30 (7), 985–1004. https://doi.org/10.1175/1520-0450-30.7.985.

Mellor, G.L., Yamada, T., 1982. Development of a turbulence closure model for geophysical fluid problems. Rev. Geophys. Space Phys. 20, 851–875 doi:0.1029/RG020i004p00851.

Merino, A., García-Ortega, E., López, L., Sánchez, J.L., Guerrero-Higueras, A.M., 2013. Synoptic environment, mesoscale configurations and forecast parameters for hailstorms in Southwestern Europe. Atmos. Res. 122, 183–198. https://doi.org/10.1016/j.atmosres.2012.10.021.

Merino, A., López, L., Sánchez, J.L., García-Ortega, E., Cattani, E., Levizzani, V., 2014. Daytime identification of summer hailstorm cells from MSG data. Nat. Hazards Earth Syst. Sci. 14 (4), 1017–1033. https://doi.org/10.5194/nhess-14-1017-2014.

Merino, A., Fernández-Vaquero, M., López, L., Fernández-González, S., Hermida, L., Sánchez, J.L., García-Ortega, E., Gascón, E., 2016. Large-scale patterns of daily precipitation extremes on the Iberian Peninsula. Int. J. Climatol. 36 (11), 3873–3891. https://doi.org/10.1002/joc.4601.

Merino, A., Fernández-González, S., García-Ortega, E., Sánchez, J.L., López, L., Gascón, E., 2018. Temporal continuity of extreme precipitation events using sub-daily precipitation: application to floods in the Ebro basin, northeastern Spain. Int. J. Climatol. 38 (4), 1877–1892. https://doi.org/10.1002/joc.5302.

Merino, A., García-Ortega, E., Navarro, A., Fernández-González, S., Tapiador, F.J., Sánchez, J.L., 2021. Evaluation of gridded rain-gauge-based precipitation datasets: impact of station density, spatial resolution, altitude gradient and climate. Int. J. Climatol. https://doi.org/10.1002/joc.7003. In press.

Mlawer, E., Taubman, S., Brown, P., Iacono, M., Clough, S., 1997. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J. Geophys. Res. 102 (D14) https://doi.org/10.1029/97JD00237, 16,663– 16,682.

Morrison, H., Thompson, G., Tatarskii, V., 2009. Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: comparison of one- and two-moment schemes. Mon. Weather Rev. 137 (3), 991–1007. https://doi.org/10.1175/2008MWR2556.1.

Moya-Álvarez, A.S., Estevan, R., Kumar, S., Flores Rojas, J.L., Ticse, J.J., Martínez-Castro, D., Silva, Y., 2020. Influence of PBL parameterization schemes in WRF_ARW model on short-range precipitation's forecasts in the complex orography of Peruvian Central Andes. Atmos. Res. 233, 104708 https://doi.org/10.1016/j.atmosres.2019.104708.

Navarro, A., García-Ortega, E., Merino, A., Sánchez, J.L., Tapiador, F.J., 2020. Orographic biases in IMERG precipitation estimates in the Ebro River basin (Spain): the effects of rain gauge density and altitude. Atmos. Res. 244, 105068 https://doi.org/10.1016/j.atmosres.2020.105068.

Richard, E., Cosma, S., Benoit, R., Binder, P., Buzzi, A., Kaufmann, P., 2003. Intercomparison of mesoscale meteorological models for precipitation forecasting. Hydrol. Earth Syst. Sci. 7, 799–811. https://doi.org/10.5194/hess-7-799-2003.

Romero, R., Ramis, C., Homar, V., 2015. On the severe convective storm of 29 October 2013 in the Balearic Islands: Observational and numerical study. Q. J. R. Meteorol. Soc. 141 (689), 1208–1222. https://doi.org/10.1002/qj.2429.

Rutledge, S., Hobbs, P., 1984. The mesoscale and microscale structure and organization of clouds and precipitation in midlatitude cyclones XII: a diagnostic modeling study of precipitation development in narrow cold-frontal rainbands. J. Atmos. Sci. 41 (20), 2949–2972. https://doi.org/10.1175/1520-0469(1984)041<2949:TMAMSA>2.0.CO;2.

Serrano-Notivoli, R., De Luis, M., Beguería, S., 2017. An R package for daily precipitation climate series reconstruction. Environ. Model. Softw. 89, 190–195. https://doi.org/10.1016/j.envsoft.2016.11.005.

Skamarock, W., Klemp, J., 2008. A time-split nonhydrostatic atmospheric model forweather research and forecasting applications. J. Comput. Phys. 227 (7), 3465–3485. https://doi.org/10.1016/j.jcp.2007.01.037.

Somos-Valenzuela, M., Manquehual-Cheuque, F., 2020. Evaluating multiple WRF configurations and forcing over the Northern Patagonian Icecap (NPI) and Baker River Basin. Atmosphere 11 (8), 815. https://doi.org/10.3390/ATMOS11080815.

Stegall, S.T., Kunkel, K.E., 2019. Simulation of daily extreme precipitation over the United States in the CMIP5 30-yr decadal prediction experiment. J. Appl. Meteorol. Climatol. 58 (4), 875–886. https://doi.org/10.1175/JAMC-D-18-0057.1.

Stensrud, D.J., 2007. Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models. Cambridge University Press, p. 459.

Suhas, E., Zhang, G.J., 2015. Evaluating convective parameterization closures using cloud-resolving model simulation of tropical deep convection. J. Geophys. Res. 120 (4), 1260–1277. https://doi.org/10.1002/2014JD022246.

Sun, C., Liang, X.Z., 2020. Improving US extreme precipitation simulation: sensitivity to physics parameterizations. Clim. Dyn. 54 (11−12), 4891–4918. https://doi.org/10.1007/s00382-020-05267-6.

Sun, J., Xue, M., Wilson, J.W., Zawadzki, I., Ballard, S.P., Onvlee-Hooimeyer, J., Joe, P., Barker, D.M., Li, P.-W., Golding, B., Xu, M., Pinto, J., 2014. Use of nwp for nowcasting convective precipitation: recent progress and challenges. Bull. Am. Meteorol. Soc. 95 (3), 409–426. https://doi.org/10.1175/BAMS-D-11-00263.1.

Tao, W.-K., Wu, D., Lang, S., Chern, J.-D., Peters-Lidard, C., Fridlind, A., Matsui, T., 2016. High-resolution NU-WRF simulations of a deep convective-precipitation system during MC3E: further improvements and comparisons between Goddard microphysics schemes and observations. J. Geophys. Res. 121 (3), 1278–1305. https://doi.org/10.1002/2015JD023986.

Tapiador, F.J., Sánchez, J.L., García-Ortega, E., 2019a. Empirical values and assumptions in the microphysics of numerical models. Atmos. Res. 215, 214–238. https://doi.org/10.1016/j.atmosres.2018.09.010.

Tapiador, F.J., Roca, R., Del Genio, A., Dewitte, B., Petersen, W., Zhang, F., 2019b. Is precipitation a good metric for model performance? Bull. Am. Meteorol. Soc. 100 (2), 223–233. https://doi.org/10.1175/BAMS-D-17-0218.1.

Thomson, G., Rasmussen, R.M., Manning, K., 2004. Explicit forecast of win- ter precipitation using an improved bulk microphysics scheme. Part I: description and sensitivity analysis. Mon. Weather Rev. 132, 519–542. https://doi.org/10.1175/1520-0493(2004)132%3c0519:EFOWPU%3e2.0.CO;2.

Thomson, G., Rasmussen, R.M., Hall, W., 2008. Explicit forecast of winter precipi- tation using an improved bulk microphysics scheme. Part II: implementation of a new snow parameterization. Mon. Weather Rev. 136 (12), 5095–5115. https://doi.org/10.1175/2008MWR2387.1.

Tramblay, Y., Somot, S., 2018. Future evolution of extreme precipitation in the Mediterranean. Clim. Chang. 151, 289–302. https://doi.org/10.1007/s10584-018-2300-5.

Trapero, L., Bech, J., Lorente, J., 2013. Numerical modelling of heavy precipitation events over Eastern Pyrenees: analysis of orographic effects. Atmos. Res. 123, 368–383. https://doi.org/10.1016/j.atmosres.2012.09.014.

Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., El Kenawy, A.M., Angulo-Martínez, M., 2009. Daily atmospheric circulation events and extreme precipitation risk in Northeast Spain: Role of the North Atlantic Oscillation, the Western Mediterranean Oscillation, and the Mediterranean Oscillation. J. Geophys. Res. 114, D08106. https://doi.org/10.1029/2008JD011492.

Vich, M., Romero, R., 2010. Multiphysics superensemble forecast applied to Mediterranean heavy precipitation situations. Nat. Hazards Earth Syst. Sci. 10 (11), 2371–2377. https://doi.org/10.5194/nhess-10-2371-2010.

Xie, S.P., Deser, C., Vecchi, G.A., Collins, M., Delworth, T.L., Hall, A., Hawkins, E., Johnson, N.C., Cassou, C., Giannini, A., Watanabe, M., 2015. Towards predictive understanding of regional climate change. Nat. Clim. Chang. 5, 921–930. https://doi.org/10.1038/nclimate2689.

Yang, Q., Yu, Z., Wei, J., Yang, C., Gu, H., Xiao, M., Shang, S., Dong, N., Gao, L., Arnault, J., Laux, P., Kunstmann, H., 2021. Performance of the WRF model in simulating intense precipitation events over the Hanjiang River Basin, China – a multi-physics ensemble approach. Atmos. Res. 248 https://doi.org/10.1016/j.atmosres.2020.105206 art. no. 105206.