Received: 19 August 2020 | Revised: 18 December 2020 | Accepted: 21 December 2020

DOI: 10.1002/joc.6985

RESEARCH ARTICLE

International Journal of Climatology RMetS

Extreme precipitation events in the Mediterranean: Spatiotemporal characteristics and connection to large-scale atmospheric flow patterns

Nikolaos Mastrantonas[1,2] | **Pedro Herrera-Lormendez**[1] | **Linus Magnusson**[2] | **Florian Pappenberger**[2] | **Jörg Matschullat**[1]

[1]Technische Universität Bergakademie Freiberg (TUBAF), Freiberg, Germany

[2]European Centre for Medium-Range Weather Forecast (ECMWF), Reading, UK

Correspondence
Nikolaos Mastrantonas, Technische Universität Bergakademie Freiberg (TUBAF), Akademiestraße 6, Freiberg 09599, Germany.
Email: nikolaos.mastrantonas@doktorand.tu-freiberg.de

Funding information
Marie Skłodowska-Curie, Grant/Award Number: 813844; European Union's Horizon 2020

Abstract

The Mediterranean region is strongly affected by extreme precipitation events (EPEs), sometimes leading to severe negative impacts on society, economy, and the environment. Understanding such natural hazards and their drivers is essential to mitigate related risks. Here, EPEs over the Mediterranean between 1979 and 2019 are analysed, using ERA5, the latest reanalysis dataset from ECMWF. EPEs are determined based on the 99th percentile of their daily distribution (P99). The different EPE characteristics are assessed, based on seasonality and spatiotemporal dependencies. To better understand their connection to large-scale atmospheric flow patterns, Empirical Orthogonal Function analysis and subsequent non-hierarchical K-means clustering are used to quantify the importance of weather regimes to EPE frequency. The analysis is performed for different variables, depicting atmospheric variability in the lower and middle troposphere. Results show a clear spatial division in EPE occurrence, with winter and autumn being the seasons of highest EPE frequency for the eastern and western Mediterranean, respectively. There is a high degree of temporal dependencies with 20% of the EPEs (median value based on all studied grid cells), occurring up to 1 week after a preceding P99 event at the same location. Local orography is a key modulator of the spatiotemporal connections and substantially enhances the probability of co-occurrence of EPEs even for distant locations. The clustering clearly demonstrates the prevalence of distinct synoptic-scale atmospheric conditions during the occurrence of EPEs for different locations within the region. Results indicate that clustering, based on a combination of sea level pressure (SLP) and geopotential height at 500 hPa (Z500), can increase the conditional probability of EPEs by more than three (3) times (median value for all grid cells) from the nominal probability of 1% for the P99 EPEs. Such strong spatiotemporal dependencies and connections to large-scale patterns can support extended-range forecasts.

KEYWORDS
extreme precipitation, large-scale/circulation patterns, Mediterranean, weather regimes

1 | INTRODUCTION

The Mediterranean lies at the crossroads of Africa, Asia, and Europe. It is a medium-scale coupled atmosphere–ocean system of unique character; a reflection of complex topography, orographic influences, and interactions with large-scale water and land bodies around the domain. As the Mediterranean extends within the Hadley and Ferrel cells of the northern hemisphere, it experiences influence of both midlatitude and tropical climate variability (Trigo *et al*., 2006).

Extreme Precipitation Events (EPEs) is one of the most frequent natural hazards that affects the domain, eventually leading to landslides, and floods (pluvial, fluvial, flash). Such events have severe negative consequences on society, environment, and economy (Jonkman, 2005; Tramblay *et al*., 2013). Given their frequency and the high vulnerability of densely populated areas, EPEs are identified as the meteorological hazard of highest negative impact for many parts of the domain (Llasat *et al*., 2010, 2013; Tramblay and Somot, 2018). Moreover, the Mediterranean is a region, where ongoing climate change is expected to have high impacts. Therefore, it is defined as a 'hot spot' to this regard (Giorgi, 2006). Some of the projected changes include increasing frequency and magnitude of EPEs (Frei *et al*., 1998; Gao *et al*., 2006; Toreti *et al*., 2013; Cardell *et al*., 2020). Recent studies have already confirmed such changes using observational data (e.g., Kostopoulou and Jones, 2005; Alexander *et al*., 2006; Vautard *et al*., 2015; Papalexiou and Montanari, 2019). These changes, together with the high vulnerability of the domain and high economic assets in many coastal areas, are expected to lead to some of the most dramatic increases globally in average annual losses due to flooding by 2050 (Hallegatte *et al*., 2013). Thus, there are large efforts and ongoing research to better understand this natural hazard and to identify ways of improving EPE predictability and increasing the resilience of the affected regions and societies (e.g., Hydrological Mediterranean Experiment—HyMeX; Drobinski *et al*., 2014). Such advances are of crucial relevance to mitigate adverse impacts and related risks.

Previous research has identified the spatiotemporal characteristics of EPEs over the Mediterranean, with most events occurring during winter half years (e.g., Houssos and Bartzokas, 2006; Lolis and Türkeş, 2016; Merino *et al*., 2016; Khodayar *et al*., 2018; Pavan *et al*., 2019; Grazzini *et al*., 2020). These results agree with the monthly accumulated precipitation over the domain, as those months record the highest precipitation amounts (Mariotti *et al*., 2002). In respect to EPEs, there is a clear seasonal differentiation between the western and eastern parts of the domain; with most events occurring during autumn and winter, respectively; results that are consistent with both observational and reanalysis datasets (Raveh-Rubin and Wernli, 2015; Cavicchia *et al*., 2018).

Many of the EPEs over the Mediterranean are closely associated with synoptic-scale atmospheric flow patterns, such as cut-off lows, troughs, warm, and cold fronts (Toreti *et al*., 2010, 2016; Merino *et al*., 2016), while others are also related to mesoscale convective systems (Rigo *et al*., 2019). Some are connected to cyclonic formations over the Mediterranean Sea (Lionello *et al*., 2006), also known as Medicanes (Cavicchia *et al*., 2014). A recent review by Dayan *et al*. (2015) provides an informative summary of atmospheric conditions inducing EPEs over the Mediterranean. Quantification of such links is crucial, as Numerical Weather Prediction (NWP) models are more skillful in predicting large-scale characteristics rather than localized EPEs, especially for extended-range forecasts (Vitart, 2014; Lavers *et al*., 2016a; Lavers *et al*., 2016b; Lavers *et al*., 2017; Lavaysse *et al*., 2018; Lavers *et al*., 2018). One way to identify EPE connections with large-scale patterns is to analyse EPE composites for various atmospheric variables and at different vertical levels. Toreti *et al*. (2016) presented such connections by analysing potential vorticity, while Merino *et al*. (2016) used a range of variables describing the dynamics of the atmosphere in the low and mid-level troposphere, an approach also used by Toreti *et al*. (2010).

Recently, progress has been achieved in connecting EPEs and precipitation regimes over the Mediterranean with climatic modes of variability, either over the specific domain, or over extended areas, covering for example, the Euro-Atlantic region. Detailed reviews about such connections is presented in Alpert *et al*. (2006), Trigo *et al*. (2006), and Xoplaki *et al*. (2012). Vicente-Serrano *et al*. (2009) analysed relationships between occurrence and magnitude of EPEs over northeast Spain and North Atlantic Oscillation (NAO; Walker and Bliss, 1932), Western Mediterranean Oscillation (WeMO; Martin-Vide and Lopez-Bustins, 2006), and Mediterranean Oscillation (MO; Conte *et al*., 1989). They found that the most extreme daily precipitation during winter is expected for negative WeMO events. Connections are stronger when analysing the impact of such oscillations

over aggregated data. NAO, for example, is found to be significantly associated with winter precipitation not only for the west Mediterranean, but also for its eastern part (Quadrelli *et al*., 2001; Türkeş and Erlat, 2005; Hoy *et al*., 2014).

Given the existing research in the area, the novelty of this work is two-fold. (i) This work applies a systematic, Mediterranean-wide, and year-round approach by identifying EPEs based on the new ERA5 dataset. (ii) This study quantifies the connection between EPEs and large-scale atmospheric flow patterns over the Mediterranean domain. The target is to subsequently use the findings for extended-range forecasting. ERA5 provides data extending up to the recent period. This allows to update the findings and to compare the results with similar studies. Moreover, physical consistency, fine spatiotemporal resolution, and completeness of the dataset (compared to rain-gauge observations that frequently have missing/unreliable data, or inconsistencies between different regions) allow analysing EPEs at fine spatial scales. The literature mainly investigates connections between EPEs and large-scale patterns by analysing composites of EPEs derived from in-situ measurements, leading to reduced spatial coverage and use of limited temporal information. Thus, by using the entire daily atmospheric variability and the ERA5 dataset, a holistic overview of such connections over the whole domain can be delivered.

In summary, we address the following research questions:

- Do EPEs over the Mediterranean dominantly occur in specific seasons?
- To which degree are EPEs spatially and temporally connected?
- Do results based on ERA5 precipitation, agree with those of previous studies using other precipitation datasets?
- How useful is weather clustering over the whole Mediterranean domain for conditioning localized EPEs?

Data and methods used for this study are described in Sections 2 and 3, respectively. Section 4 presents results and discusses the findings. Section 5 summarizes the main conclusions, points out limitations and suggests future pathways.

2 | DATA

Data from ERA5, the latest reanalysis dataset of ECMWF (Hersbach *et al*., 2020), are used for the years 1979–2019, providing a complete record. The data are generated at hourly resolution in a horizontal grid of roughly 30 km × 30 km, using 137 vertical levels from the surface up to 80 km height to resolve the atmosphere. Four atmospheric variables are used: Total precipitation, sea level pressure (SLP), temperature at 850 hPa (T850), and geopotential height at 500 hPa (Z500).

Total precipitation is analysed at a horizontal resolution of 0.25° × 0.25° to derive statistical information of EPEs at fine spatial scales. This resolution is the closest to the original one. The selected domain covers the area 29°/47°N and −8°/38°E, referred to as *Mediterranean* from now on. Precipitation data within ERA5 are calculated from short-term forecasts of hourly accumulations. To reduce the influence of possible spin-up errors of the forecast model outputs (Dee *et al*., 2011), the daily precipitation is calculated based on the accumulation of the forecast steps 7–18 for the models initiated at 18:00 UTC of the previous day, and at 06:00 UTC of the target day. This method results in a spin-up time of 6 hr before incorporating any of the model outputs into the analysis. Although ERA5 provides modelled data, this dataset has already been used for extreme precipitation-related studies over the Mediterranean (De Luca *et al*., 2020). Appendix A provides the analysis of connecting localized EPEs to large-scale patterns using the precipitation derived from the E-OBS dataset (Cornes *et al*., 2018), a product based on observational data. The results had no substantial differences, and, moreover, most of the identified EPEs, based on ERA5, are also identified as EPEs, based on the E-OBS dataset.

In order to understand the connection of large-scale atmospheric patterns to EPE occurrence, atmospheric variability in the lower and middle troposphere is analysed, based on SLP, T850, and Z500. These variables were selected since previous work demonstrated their importance in defining synoptic environments connected to EPEs (e.g., Xoplaki *et al*., 2004; Hidalgo-Muñoz *et al*., 2011; Catto *et al*., 2014; Greco *et al*., 2020). The selected spatial resolution for these variables is 1° × 1°. Mean daily values were calculated by averaging all 24 available hourly data, and the spatial domain used for deriving the data is 26°/50°N and −11°/41°E. Thus, the analysis captures the influence of adjacent areas, for example, the Atlantic Ocean and the Alps. During initial exploratory analysis, domains of larger spatial extend were analysed (e.g., Euro-Atlantic domain, Mediterranean further extended to the Atlantic), resulting in connections between EPEs and large-scale patterns being weaker. This corroborates previous studies, demonstrating that larger domains are not efficient to optimize the relationship between circulation types and precipitation (Beck *et al*., 2016).

EPEs are highly affected by humidity-related variables (e.g., water vapour flux, specific humidity, and equivalent potential temperature), and previous research

has already confirmed their importance over Mediterranean (Hertig and Jacobeit, 2013; Hertig *et al.*, 2014). Nevertheless, these variables have complex spatiotemporal structure. This, in combination with the high spatial coverage of the domain, would result in challenges when implementing Empirical Orthogonal Function (EOF) analysis (necessity for many components to explain a substantial amount of the total variance) and subsequent K-means clustering (described in the Methodology section). Moreover, such variables have lower predictability at extended-range forecasts compared to the three selected variables (SLP, T850, Z500). Thus, humidity-related variables are not considered in this study.

3 | METHODOLOGY

3.1 | EPEs and spatiotemporal analysis

EPEs are identified as instances exceeding the 99th percentile (P99) of daily distribution, considering all available data for each grid cell. EPEs defined on seasonal data percentiles are not considered, since this work aims at analysing EPEs that are closely associated with negative consequences to society. Moreover, because of significant differences in precipitation amounts and the number of dry days between various locations of the domain (especially north vs. south and highlands vs. lowlands), we computed EPEs considering all available data, including non-rainy days, instead of EPEs computed only on wet days. Percentiles of lower frequency (P95, P97) were also tested, providing no substantial differences in the conclusions for spatiotemporal dependencies and connection to large-scale patterns.

The identified events are analysed based on their seasonality, in order to quantify the importance of seasons in the occurrence of EPEs. As the data used cover the period January 1979 to December 2019, the available number of months is the same for all seasons (4×3 months $\times$ 41 years).

The degree of temporal dependence is crucial, since multiple EPEs within short temporal intervals can significantly increase flood and landslide risks. Here, four different temporal intervals are analysed. At each grid cell, the EPE percentage that occurred within the selected interval from the preceding EPE at the same grid cell is calculated. The intervals are 1, 3, 7, and 15 days, for understanding EPE persistence at short, medium, and extended range temporal scales. Each of these scales yields different impacts and is associated with different meteorological and climatological drivers. Other intervals have been tested and did not lead to different conclusions.

3.2 | Large-scale patterns

EOF analysis on the covariance matrix (Wilks, 2011) is performed on the derived daily anomalies, independent for each of the selected variables. The analysis is conducted with the Python package *eofs* (Dawson, 2016). Data refer to all daily instances and not only to the temporal subset of EPEs. The daily anomalies are from now on simply referred to by their variable name. The square root of the cosine of latitude is used for weighting the data and giving equal-area weighting on the covariance matrix. The climatology of these variables is calculated with a 5-day smoothing window. To derive the anomalies, each day of the 1979–2019 period was subtracted from its corresponding daily climatology. The necessary number of modes (principal components) that explain at least 90% of the total variance, was retained to obtain a compressed dataset providing most of the available information.

Non-hierarchical K-means clustering (Hartigan and Wong, 1979) is implemented on the data projections on the retained principal components, and the Euclidean distance is used as the similarity measure. The analysis is conducted with the Python package *scikit-learn* (Pedregosa *et al.*, 2011). Clustering is performed independently for each variable, as well as for all different combinations of the three variables used. For the latter, the projections are pre-processed to show comparable units, and weights relevant to the explained variance. More specifically, standardization was implemented so that all projections have a standard deviation of 1. In the next step, each projection was multiplied with the square root of the percentage of the total variance of the corresponding variable that it explains, so that the projections are weighted based on their relative importance. The 95% uncertainty intervals of the climatological frequencies of the clusters are computed according to Lee *et al.* (2019), using cluster persistence and the effective sample size (Wilks, 2011) based on self-transition probability.

The number of clusters can be constrained and selected based on (semi)objective criteria. Yet, there is a level of subjectivity introduced (Gong and Richman, 1995). Moreover, the particular use of the derived clusters is crucial for the final selection. Given that this work aims at conditioning EPEs for further use of the results on subseasonal forecasting, a large number of clusters is not recommended, since only a broad indication of the expected synoptic-scale pattern is possible for such timescales (Neal *et al.*, 2016). On the other hand, a very small number of clusters cannot provide high connection with EPEs at the grid cells spatial scale, as each cluster will include daily patterns of substantial differences (high intra-cluster distances

amongst the samples). This trade-off problem is addressed by performing K-means clustering for 7 up to 12 clusters, and finally by selecting the preferred number of clusters based on their connections to EPEs (explained below).

3.3 | Connection between large-scale patterns and EPEs

The importance of each cluster in EPE occurrence was assessed, based on the conditional probability of EPEs for each cluster. In this context, we refer to the largest conditional probability over all clusters (for each combination of used variables and selected number of clusters) as the Maximum Conditional Probability (MCP). Conditional probability is highly beneficial for this study since it considers any potential differences in the climatological frequency of the identified regimes. Moreover, it can be directly compared with the nominal probability of the selected EPEs (meaning 1%). To obtain a more holistic understanding of the classification benefits for the different variables and the number of clusters, additional indicators were analysed, namely the EPE percentage allocated to each cluster (e.g., as in Yiou and Nogaj, 2004), and the percentage of grid cells that exhibit statistically significant connections with at least one cluster. Statistically significant connections between EPEs and clusters are assessed with the two-tailed 95% confidence interval of binomial distribution (e.g., Olmo *et al*., 2020). Although the interest of this work lies on positive relationships between EPEs and clusters, two-tailed test is preferred over one-tailed test, as the latter has higher type 1 errors. The probability of occurrence that is introduced for the significance testing is the upper 95th confidence interval of the relevant climatological frequency, so that strict criteria are used due to the inherent uncertainties associated with clustering.

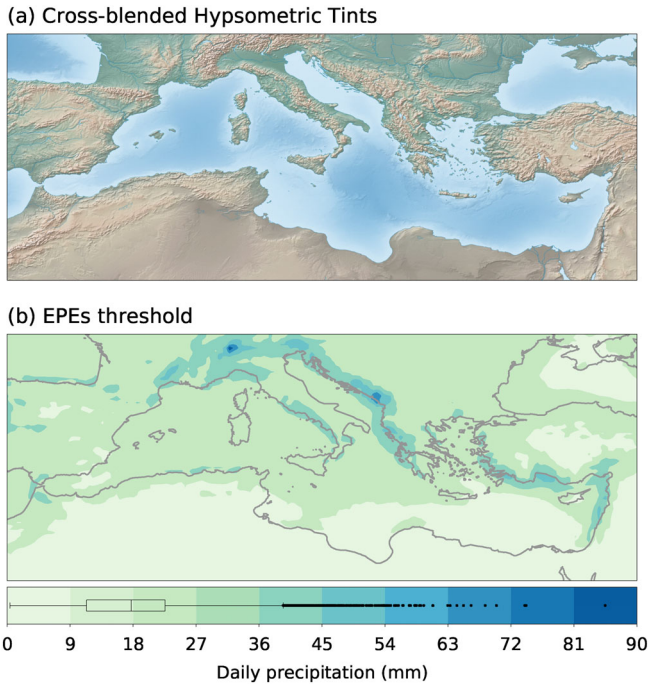

FIGURE 1 Orography (a), and P99 (b) EPEs threshold for the study area. The orography subplot (a) uses multiple colour scales to distinguish between elevation and climatic differences (humid areas vs arid areas). More information about the colour scales can be found at www.naturalearthdata.com, where the data were downloaded from (file name 'HYP_LR_OB_DR.tif'). The boxplot in subplot (b) presents the P99 distribution of all grid cells, indicating the median value and extending from the lower to upper quartile. The whiskers extend to the further available value up to 1.5*IQR from the lower and upper quartile, and all other values outside this range are presented as outliers [Colour figure can be viewed at wileyonlinelibrary.com]

4 | RESULTS AND DISCUSSION

4.1 | Spatiotemporal characteristics

Results indicate the strong influence of orography (Figure 1a) on precipitation intensity (e.g., Atlas Mountains, Alps, coast of west Balkans), as well as the importance of latitude, with locations closer to the sinking air masses of the Hadley cell (northern Africa) having significantly smaller thresholds than locations at northern latitudes. Figure 1b presents the derived P99 EPEs thresholds within the domain. The spatial EPE pattern, as well as their magnitudes are very similar to the study of Cavicchia *et al*. (2018), who used the E-OBS dataset with same spatial resolution, and identified wet-days 99th percentile intensities. The greater intensity differences in the southern Mediterranean compared to the rest of the domain can be attributed to the large number of dry days in that region. Since only wet days were used in the work of Cavicchia *et al*. (2018), it is expected that the derived percentile magnitudes over those areas will be larger compared to the current study that uses all daily values.

The EPE seasonality demonstrates a west/east divide, with most EPEs occurring during winter (autumn) for the eastern (western) parts of the domain (Figure 2, Figure S1). These findings agree with results from other studies using different datasets (Raveh-Rubin and Wernli, 2015; Cavicchia *et al*., 2018), and indicate that different synoptic-scale configurations are likely to generate EPEs at the different regions of the domain (Raveh-Rubin and Wernli, 2015). More than 70% of the EPEs in

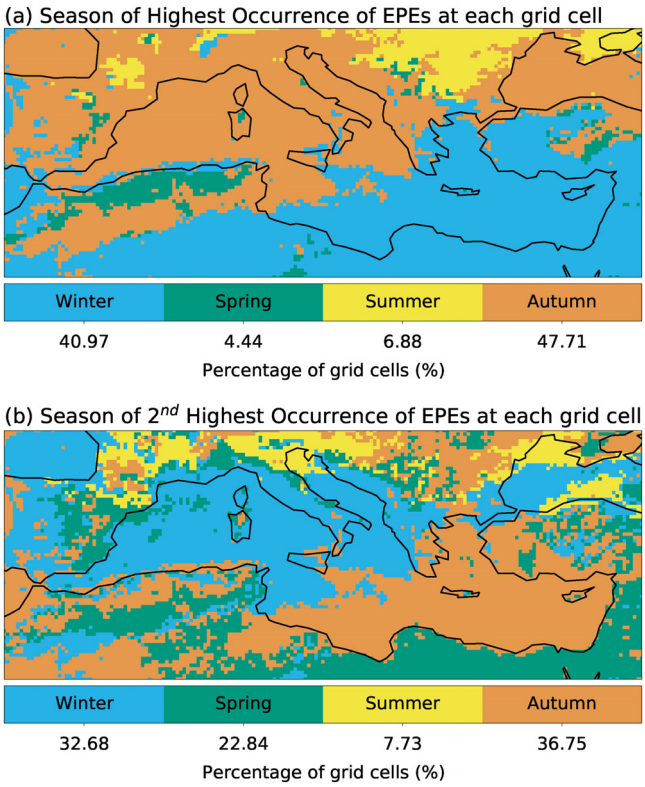

FIGURE 2 Season of highest (a) and second highest (b) occurrence of P99 EPEs per grid cell [Colour figure can be viewed at wileyonlinelibrary.com]

parts of southeast Mediterranean occur during winter, whereas over 60% of events in parts of west Mediterranean, Italy, and west Balkans occur during autumn. These two seasons interchange between first and second place in terms of EPE occurrence for most of the domain. The north Balkans and southeast Europe exhibit a different pattern, with summer and spring being the two seasons of highest EPE occurrence for most of the area. This indicates that the Mediterranean Sea has less influence on those areas, where EPEs are likely associated with different large-scale patterns. This can also be explained by orography, with mountain ranges over the south Balkans minimizing direct interactions between the Mediterranean Sea and north Balkans/southeast Europe. Spring and summer are important seasons for mountainous locations within the domain (e.g., Alps, Pyrenees). The above can be attributed to EPEs of convective nature that occur in such areas (Romero *et al*., 2001). These events, resulting from thermal low pressure systems over the region (Campins *et al*., 2000), are further enhanced by orography. Because of the generally small spatial scale of such convective events, this was not explored in the study of Raveh-Rubin and Wernli (2015), who used ERA-Interim with 1° spatial resolution and additionally implemented spatiotemporal smoothing. This shows that recent advances in the resolution of reanalysis data can bring additional benefits and more realistic information even for finer scales, due to increased computational efficiency and computation power. Finally, spring is important for north Africa and is the season of the highest/second highest occurrence of EPEs for most of this area.

Temporal EPE dependence is strong even for the 1-day interval, especially so for the dry parts of the domain in northeast Africa and adjacent parts of the Mediterranean Sea (Figure 3). Orography enhances such dependencies, with locations in Atlas Mountains, Alps, and west Balkans coast, differentiating from their neighbouring locations in many of the presented results. This can be attributed to orographic lifting and forced convection that occurs on the windward side of the mountains. These processes can trigger EPEs, even when large-scale systems are in distant areas, as long as moisture advection is directed towards the mountains (Pfahl, 2014). Thus, for 1- and 3-day intervals, associated with eastward propagation of synoptic-scale weather, these regions strongly differ from surrounding locations. The percentage of events that occur within 1 day of a preceding P99 EPE is about 11% (median value from all grid cells). This value increases to about 21% for 7 days interval, and almost 30% for 15 days interval. Such strong dependencies suggest persistent meteorological (e.g., troughs, cut-off lows, storm tracks, cyclones) and climatological (e.g., weather regimes; Vautard, 1990; Barnes and Hartmann, 2010) conditions.

4.2 | Connection to large-scale patterns

Before presenting EPE connections to large-scale patterns, it is worth commenting on the EOFs for the three analysed atmospheric variables. The number of EOFs explaining at least 90% of the total variability for SLP, T850, and Z500 are 6, 12, and 7, respectively. These EOFs are derived based on the entire dataset (all days; not only the P99 temporal subset). We also performed the same analysis for some humidity-related variables (water vapour flux, specific humidity at 850 hPa, and equivalent potential temperature at 850 hPa), with all of them having a substantially larger number of EOFs (49, 115, 42, respectively). This finding supports the decision not to use such variables in this work.

Figure S2 presents the EOFs for SLP (subplots a–f) and Z500 (subplots g–m). Their combination provides the best results for subsequent clustering (elaborated in the next paragraph). These EOFs do not have any substantial difference (in patterns and percentage of explained variance) with EOFs derived by considering only the autumn and winter months (the main P99 seasons; EOFs not

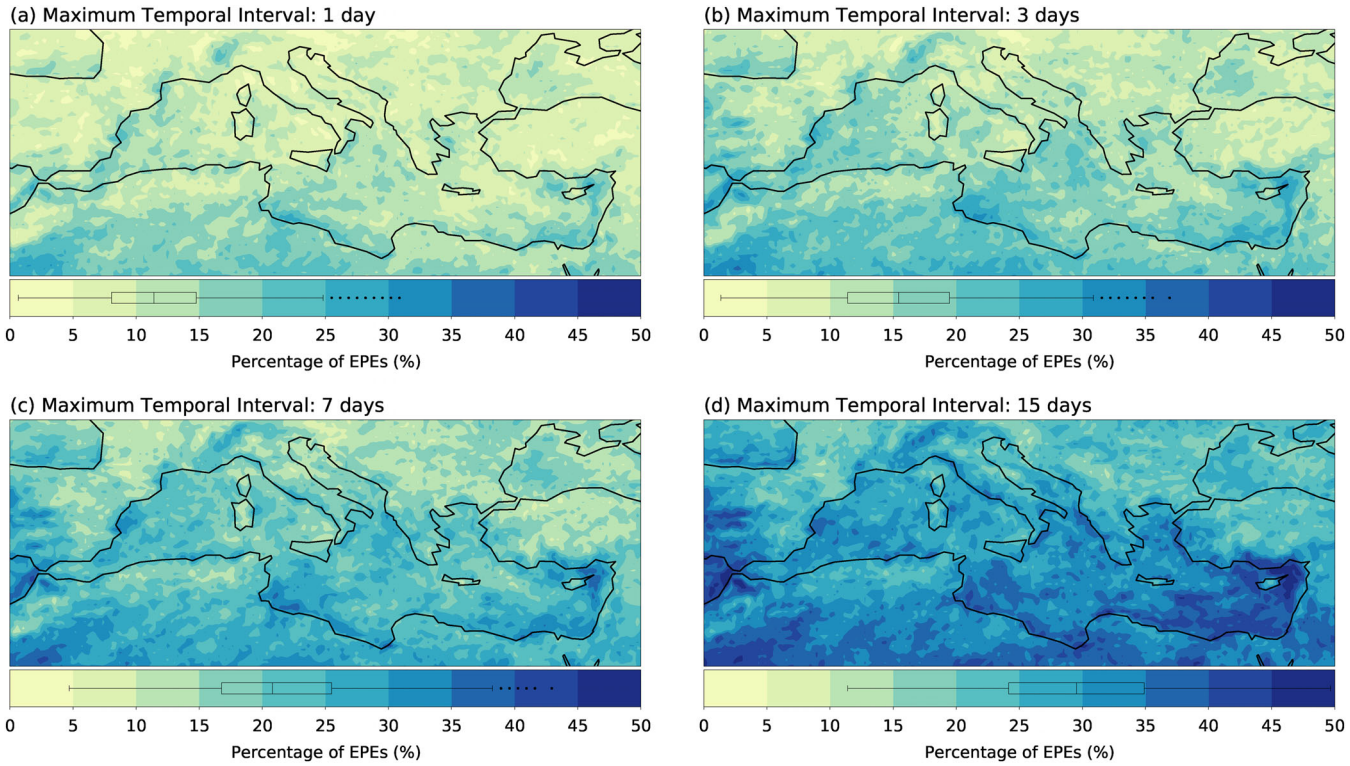

FIGURE 3 Temporal EPE dependencies. The subplots present the percentage of P99 EPEs that occur within the selected temporal intervals (a: 1 day; b: 3 days; c: 7 days; d: 15 days) from a preceding EPE at the same grid cell. The boxplots in the subplots are as in Figure 1 [Colour figure can be viewed at wileyonlinelibrary.com]

shown). The title of each subplot indicates the percentage of total variance explained by each of the EOFs (independently for each variable). EOFs are similar in spatial patterns for both SLP and Z500, with the latter being smoother, as it is not influenced by the orography. The first component explains almost half (one third) of the total variance of SLP (Z500). Its core is located over France. The pattern shows that east and west Mediterranean have opposite behaviour for Z500. The east/west divide of the domain is not only depicted at the seasonality of EPEs (presented at the previous section), but also at the EOF analysis. It is mainly noticeable in the second EOF of both variables, while the fourth (third) EOF of SLP (Z500) exhibits an omega pattern with the central parts of the domain showing opposite behaviour from the east and west subdomains. These four EOFs are associated with the Mediterranean Oscillation (MO; difference in SLP anomalies between Gibraltar and Lod, Israel; EOF1 for SLP and Z500) and WeMO (difference in SLP anomalies between San Fernando, Spain and Padova, Italy, EOF4 for SLP and EOF3 for Z500).

Figure 4 presents the results on the connection of P99 EPEs to the identified weather regimes (clusters) for all variables and number of clusters. The features used for clustering are the projections of the daily patterns to the retained EOFs for each of the analysed combinations of variables. For example, for K-means clustering based on Z500 the used features are 7, and for the combined variables (SLP, T850, Z500) the used features are 25 (6 + 12 + 7). Figure 4a presents the MCP and Figure 4b refers to the EPE percentages that occur in the cluster of MCP. Both plots show the median value from all grid cells with statistically significant connections to the generated weather regimes. Conditional probability increases for all variables with increasing number of clusters (Figure 4a). Clustering based on the combined information from SLP and Z500 outperforms all other clustering results. The other variables perform similarly, except for clustering based on T850, which is substantially worse. This is expected, as many EPEs over the Mediterranean are connected to troughs and cut-off lows, which mainly have a strong signal in geopotential height and surface pressure. Such patterns also show a signal in the temperature fields due to the frequent generation of cold and warm fronts. Yet, as these formations are of smaller spatial extent, clustering based solely on temperature fields for such a large domain is less effective compared to clustering based on the other variables.

The EPE percentage associated with the cluster of MCP decreases with a higher number of clusters. Besides

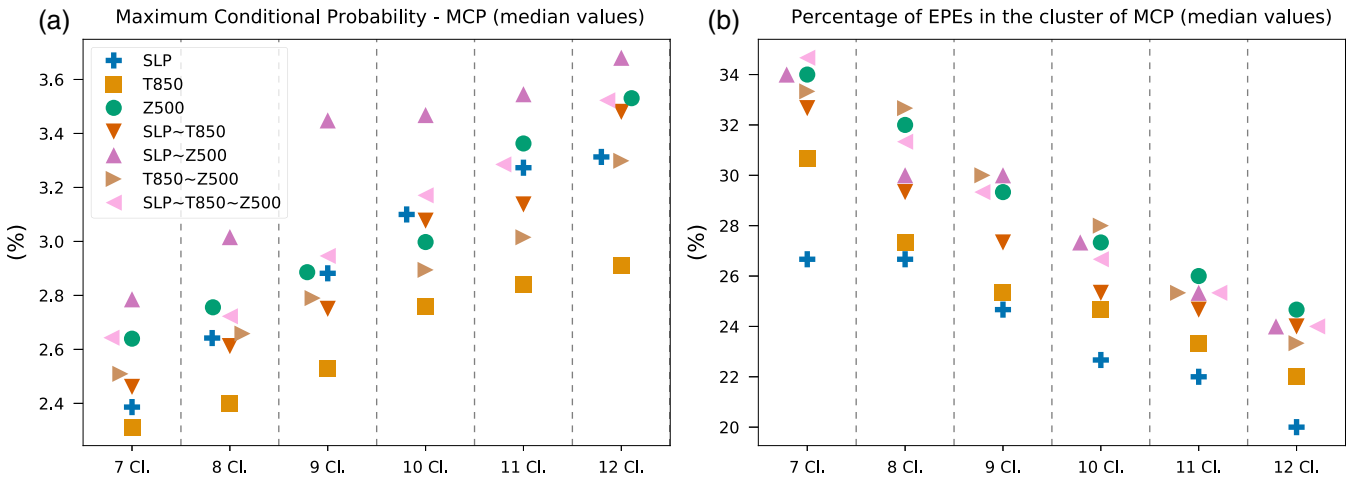

FIGURE 4 (a) Maximum Conditional Probability (MCP) of P99 EPEs for the studied variables and number of clusters. (b) Percentage of P99 EPEs associated with the cluster corresponding to MCP. For both plots, values represent the median value of all grid cells that have a statistically significant connection to the generated weather regimes [Colour figure can be viewed at wileyonlinelibrary.com]

SLP and T850, which have a weak connection, all other variables perform similarly. This tradeoff between conditional probability and percentage of EPEs is expected, as by increasing the number of clusters, the connection between EPEs and some of the derived clusters becomes stronger. At the same time, as clusters correspond to smaller number of days, the associated EPE percentages decrease. It can be noticed that for SLP-Z500 combination the MCP saturates from 9 to 12 clusters. Nine clusters perform very similar to 8 in terms of EPE percentages associated with the MCP cluster. Thus, the combination of SLP and Z500 for 9-clusters K-means clustering is selected as the preferred classifier to connect EPEs to large-scale patterns. This selection is further justified by results shown in Figure S3, presenting two additional indicators for the clusters and variables studied. These are the percentage of grid cells that have statistically significant connections with at least one cluster (subplot a), and the percentage of EPEs (median value from all grid cells with statistically significant connections) that are significantly connected with any of the clusters (subplot b). For both indicators, the order of magnitude of all variables is generally similar and results do not change substantially by changing the number of clusters. This means that the final selection of the preferred combination is not notably affected by these two indicators. Note that at least half of the total P99 EPEs are significantly associated with preferential weather regimes for most of the grid cells and for all number of clusters and selected variables (expect SLP and T850). Finally, for the selected combination of SLP-Z500 and nine clusters, the generated composites (cluster centroids, defined by averaging all data corresponding to each cluster) do not change substantially in spatial pattern and magnitude when more clusters are considered (not shown). This indicates that the clusters demonstrate coherent similarities between individual samples. Detailed results about this classifier are presented below.

Figure 5 presents the composites of the 9-classes K-means clustering, based on the combined SLP (colour shading) and Z500 (contours) fields. The derived composites have a naming convention based on the location of negative anomalies (if existing), so that the connections between clusters and EPEs are more intuitive. The composites exhibit noticeable differences in type, magnitude, or location of the large-scale patterns. The importance of the Atlantic, and the storms generated over that area is very clear. It can be noticed that Clusters 1 (Atlantic Low), 2 (Biscay Low), and 3 (Iberian Low) are associated with negative anomalies that relate to unstable conditions centred over/near the Atlantic. Cluster 4 (Sicilian Low) has a negative anomaly centred over the central Mediterranean, albeit of low magnitude. For Cluster 5 (Balkan Low), the negative anomalies are centred over (west) Balkans. Cluster 6 (Black Sea Low) has a dipole structure with positive anomalies over west Mediterranean and negative over east, while Cluster 7 (Mediterranean High) corresponds to positive anomalies and stable conditions over the whole domain. Finally, Clusters 8 (Minor Low) and 9 (Minor High) correspond to negative and positive anomalies of low magnitude over most of the domain and are associated with days that do not indicate distinct cyclonic or anticyclonic conditions of synoptic scale over the area. They can be therefore considered as the 'no-anomaly' clusters in the studied domain. Some of these clusters (e.g., Iberian Low, Balkan Low, Black Sea Low) have remarkably resemblance with the patterns identified by Toreti *et al.* (2010), who

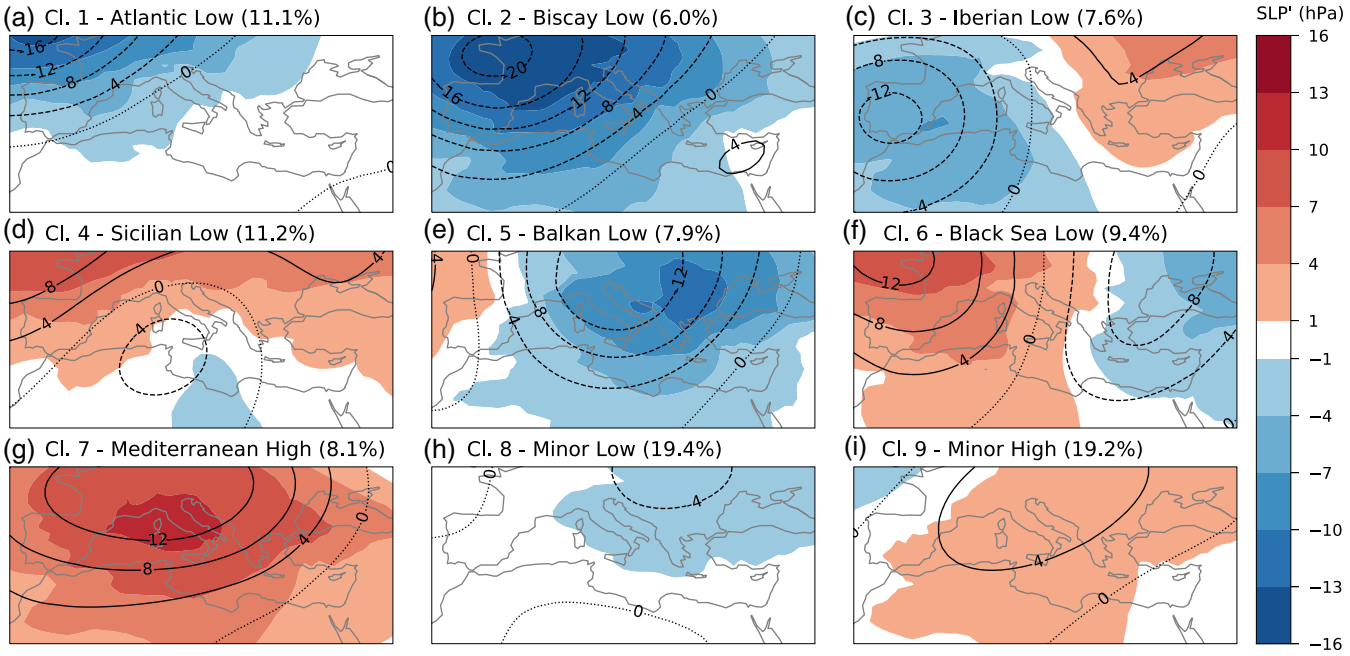

FIGURE 5 Composite of clusters derived with K-means clustering on the principal components' projections of SLP and Z500 anomalies. Colour shading refers to SLP anomalies (hPa), and contours to Z500 anomalies (dam). Percentages indicate the climatological frequency of each cluster [Colour figure can be viewed at wileyonlinelibrary.com]

calculated patterns of Z500 and SLP anomalies solely based on days with EPEs over 20 coastal stations across the Mediterranean. The clusters of the Iberian Low and Atlantic Low show many similarities with the clusters identified by Merino *et al*. (2016), while the Biscay Low is similar to one of the clusters identified by Grazzini *et al*. (2020). These two studies analysed days of EPEs over Iberian Peninsula and north Italy respectively. In these works, the composites are based on the actual fields and not on the anomalies from the daily climatological conditions.

Seasonal and annual variability in the occurrence of most clusters is very high (Figure S4). Atlantic Low and Sicilian Low have median occurrences of similar magnitude throughout the different seasons, while Minor Low, Minor High, and Mediterranean High show very high seasonal variability. The Mediterranean High is very frequent during winter. This is crucial for the occurrence of cold spells (not studied in this work), which can be generated when such conditions persist for many days (Grams *et al*., 2017; Ferranti *et al*., 2018). Biscay Low, Iberian Low, Balkan Low, Black Sea Low and Mediterranean High, with negative/positive anomalies over large parts of the domain, barely occur in summer. Atmospheric characteristics associated with these clusters are mainly driven by storm tracks and unstable conditions that prevail during winter and the intermediate seasons of autumn and spring. The high annual variability for all clusters suggests that their occurrence is modulated by climatic variability and larger-scale phenomena, for example, NAO (e.g., Dünkeloh and Jacobeit, 2003; Xoplaki *et al*., 2012).

Statistically significant clusters of highest (MCP; a) and second highest (b) conditional probability of P99 EPEs for each grid cell are presented in Figure 6. The top panels show the associated cluster. The legend indicates the percentage of grid cells allocated to each cluster. Subplots c and d present the corresponding conditional probabilities, and subplots e and f the percentage of total EPEs for each grid cell that occur during the associated cluster.

The clusters corresponding to the highest conditional probability for each grid cell show very smooth spatial behaviour, with neighbouring regions allocated mainly to the same cluster. Few discontinuities (e.g., west Italy and west Balkans vs. East Italy, west Greece and west Turkey vs. east Greece and Aegean) are mainly associated with orography (explained below). Only 0.4% of the total grid cells (located at the edge of the domain in the Middle East) cannot be significantly related to any of the derived weather regimes, while this percentage increases to 23% for the second 'best' cluster. Results for the second highest conditional probability are patchier, yet closely relate to the results of subplot a. Less than 25% of the total grid cells are significantly associated with more than two clusters (not shown), with conditional probabilities

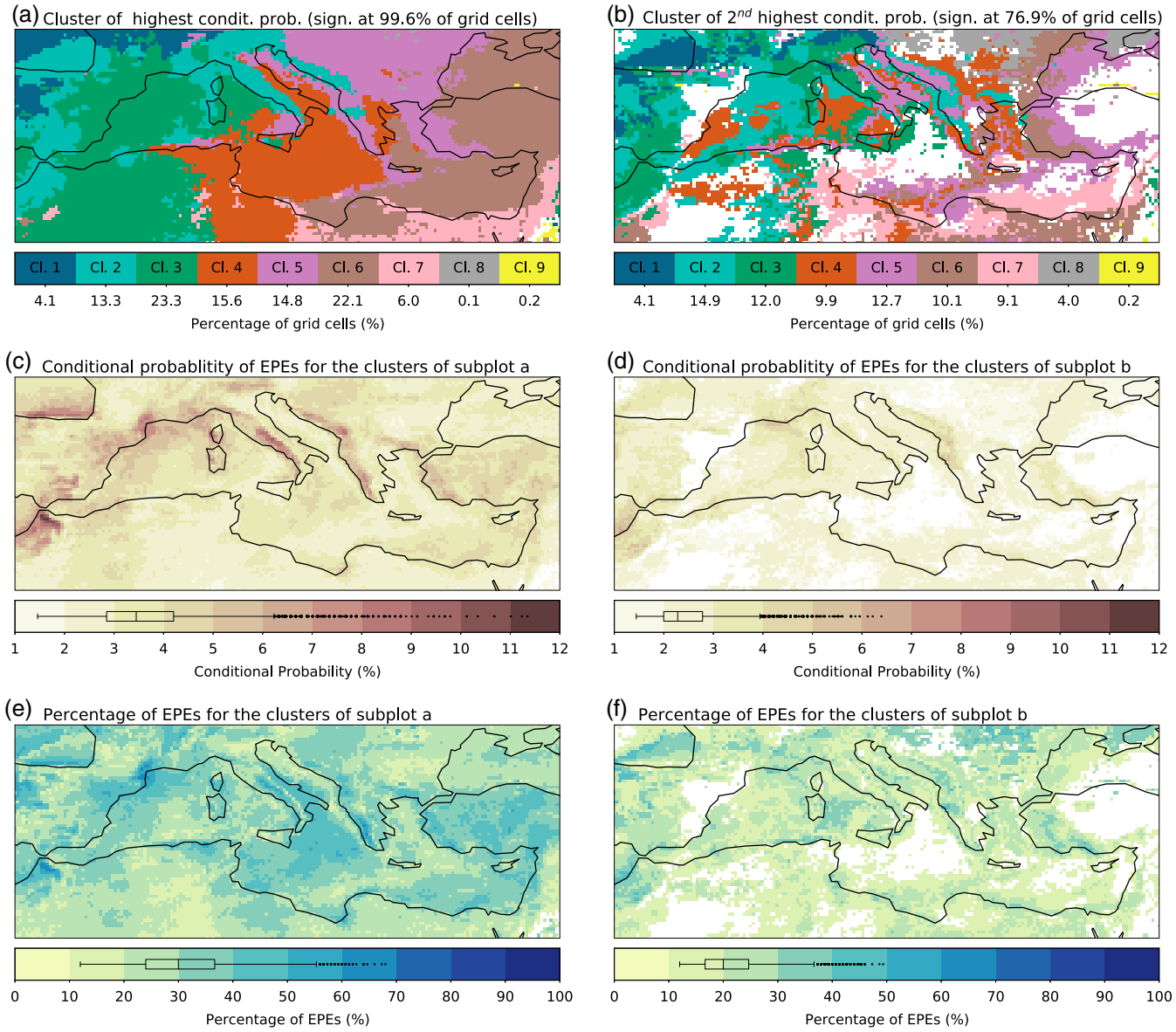

FIGURE 6 Connection of P99 EPEs and weather regimes. The subplots depict: (a) and (b) the clusters of highest and second highest occurrence probability for each grid cell, respectively; (c) and (d) the associated conditional probability of the clusters at subplots (a) and (b), respectively; (e) and (f) the associated percentage of total EPEs occurring at the clusters of subplots (a) and (b), respectively. The boxplots in subplots (c)–(f) are as in Figure 1, but only for statistically significant grid cells [Colour figure can be viewed at wileyonlinelibrary.com]

for each weather regime at the statistically significant grid cells being presented in Figure 7.

Cluster composites can physically explain connections with EPEs. Negative anomalies of pressure and geopotential height centred over the Atlantic (Atlantic Low) correspond to high chance of EPEs over the west Iberian Peninsula and southern France (Figure 7a). As these anomalies move further east (Biscay Low—Figure 7b and Iberian Low—Figure 7c), the most impacted locations are the west and west-central Mediterranean. The Biscay Low has significant connections over a large area. It is the cluster of the highest/second highest conditional probabilities for regions spanning from the Atlantic (west Mediterranean) to the west Balkans (east Mediterranean), and for both high and low latitudes. The significance of orography is demonstrated in this cluster. As cyclonic flow and existing moisture approaches the Cantabrian mountains, Atlas, Alps, Apennines, and west Balkans, windward locations are preferentially affected by EPEs. The conditional probability for EPEs over such regions is more than seven times higher from the nominal probability of 1%. Such results can be further explained by the importance of the moisture advection towards the mountains in the generation of extreme precipitation (Pfahl, 2014). Such orographic influence is also demonstrated in the association of the Sicilian Low

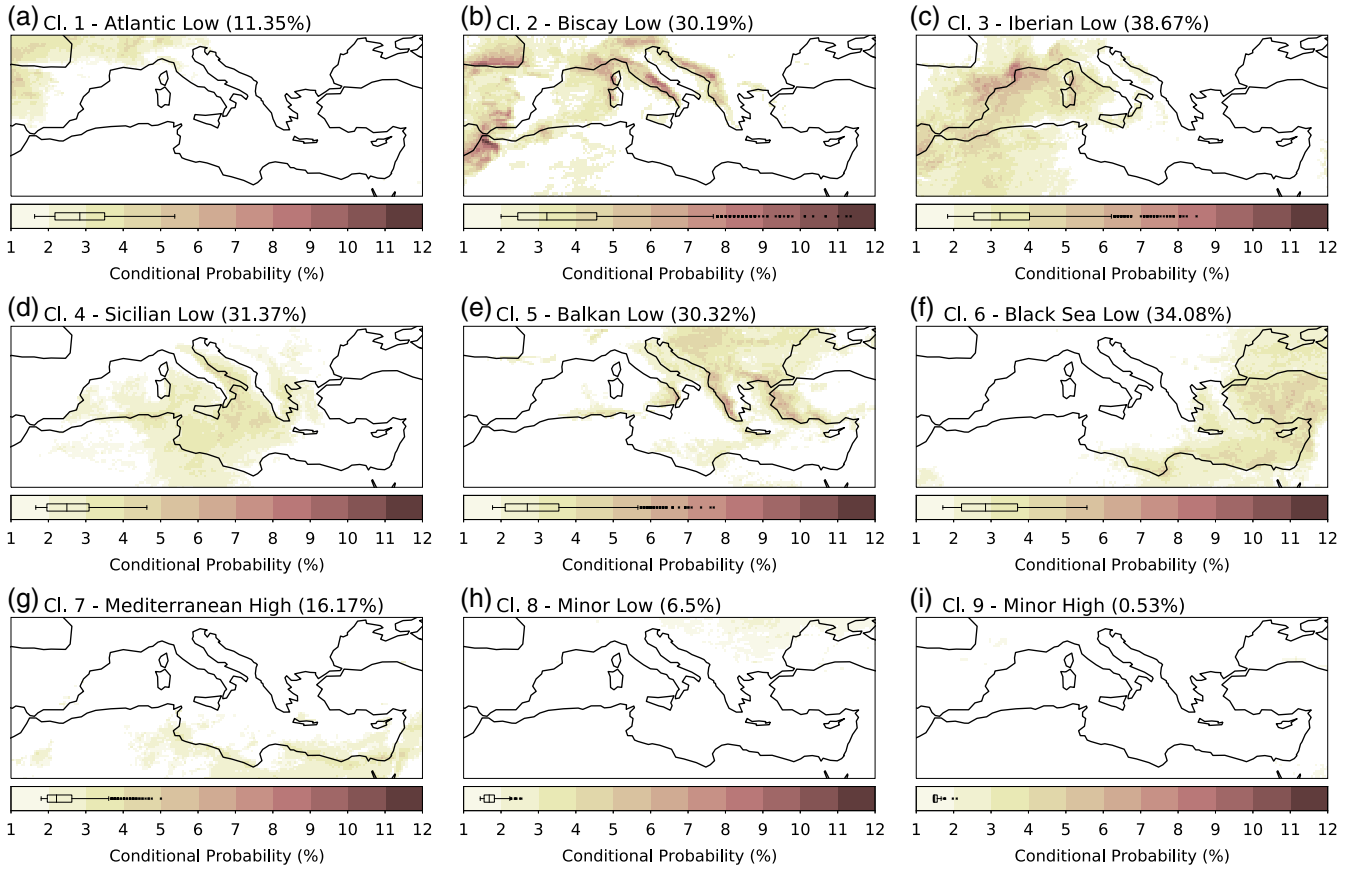

FIGURE 7 Connection of weather regimes with P99 EPEs. The conditional probability of EPEs for each weather regime is presented. Results show only areas with statistically significant connections. The percentages in the parenthesis, refer to the percentage of grid cells that have statistically significant connections with each weather regime. The boxplots are as in Figure 1, but only considering the statistically significant grid cells [Colour figure can be viewed at wileyonlinelibrary.com]

(Figure 7d) with EPEs. In this cluster, the central Mediterranean, and locations east of the Apennines and east of the mountain ranges of the west Balkans (Dinaric Alps, Pindos) are affected. Balkan Low (Figure 7e) and Black Sea Low (Figure 7f) are highly associated with EPEs over the Balkans and Turkey, respectively, while the Mediterranean High (Figure 7g), with the anticyclonic flow centred over the domain, generates EPEs over eastern north Africa and the Middle East. Minor Low (Figure 7h) and Minor High (Figure 7i) show minor association with EPEs over the domain. These results agree with previous works focusing on connection of large-scale patterns to EPEs over locations across the Mediterranean or its subdomains (Toreti *et al.*, 2010; Merino *et al.*, 2016; Grazzini *et al.*, 2020). One well-established link between EPEs and large-scale patterns over east Mediterranean is the Cyprus Low (Toreti *et al.*, 2010). However, as this work studies the connection of EPEs with large-scale patterns derived from the whole temporal dataset over the Mediterranean domain, the pattern associated with the Cyprus Low is of minor importance. Thus, this pattern is not identified based on the performed clustering. This is further corroborated by the EOF analysis of Z500, which indicates only at the fourth EOF (explaining 10.32% of total variance) a coherent structure resembling the Cyprus Low (Figure S2j).

The conditional probabilities associated with the MCP cluster are almost 50% higher compared to the second cluster in general, with the latter still showing twice the probabilities compared to the nominal of 1% for most of the grid cells. Subplots e (f) show that for most grid cells at least 30% (20%) of the total EPEs occur during the cluster of MCP (second MCP). This indicates that such classification does not only significantly increase the conditional probabilities but can also explain a large amount of the total EPEs, even when using only two of the nine clusters (around 50% for most of the grid cells). The remaining EPEs are not allocated to large-scale patterns that indicate substantial instabilities (negative anomalies of SLP/Z500) over the affected regions. This suggests that the main drivers of these events are associated with mesoscale and/or localized atmospheric phenomena, which cannot notably influence the atmospheric pattern over the whole studied domain.

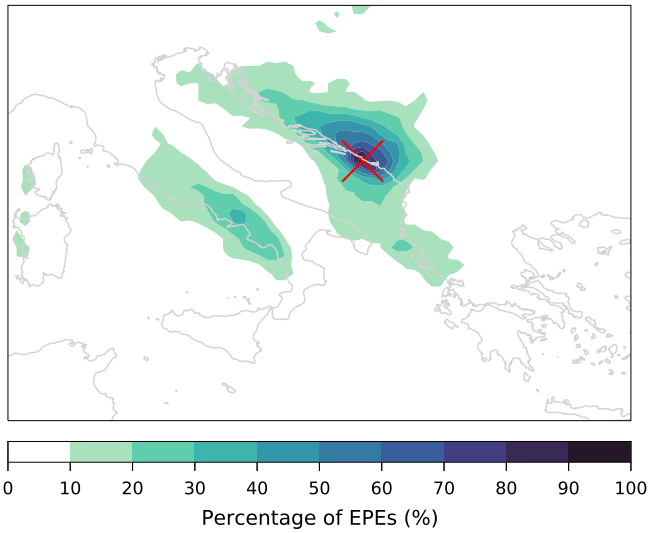

FIGURE 8 Percentage of co-occurrence of P99 EPEs between the reference grid cell at the coast of Croatia/Montenegro (Lat:42.50—Lon:18.25; cross) and the rest of the presented locations [Colour figure can be viewed at wileyonlinelibrary.com]

For quantifying the connection between lower-threshold EPEs and the derived large-scale patterns, Figure S5 presents some of the results presented in Figure 6 (only the subplots related to clusters of MCP), for EPEs based on P95, P97, and P99. It can be noticed that the conclusions do not change, indicating that the identified 9 weather clusters are strongly related to the occurrence of EPEs over the studied area. As the EPEs threshold becomes less severe (e.g., P97 compared to P99), the median value and upper tail of the MCP (second row plots) and associated percentage of EPEs (third row plots) decreases (not substantially though). This is because, by reducing the EPEs threshold, the importance of more localized phenomena in generating EPEs is increasing.

Coming back to the results of clusters of highest conditional probability (Figure 6a), a very interesting connection between western Italy and the west Balkan coast is demonstrated. Not only do these regions correspond to the same cluster, but there is also a high degree of temporal overlap (Figure 8). More than 30% of EPEs over parts of west Italy occur on the same day as EPEs over the coast of Croatia and Montenegro (and nearby areas). This is the effect of the Apennines that block the westerly airflow and force the moisture to precipitate in west Italy. Thus, a strong connection in the co-occurrence of EPEs between west Italy and west Balkans is established, whereas east Italy and the Adriatic Sea have a clear dissociation.

Figure S6 presents some additional locations, demonstrating the impact of orography on the level of temporal connectivity among nearby/distant areas for various parts of the Mediterranean. For example, subplots d and f present the importance of orography in smaller regions of high complexity. As the (generally) westerly flow intercepts with high mountains (e.g., Alps and Troodos) the rain-shadow areas (e.g., Po river valley and northeast Cyprus) have notably lower connectivity with the surrounding locations. Subplot a illustrates how the Cantabrian mountains, stretching along the coast of north Spain, modulate the regions of high connectivity, whereas for subplot b the area of high overlap takes the shape of the high orography of the Balkan Peninsula. It can be concluded that orography is a key modulator of the spatiotemporal connections between locations.

5 | CONCLUSIONS

Spatiotemporal characteristics of daily EPEs (defined based on the 99th percentile, P99) over the Mediterranean are analysed using the ERA5 dataset, and their connection to large-scale atmospheric flow patterns is quantified. The patterns are generated with EOF analysis and subsequent K-means clustering of atmospheric variability in the lower and middle troposphere.

Results indicate that the Mediterranean region is divided into two domains in terms of EPE seasonality, with autumn being the prevailing season for the western domain, and winter for the eastern one. This is consistent with previous studies in the domain (e.g., Raveh-Rubin and Wernli, 2015; Cavicchia *et al.*, 2018). Spatiotemporal connections between EPEs are very strong and are furthermore modulated by orography and regional climate. For most of the analysed grid cells, at least 20% (30%) of their EPEs occur up to 7 (15) days after a preceding EPE at the same grid cell. These high values suggest persistent meteorological and climatological conditions that favour the generation of EPEs. Such connections are even stronger at high-altitude locations (e.g., Alps and Atlas Mountains), as well as regions characterized by their dry climatic conditions (north Africa and Middle East). West Italy and the west Balkans demonstrate a remarkable temporal EPE connection with more than 30% of EPEs occurring on the same day; an effect of the Apennine Mountains.

Clustering, based on selected atmospheric variables, demonstrates that a combination of SLP and Z500 anomalies leads to the highest association of EPEs with the derived clusters. 9-class K-means clustering is the preferred classifier. The importance of these two atmospheric variables and the derived spatial patterns of the identified clusters are in accordance with previous studies on the domain (Toreti *et al.*, 2010; Merino *et al.*, 2016; Grazzini *et al.*, 2020). The clusters correspond to negative or positive anomalies over west, central, or east Mediterranean, and can be associated with cyclonic and

anticyclonic conditions on the synoptic scale. These clusters show a clear connection with the observed EPEs over most of the studied domain; a relationship that can be explained by atmospheric dynamics associated with each cluster. Conditional EPE probability increases more than three times for most locations, compared to the nominal probability of 1% for the studied EPEs. Additionally, more than 30% of the P99 EPEs preferentially occur during the cluster of highest conditional probability for most grid cells. Orography further enhances these relationships. Locations windward of the Apennine Mountains, Cantabrian Mountains, and Atlas Mountains, show MCP of P99 EPEs over 7%, and more than 40% of the EPEs occur during the associated cluster.

This study delivers additional benefits when analysing EPEs predictability at extended-range forecasts. It has been demonstrated that many EPEs can be associated to large-scale atmospheric flow patterns. This is important, since NWP models are more skillful in predicting such large-scale patterns compared to localized EPEs over extended-range forecasts (Vitart, 2014). Nevertheless, the optimal selection of weather clustering for applications targeting extended-range forecasts does also depend on the performance of NWP models for the selected lead time. For example, a combination of different atmospheric variables, that has less connection to EPEs but is more skillful compared to the presented combination in this work, can provide a higher added value. At that stage, it would be useful to analyse the additional benefits of humidity-related variables in the extended-range predictability of EPEs. The use of such variables can be considered as a second step on a forecast product, where humidity-related information can further specify potential locations of high risk of the derived daily pattern. These points will be studied in future work.

This work uses ERA5 reanalysis data to delineate EPEs. ERA5 is produced using a high-resolution model that incorporates state of the art physics and assimilates multiple observations to calculate the atmospheric variables. Despite challenges of reanalysis data (e.g., limitations in Hersbach *et al.*, 2020), our approach provides the opportunity to use a physically-consistent dataset for the performed analysis. Nevertheless, as precipitation is a complex phenomenon, highly varying across spatiotemporal scales, it would be useful to perform this analysis utilizing observational data and assess whether EPEs of finer spatiotemporal resolutions have similar characteristics.

ACKNOWLEDGEMENTS

This work is part of the Climate Advanced Forecasting of sub-seasonal Extremes (CAFE) project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813844. The authors would like to thank David Richardson and three anonymous reviewers for their fruitful suggestions that improved the quality of this work. The ERA5 data used are available on the ECMWF MARS server. The authors acknowledge the E-OBS dataset from the EU-FP6 project UERRA (https://www.uerra.eu) and the Copernicus Climate Change Service, and the data providers in the ECA&D project (https://www.ecad.eu). Open Access funding enabled and organized by ProjektDEAL.

AUTHOR CONTRIBUTIONS

Nikolaos Mastrantonas conducted literature review, data analysis and wrote the initial manuscript. Linus Magnusson, Florian Pappenberger, and Jörg Matschullat followed the work and reviewed earlier versions. Pedro Herrera-Lormendez contributed with inputs on physical mechanisms relating large-scale patterns to extreme precipitation. All authors contributed to discussions throughout the development of this study and improved the initial manuscript. All authors have read and approved the content of the final manuscript.

CODE AVAILABILITY

The scripts used for this study are publicaly available at: https://github.com/ecmwf-lab/med-extreme-prec-atm-patterns

ORCID

Nikolaos Mastrantonas https://orcid.org/0000-0002-2430-3634
Pedro Herrera-Lormendez https://orcid.org/0000-0003-0982-0032
Linus Magnusson https://orcid.org/0000-0003-4707-2231
Florian Pappenberger https://orcid.org/0000-0003-1766-2898
Jörg Matschullat https://orcid.org/0000-0003-0549-7354

REFERENCE

Alexander, L.V., Zhang, X., Peterson, T.C., Caesar, J., Gleason, B., Tank, A.M.G.K., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., Tagipour, A., Kumar, K.R., Revadekar, J., Griffiths, G., Vincent, L., Stephenson, D.B., Burn, J., Aguilar, E., Brunet, M., Taylor, M., New, M., Zhai, P., Rusticucci, M. and Vazquez-Aguirre, J.L. (2006) Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research Atmospheres*, 111. https://doi.org/10.1029/2005JD006290.

Alpert, P., Baldi, M., Ilani, R., Krichak, S., Price, C., Rodó, X., Saaroni, H., Ziv, B., Kishcha, P., Barkan, J., Mariotti, A. and Xoplaki, E. (2006) Chapter 2. Relations between climate variability in the Mediterranean region and the tropics: ENSO,

South Asian and African monsoons, hurricanes and Saharan dust. In: Lionello, P., Malanotte-Rizzoli, P. and Boscolo, R. (Eds.) *Developments in Earth and Environmental Sciences, Mediterranean*. Amsterdam: Elsevier, pp. 149–177. https://doi.org/10.1016/S1571-9197(06)80005-4.

Barnes, E.A. and Hartmann, D.L. (2010) Dynamical feedbacks and the persistence of the NAO. *Journal of the Atmospheric Sciences*, 67, 851–865. https://doi.org/10.1175/2009JAS3193.1.

Beck, C., Philipp, A. and Streicher, F. (2016) The effect of domain size on the relationship between circulation type classifications and surface climate. *International Journal of Climatology*, 36, 2692–2709. https://doi.org/10.1002/joc.3688.

Campins, J., Genovés, A., Jansà, A., Guijarro, J.A. and Ramis, C. (2000) A catalogue and a classification of surface cyclones for the Western Mediterranean. *International Journal of Climatology*, 20, 969–984. https://doi.org/10.1002/1097-0088(200007)20:9<969::AID-JOC519>3.0.CO;2-4.

Cardell, M.F., Amengual, A., Romero, R. and Ramis, C. (2020) Future extremes of temperature and precipitation in Europe derived from a combination of dynamical and statistical approaches. *International Journal of Climatology*, 40, 1–28. https://doi.org/10.1002/joc.6490.

Catto, J.L., Nicholls, N., Jakob, C. and Shelton, K.L. (2014) Atmospheric fronts in current and future climates. *Geophysical Research Letters*, 41, 7642–7650. https://doi.org/10.1002/2014GL061943.

Cavicchia, L., Scoccimarro, E., Gualdi, S., Marson, P., Ahrens, B., Berthou, S., Conte, D., Dell'Aquila, A., Drobinski, P., Djurdjevic, V., Dubois, C., Gallardo, C., Li, L., Oddo, P., Sanna, A. and Torma, C. (2018) Mediterranean extreme precipitation: a multi-model assessment. *Climate Dynamics*, 51, 901–913. https://doi.org/10.1007/s00382-016-3245-x.

Cavicchia, L., von Storch, H. and Gualdi, S. (2014) A long-term climatology of medicanes. *Climate Dynamics*, 43, 1183–1195. https://doi.org/10.1007/s00382-013-1893-7.

Conte, M., Giuffrida, A., and Tedesco, S. (1989) The mediterranean oscillation: impact on precipitation and hydrology in Italy, pp. 121–137.

Cornes, R.C., van der Schrier, G., van den Besselaar, E.J.M. and Jones, P.D. (2018) An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123, 9391–9409. https://doi.org/10.1029/2017JD028200.

Dawson, A. (2016) eofs: a library for EOF analysis of meteorological, oceanographic, and climate data. *Journal of Open Research Software*, 4, e14. https://doi.org/10.5334/jors.122.

Dayan, U., Nissen, K. and Ulbrich, U. (2015) Review article: atmospheric conditions inducing extreme precipitation over the eastern and western Mediterranean. *Natural Hazards and Earth System Sciences*, 15, 2525–2544. https://doi.org/10.5194/nhess-15-2525-2015.

De Luca, P., Messori, G., Faranda, D., Ward, P.J. and Coumou, D. (2020) Compound warm–dry and cold–wet events over the Mediterranean. *Earth System Dynamics*, 11, 793–805. https://doi.org/10.5194/esd-11-793-2020.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F. (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597. https://doi.org/10.1002/qj.828.

Drobinski, P., Ducrocq, V., Alpert, P., Anagnostou, E., Béranger, K., Borga, M., Braud, I., Chanzy, A., Davolio, S., Delrieu, G., Estournel, C., Boubrahmi, N.F., Font, J., Grubišić, V., Gualdi, S., Homar, V., Ivančan-Picek, B., Kottmeier, C., Kotroni, V., Lagouvardos, K., Lionello, P., Llasat, M.C., Ludwig, W., Lutoff, C., Mariotti, A., Richard, E., Romero, R., Rotunno, R., Roussot, O., Ruin, I., Somot, S., Taupier-Letage, I., Tintore, J., Uijlenhoet, R. and Wernli, H. (2014) HyMeX: a 10-year multidisciplinary program on the Mediterranean water cycle. *Bulletin of the American Meteorological Society*, 95, 1063–1082. https://doi.org/10.1175/BAMS-D-12-00242.1.

Dünkeloh, A. and Jacobeit, J. (2003) Circulation dynamics of Mediterranean precipitation variability 1948–98. *International Journal of Climatology*, 23, 1843–1866. https://doi.org/10.1002/joc.973.

Ferranti, L., Magnusson, L., Vitart, F. and Richardson, D.S. (2018) How far in advance can we predict changes in large-scale flow leading to severe cold conditions over Europe? *Quarterly Journal of the Royal Meteorological Society*, 144, 1788–1802. https://doi.org/10.1002/qj.3341.

Frei, C., Schär, C., Lüthi, D. and Davies, H.C. (1998) Heavy precipitation processes in a warmer climate. *Geophysical Research Letters*, 25, 1431–1434. https://doi.org/10.1029/98GL51099.

Gao, X., Pal, J.S. and Giorgi, F. (2006) Projected changes in mean and extreme precipitation over the Mediterranean region from a high resolution double nested RCM simulation. *Geophysical Research Letters*, 33. https://doi.org/10.1029/2005GL024954.

Giorgi, F. (2006) Climate change hot-spots. *Geophysical Research Letters*, 33. https://doi.org/10.1029/2006GL025734.

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

Grams, C.M., Beerli, R., Pfenninger, S., Staffell, I. and Wernli, H. (2017) Balancing Europe's wind-power output through spatial deployment informed by weather regimes. *Nature Climate Change*, 7, 557–562. https://doi.org/10.1038/nclimate3338.

Grazzini, F., Craig, G.C., Keil, C., Antolini, G. and Pavan, V. (2020) Extreme precipitation events over northern Italy. Part I: a systematic classification with machine-learning techniques. *Quarterly Journal of the Royal Meteorological Society*, 146, 69–85. https://doi.org/10.1002/qj.3635.

Greco, A., De Luca, D.L. and Avolio, E. (2020) Heavy precipitation systems in Calabria region (Southern Italy): High-resolution observed rainfall and large-scale atmospheric pattern analysis. *Water*, 12, 1468. https://doi.org/10.3390/w12051468.

Hallegatte, S., Green, C., Nicholls, R.J. and Corfee-Morlot, J. (2013) Future flood losses in major coastal cities. *Nature Climate Change*, 3, 802–806. https://doi.org/10.1038/nclimate1979.

Hartigan, J.A. and Wong, M.A. (1979) A K-means clustering algorithm. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, 28, 100–108. https://doi.org/10.2307/2346830.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G.D., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.-N. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049. https://doi.org/10.1002/qj.3803.

Hertig, E. and Jacobeit, J. (2013) A novel approach to statistical downscaling considering nonstationarities: application to daily precipitation in the Mediterranean area. *Journal of Geophysical Research: Atmospheres*, 118, 520–533. https://doi.org/10.1002/jgrd.50112.

Hertig, E., Seubert, S., Paxian, A., Vogt, G., Paeth, H. and Jacobeit, J. (2014) Statistical modelling of extreme precipitation indices for the Mediterranean area under future climate change. *International Journal of Climatology*, 34, 1132–1156. https://doi.org/10.1002/joc.3751.

Hidalgo-Muñoz, J.M., Argüeso, D., Gámiz-Fortis, S.R., Esteban-Parra, M.J. and Castro-Díez, Y. (2011) Trends of extreme precipitation and associated synoptic patterns over the southern Iberian Peninsula. *Journal of Hydrology*, 409, 497–511. https://doi.org/10.1016/j.jhydrol.2011.08.049.

Houssos, E.E., Bartzokas, A. (2006) Extreme precipitation events in NW Greece. In: *Advances in Geosciences. Presented at the 7th Plinius Conference on Mediterranean Storms (2005) - 7th Plinius Conference on Mediterranean Storms, Crete, Greece, October 5–7, 2005*. Copernicus GmbH. pp. 91–96. https://doi.org/10.5194/adgeo-7-91-2006

Hoy, A., Schucknecht, A., Sepp, M. and Matschullat, J. (2014) Large-scale synoptic types and their impact on European precipitation. *Theoretical and Applied Climatology*, 116, 19–35. https://doi.org/10.1007/s00704-013-0897-x.

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

Khodayar, S., Kalthoff, N. and Kottmeier, C. (2018) Atmospheric conditions associated with heavy precipitation events in comparison to seasonal means in the western mediterranean region. *Climate Dynamics*, 51, 951–967. https://doi.org/10.1007/s00382-016-3058-y.

Kostopoulou, E. and Jones, P.D. (2005) Assessment of climate extremes in the Eastern Mediterranean. *Meteorology and Atmospheric Physics*, 89, 69–85. https://doi.org/10.1007/s00703-005-0122-2.

Lavaysse, C., Vogt, J., Toreti, A., Carrera, M.L. and Pappenberger, F. (2018) On the use of weather regimes to forecast meteorological drought over Europe. *Natural Hazards and Earth System Sciences*, 18, 3297–3309. https://doi.org/10.5194/nhess-18-3297-2018.

Lavers, D.A., Pappenberger, F., Richardson, D.S. and Zsoter, E. (2016a) ECMWF Extreme Forecast Index for water vapor transport: a forecast tool for atmospheric rivers and extreme precipitation. *Geophysical Research Letters*, 43, 11852–11858. https://doi.org/10.1002/2016GL071320.

Lavers, D.A., Richardson, D.S., Ramos, A.M., Zsoter, E., Pappenberger, F. and Trigo, R.M. (2018) Earlier awareness of extreme winter precipitation across the western Iberian Peninsula. *Meteorological Applications*, 25, 622–628. https://doi.org/10.1002/met.1727.

Lavers, D.A., Waliser, D.E., Ralph, F.M. and Dettinger, M.D. (2016b) Predictability of horizontal water vapor transport relative to precipitation: enhancing situational awareness for forecasting western U.S. extreme precipitation and flooding. *Geophysical Research Letters*, 43, 2275–2282. https://doi.org/10.1002/2016GL067765.

Lavers, D.A., Zsoter, E., Richardson, D.S. and Pappenberger, F. (2017) An assessment of the ECMWF extreme forecast index for water vapor transport during Boreal winter. *Weather and Forecasting*, 32, 1667–1674. https://doi.org/10.1175/WAF-D-17-0073.1.

Lee, S.H., Furtado, J.C. and Charlton-Perez, A.J. (2019) Wintertime North American weather regimes and the arctic stratospheric polar vortex. *Geophysical Research Letters*, 46, 14892–14900. https://doi.org/10.1029/2019GL085592.

Lionello, P., Bhend, J., Buzzi, A., Della-Marta, P.M., Krichak, S.O., Jansà, A., Maheras, P., Sanna, A., Trigo, I.F. and Trigo, R. (2006) Chapter 6. Cyclones in the Mediterranean region: Climatology and effects on the environment. In: Lionello, P., Malanotte-Rizzoli, P. and Boscolo, R. (Eds.) *Developments in Earth and Environmental Sciences, Mediterranean*. Amsterdam: Elsevier, pp. 325–372. https://doi.org/10.1016/S1571-9197(06)80009-1.

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

Llasat, M.C., Llasat-Botija, M., Prat, M.A., Porcú, F., Price, C., Mugnai, A., Lagouvardos, K., Kotroni, V., Katsanos, D., Michaelides, S., Yair, Y., Savvidou, K., Nicolaides, K. (2010). High-impact floods and flash floods in Mediterranean countries: the FLASH preliminary database. In: *Advances in Geosciences. Presented at the 10th EGU Plinius Conference on Mediterranean Storms (2008)— 10th Plinius Conference on Mediterranean Storms, Nicosia, Cyprus, September 22–24, 2008*. Copernicus GmbH, pp. 47–55. https://doi.org/10.5194/adgeo-23-47-2010

Lolis, C.J. and Türkeş, M. (2016) Atmospheric circulation characteristics favouring extreme precipitation in Turkey. *Climate Research*, 71, 139–153. https://doi.org/10.3354/cr01433.

Mariotti, A., Struglia, M.V., Zeng, N. and Lau, K.-M. (2002) The hydrological cycle in the Mediterranean region and implications for the water budget of the Mediterranean Sea. *Journal of Climate*, 15, 1674–1690. https://doi.org/10.1175/1520-0442(2002)015<1674:THCITM>2.0.CO;2.

Martin-Vide, J. and Lopez-Bustins, J.-A. (2006) The western Mediterranean oscillation and rainfall in the Iberian Peninsula. *International Journal of Climatology*, 26, 1455–1475. https://doi.org/10.1002/joc.1388.

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

Neal, R., Fereday, D., Crocker, R. and Comer, R.E. (2016) A flexible approach to defining weather patterns and their application in weather forecasting over Europe. *Meteorological Applications*, 23, 389–400. https://doi.org/10.1002/met.1563.

Olmo, M., Bettolli, M.L. and Rusticucci, M. (2020) Atmospheric circulation influence on temperature and precipitation individual and compound daily extreme events: spatial variability and trends over southern South America. *Weather and Climate Extremes*, 29 100267. https://doi.org/10.1016/j.wace.2020.100267.

Papalexiou, S.M. and Montanari, A. (2019) Global and regional increase of precipitation extremes under global warming. *Water Resources Research*, 55, 4901–4914. https://doi.org/10.1029/2018WR024067.

Pavan, V., Antolini, G., Barbiero, R., Berni, N., Brunier, F., Cacciamani, C., Cagnati, A., Cazzuli, O., Cicogna, A., De Luigi, C., Di Carlo, E., Francioni, M., Maraldo, L., Marigo, G., Micheletti, S., Onorato, L., Panettieri, E., Pellegrini, U., Pelosini, R., Piccinini, D., Ratto, S., Ronchi, C., Rusca, L., Sofia, S., Stelluti, M., Tomozeiu, R. and Torrigiani Malaspina, T. (2019) High resolution climate precipitation analysis for north-central Italy, 1961–2015. *Climate Dynamics*, 52, 3435–3453. https://doi.org/10.1007/s00382-018-4337-6.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, É. (2011) Scikit-learn: machine Learning in python. *Journal of Machine Learning Research*, 12, 2825–2830.

Pfahl, S. (2014) Characterising the relationship between weather extremes in Europe and synoptic circulation features. *Natural Hazards and Earth System Sciences*, 14, 1461–1475. https://doi.org/10.5194/nhess-14-1461-2014.

Quadrelli, R., Pavan, V. and Molteni, F. (2001) Wintertime variability of Mediterranean precipitation and its links with large-scale circulation anomalies. *Climate Dynamics*, 17, 457–466. https://doi.org/10.1007/s003820000121.

Raveh-Rubin, S. and Wernli, H. (2015) Large-scale wind and precipitation extremes in the Mediterranean: a climatological analysis for 1979–2012. *Quarterly Journal of the Royal Meteorological Society*, 141, 2404–2417. https://doi.org/10.1002/qj.2531.

Rigo, T., Berenguer, M. and Llasat, M.d.C. (2019) An improved analysis of mesoscale convective systems in the western Mediterranean using weather radar. *Atmospheric Research*, 227, 147–156. https://doi.org/10.1016/j.atmosres.2019.05.001.

Romero, R., Doswell, C.A. and Riosalido, R. (2001) Observations and fine-grid simulations of a convective outbreak in Northeastern Spain: importance of diurnal forcing and convective cold pools. *Monthly Weather Review*, 129, 2157–2182. https://doi.org/10.1175/1520-0493(2001)129<2157:OAFGSO>2.0.CO;2.

Schulzweida, U. (2019) CDO user guide. https://doi.org/10.5281/zenodo.3539275

Toreti, A., Giannakaki, P. and Martius, O. (2016) Precipitation extremes in the Mediterranean region and associated upper-level synoptic-scale flow structures. *Climate Dynamics*, 47, 1925–1941. https://doi.org/10.1007/s00382-015-2942-1.

Toreti, A., Naveau, P., Zampieri, M., Schindler, A., Scoccimarro, E., Xoplaki, E., Dijkstra, H.A., Gualdi, S. and Luterbacher, J. (2013) Projections of global changes in precipitation extremes from coupled model intercomparison project phase 5 models. *Geophysical Research Letters*, 40, 4887–4892. https://doi.org/10.1002/grl.50940.

Toreti, A., Xoplaki, E., Maraun, D., Kuglitsch, F.G., Wanner, H. and Luterbacher, J. (2010) Characterisation of extreme winter precipitation in Mediterranean coastal sites and associated anomalous atmospheric circulation patterns. *Natural Hazards and Earth System Sciences*, 10, 1037–1050. https://doi.org/10.5194/nhess-10-1037-2010.

Tramblay, Y., El Adlouni, S. and Servat, E. (2013) Trends and variability in extreme precipitation indices over Maghreb countries. *Natural Hazards and Earth System Sciences*, 13, 3235–3248. https://doi.org/10.5194/nhess-13-3235-2013.

Tramblay, Y. and Somot, S. (2018) Future evolution of extreme precipitation in the Mediterranean. *Climatic Change*, 151, 289–302. https://doi.org/10.1007/s10584-018-2300-5.

Trigo, R., Xoplaki, E., Zorita, E., Luterbacher, J., Krichak, S.O., Alpert, P., Jacobeit, J., Sáenz, J., Fernández, J., González-Rouco, F., Garcia-Herrera, R., Rodo, X., Brunetti, M., Nanni, T., Maugeri, M., Türke, M., Gimeno, L., Ribera, P., Brunet, M., Trigo, I.F., Crepon, M. and Mariotti, A. (2006) Chapter 3. Relations between variability in the Mediterranean region and mid-latitude variability. In: Lionello, P., Malanotte-Rizzoli, P. and Boscolo, R. (Eds.) *Developments in Earth and Environmental Sciences, Mediterranean*. Amsterdam: Elsevier, pp. 179–226. https://doi.org/10.1016/S1571-9197(06)80006-6.

Turco, M., Zollo, A.L., Ronchi, C., De Luigi, C. and Mercogliano, P. (2013) Assessing gridded observations for daily precipitation extremes in the Alps with a focus on northwest Italy. *Natural Hazards and Earth System Sciences*, 13, 1457–1468. https://doi.org/10.5194/nhess-13-1457-2013.

Türkeş, M. and Erlat, E. (2005) Climatological responses of winter precipitation in Turkey to variability of the North Atlantic Oscillation during the period 1930–2001. *Theoretical and Applied Climatology*, 81, 45–69. https://doi.org/10.1007/s00704-004-0084-1.

Vautard, R. (1990) Multiple weather regimes over the North Atlantic: analysis of precursors and successors. *Monthly Weather Review*, 118, 2056–2081. https://doi.org/10.1175/1520-0493(1990)118<2056:MWROTN>2.0.CO;2.

Vautard, R., Yiou, P., van Oldenborgh, G.-J., Lenderink, G., Thao, S., Ribes, A., Planton, S., Dubuisson, B. and Soubeyroux, J.-M. (2015) Extreme fall 2014 precipitation in the Cévennes Mountains. *Bulletin of the American Meteorological Society*, 96, S56–S60. https://doi.org/10.1175/BAMS-D-15-00088.1.

Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., El Kenawy, A.M. and 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. *Journal of Geophysical Research*, 114, D08106. https://doi.org/10.1029/2008JD011492.

Vitart, F. (2014) Evolution of ECMWF sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society*, 140, 1889–1899. https://doi.org/10.1002/qj.2256.

Walker, G.T. and Bliss, E.W. (1932) World weather V. *Memoirs of the Royal Meteorological Society*, 4, 53–84.

Wilks, D.S. (2011) *Statistical Methods in the Atmospheric Sciences*. San Diego, CA: Academic Press.

Xoplaki, E., González-Rouco, J.F., Luterbacher, J. and Wanner, H. (2004) Wet season Mediterranean precipitation variability: influence of large-scale dynamics and trends. *Climate Dynamics*, 23, 63–78. https://doi.org/10.1007/s00382-004-0422-0.

Xoplaki, E., Trigo, R.M., García-Herrera, R., Barriopedro, D., D'Andrea, F., Fischer, E.M., Gimeno, L., Gouveia, C., Hernández, E., Kuglitsch, F.G., Mariotti, A., Nieto, R., Pinto, J. G., Pozo-Vázquez, D., Saaroni, H., Toreti, A., Trigo, I.F., Vicente-Serrano, S.M., Yiou, P. and Ziv, B. (2012) Large-scale atmospheric circulation driving extreme climate events in the Mediterranean and its related impacts. In: Lionello, P. (Ed.) *The Climate of the Mediterranean Region*. Oxford: Elsevier, pp. 347–417. https://doi.org/10.1016/B978-0-12-416042-2.00006-9.

Yiou, P. and Nogaj, M. (2004) Extreme climatic events and weather regimes over the North Atlantic: when and where? Weather regimes and extremes. *Geophysical Research Letters*, 31. https://doi.org/10.1029/2003GL019119.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Mastrantonas N, Herrera-Lormendez P, Magnusson L, Pappenberger F, Matschullat J. Extreme precipitation events in the Mediterranean: Spatiotemporal characteristics and connection to large-scale atmospheric flow patterns. *Int J Climatol*. 2021;1–19. https://doi.org/10.1002/joc.6985

APPENDIX

This appendix presents information regarding the connection of the identified large-scale patterns and localized EPEs derived from E-OBS (Cornes *et al*., 2018), to complement the ERA5 results. It moreover presents a brief analysis in the percentage of ERA5 EPEs that are identified as EPEs also within E-OBS dataset.

E-OBS is a gridded dataset that provides daily precipitation at 0.10°× 0.10° resolution grid over Europe. These values are produced using interpolated observations from quality-controlled stations. To improve the precipitation estimates, the dataset uses a 100-member ensemble of realizations of each daily field (Cornes *et al*., 2018). E-OBS data for the period 1979–2019, regridded to the same grid of ERA5 using first-order conservative regridding with *cdo* tool (Schulzweida, 2019), are used in this comparison. Note that the E-OBS data provide information only over land, and moreover the results of this analysis refer to locations that have less than 5% missing data on the E-OBS dataset (all other locations are hatched in the relevant plots).

Figures A1, S7, and S8 reproduce the diagnostics of Figure 6, 7, and S5, respectively, but for EPEs based on E-OBS dataset. The results demonstrate that the findings are very similar for both datasets, a further indication that the derived connections between localized EPEs and the identified 9 patterns are robust. For most grid cells the cluster of Maximum Conditional Probability (MCP) is the same for both precipitation datasets. Moreover, as for the ERA5 EPEs, orography enhances the connections to large-scale patterns, and the median value of the MCP is more than three times higher compared to the nominal ones.

Figure A2 presents an analysis in the percentage of ERA5 EPEs that are identified as EPEs also within E-OBS dataset. The subplots present results for overlap when considering a 1-day flexibility window in the timing of EPEs from E-OBS. This is considered because the daily aggregations for E-OBS have different UTC starting time for each country, so the majority of the precipitation could have occurred in the previous, same, or next day when comparing to ERA5 data that has a consistent timing based on UTC. Findings of such temporal shift have already being identified (Turco *et al*., 2013).

Figure A2a,c,e refer to the percentage of P99, P97, and P95 events of ERA5 that overlaps with P99, P97, and P95 events of E-OBS respectively, while Figure A2b,d,f present the percentage of P99, P97, and P95 events of ERA5 that overlaps with P97, P95, and P93 events of E-OBS, respectively.

It can be noticed that the temporal agreement between the two datasets is substantial, especially when relaxing the EPEs threshold for the E-OBS data. More than 40% of the P99 EPEs of ERA5 are also identified as P99 EPEs based on the E-OBS dataset in most of the domain. This overlap is higher for P97 (50–60%) and P95 (60–70%) EPEs, while when the E-OBS threshold is relaxed by 2% (right column subplots), the degree of overlap is 70–90% for most of the domain.

The higher differences between the two datasets can be identified in locations of sparse stations, like Africa, Asia (except Israel), Greece and Sicily (Figure S9). In data sparse regions, very localized extremes can result in EPEs within E-OBS dataset, even for larger domains, as an artefact of the interpolation techniques used. Another reason that could result in reduction of the overlap, is that localized extremes of convective nature are not expected to be identified by the current grid resolution of ERA5, as it represents a grid-box average.

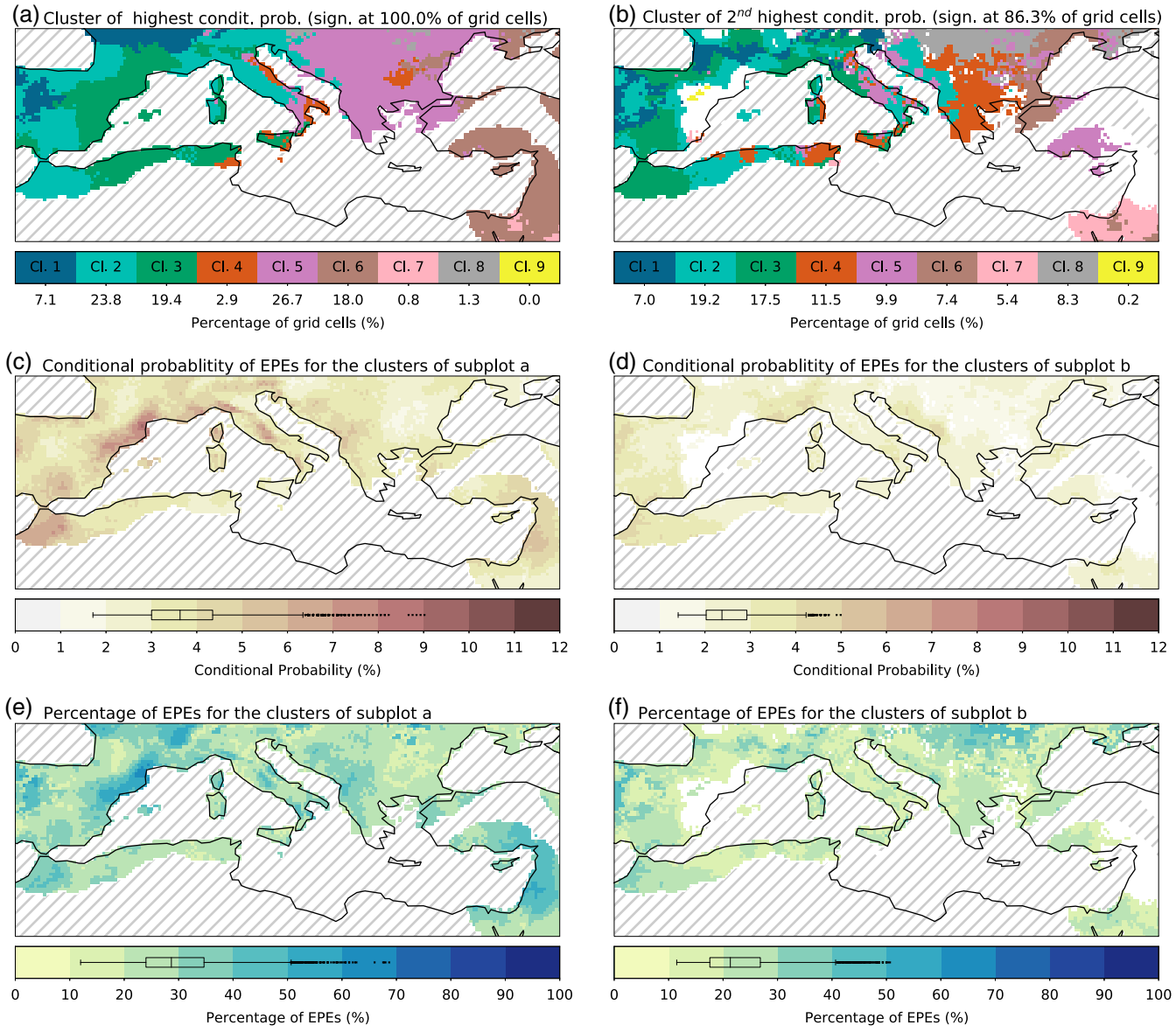

FIGURE A1 As Figure 6, but for EPEs based on EOBS dataset. The locations that have at least 5% missing data based on EOBS dataset are hatched. Note that, the colour scale for the Conditional Probabilities is the same as Fig6 with the addition of grey colour for probabilities between 0% and 1% that are identified as significant. This is valid in a very small section of the southeast part of the domain (less than 100 grid cells) and for EPEs defined based on 97th and 99th percentile (see Figure S8). For those percentiles the precipitation threshold in these areas is zero. As for these locations non-zero days are fewer than 3% (extremely dry locations according to E-OBS), the total number of EPEs is substantially lower (less than 40%) compared to the expected one. This leads to significant connections between EPEs and large-scale patterns that demonstrate conditional probabilities lower than the nominal ones [Colour figure can be viewed at wileyonlinelibrary.com]

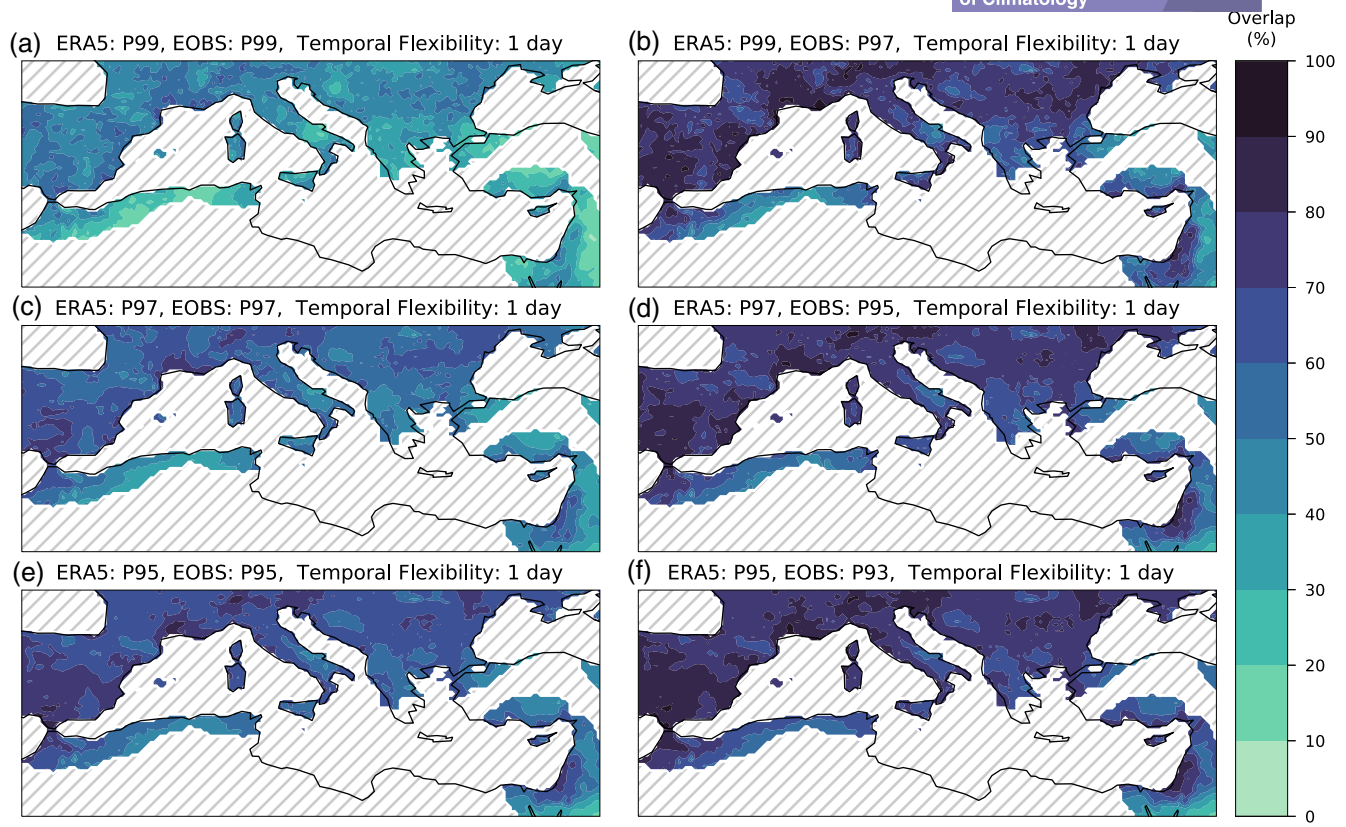

FIGURE A2 Percentage of ERA5 EPEs that are identified as EPEs also within E-OBS dataset. The locations that have at least 5% missing data based on EOBS dataset are hatched. The subplots refer to various combinations of ERA5 and EOBS percentile thresholds. For all subplots a 1-day flexibility window in the timing of EPEs from E-OBS is considered, because the daily aggregations for E-OBS have different UTC starting time for different locations. Thus, the majority of the precipitation could have occurred in the previous, same, or next day when comparing to ERA5 data that has a consistent timing based on UTC [Colour figure can be viewed at wileyonlinelibrary.com]

Given the inherent uncertainties in the identification of EPEs for aggregated areas, these results suggest that ERA5 can be considered as a reliable source of information regarding the timing of daily EPEs for the majority of the domain, especially when considering a relaxed percentile threshold from E-OBS. Albeit of high scientific interest, a more detailed analysis regarding the timing (and intensity) of EPEs for the two datasets is not in the scope of this investigation.