ARTICLE   OPEN

Check for updates

Inter-seasonal connection of typical European heatwave patterns to soil moisture

Elizaveta Felsche[1,2,3 ✉], Andrea Böhnisch [2] and Ralf Ludwig [2]

Although prolonged heat periods have become a recurring feature of European climate, little knowledge is available on dominant spatial patterns of heatwaves and their influence on moisture-related processes. Increased knowledge will help to improve heatwave and drought prediction and mitigation. This study uses hierarchical agglomerative clustering to derive nine dominating spatial heatwave patterns from a 50-member regional climate model (Canadian Regional Climate Model version 5, CRCM5-LE). The heatwave patterns correspond well with clusters derived from an observational data set (E-OBS) and with extreme historical heatwave events. Moreover, we analyse the occurrence of heatwaves in the identified spatial patterns regarding a soil moisture deficit present before and after the event. We show that negative soil moisture anomalies in the preceding winter/spring (JFMA) can serve as a predictor for heatwaves in South Europe. For North Europe, we find a negative correlation between the number of heatwave days in summer and autumn (OND) soil moisture content.

INTRODUCTION

Heatwaves and droughts substantially impact human mortality, economic well-being, infrastructure, and natural ecosystems[1–3]. For example, the 2003 heatwave in Europe is estimated to have caused more than 70,000 deaths[4]. Globally, 2% of working hours are lost due to too-hot conditions[5]. Droughts that often accompany a heatwave have been estimated to cause losses of USD 621 Million on average per event between 1950 and 2014 in Europe[6]. The 2010 heatwave in Russia caused USD 15 Billion (1% gross domestic product) in total economic losses[7]. Since 1950, most regions worldwide have observed a significant increase in the number of heatwave days, maximum duration, and cumulative heat[8]. Climate model projections estimate that the described trends will continue throughout the 21st century[9,10]. There is a high need for operational seasonal forecasts of heatwaves and droughts to mitigate their impacts, e.g., to introduce measures for water saving or prepare navigation infrastructure for low flows[11,12]. Current forecasts offer limited predictive capabilities, underlining the importance of future studies to increase understanding of mechanisms causing the events[13–15]. Identifying heatwave patterns allows a meaningful way of dimensionality reduction, which is important for further research on driving physical mechanisms for heatwave occurrence.

Current research highlights that heatwaves and droughts are highly interrelated and caused by similar persistent large-scale atmospheric circulation patterns[16–18]. Moreover, the self-intensifying nature of extreme droughts and heatwaves has been suggested as central to their evolution[19–21]. There is a two-fold relationship. On the one hand, soil and vegetation dry with the occurrence of a heatwave, leading to reduced evaporation. Therefore, the likelihood of rainfall decreases, favouring the formation of drought[20,22]. On the other hand, evaporation decreases with the onset of drought. The reduced cloud cover leads to a larger fraction of solar radiation reaching the land surface, increasing the likelihood of heatwave formation[20,22]. Global-warming-induced changes in thermodynamic conditions account for 57.3% of Europe's increase in extreme heat occurrence[23].

The influence of precipitation and soil moisture anomalies on heatwave formation has been studied in different European regions. A rainfall deficit in the Mediterranean in spring is found to favour the formation of heatwaves in Northern Europe as the rainfall deficit propagates northward throughout the summer[16,24]. A recent study[17] confirmed that dry conditions in winter/spring seasons prevail prior to hot summers over Southern Europe. Other studies confirm that anomalously dry Western and Northern European summers significantly correlate with the occurrence of heatwaves in those regions[25]. Soil moisture and other precipitation-related indices correlate with the temperature extremes in South-Eastern Europe[26,27]. For the European heatwave of July 2019, land-atmosphere feedback and influences of northward propagation of dryness contributed to the exceptional intensity of the event[28].

Most research on heatwaves investigates historical events based on observational data[17,24,26,29], which rarely happen by definition. Moreover, only few studies have analysed generalised patterns of heatwaves to derive scientific findings applicable to coherent regions[24,29] instead of focusing on the causes of single events[30,31]. By 'coherent regions', we here and thereafter mean regions connected to similar atmospheric circulation patterns, such that the heatwaves occur simultaneously and over the same geographical region. Large climate model ensembles have proven their usefulness for the investigation of extreme events both in terms of extreme cold and wet, as well as in terms of hot and dry events[32–36]. They allow the assessment of the natural variability of extreme weather events and therefore facilitate to derive of statistically reliable findings. Moreover, regional climate models offer a finer spatial resolution, which allows for the resolution of finer spatiotemporal processes and therefore obtains spatial patterns on a regional and subregional level when compared to Global Climate Models[37,38].

[1]Technical University of Munich, Munich, Germany. [2]Department of Geography, Ludwig-Maximilians-Universität München, Munich, Germany. [3]Center for Digital Technology and Management, Munich, Germany. ✉email: felsche@cdtm.de

In this study, we want to take a regional approach to the investigation of heatwaves, as those usually cover a fraction of the continent. Therefore, we aim to find stable spatial patterns of heatwaves using the 50 members of a Single Model Initial-condition Large Ensemble (SMILE) CRCM5-LE over Europe. Canadian Earth System Model 2-Large Ensemble (CanESM2-LE) during the period 1950-2099 is used to derive the boundary conditions for the Canadian Regional Climate Model version 5-Large Ensemble (CRCM5-LE)[39]. CRCM5-LE obtains more realistic representations of climate over complex topologies, as in the southwest part of Scandinavia, the Iberian Peninsula, the Alps and the Pyrenees[39]. A study by Trentini et al.[40] confirms the applicability of the chosen model for heatwave research in the European domain. It investigates the interannual variability of three different large ensembles, with CRCM5-LE being one of them, and compares it to E-OBS data. The study shows that CRCM5-LE has a good representation of JJA temperature and the number of heatwave days per year. Another study[41] compares CRCM5-LE with the EURO-CORDEX ensemble and confirms the added value of the ensemble for the European domain. Moreover, CRCM5-LE was already used in a multitude of studies on European extreme events[32,42,43].

We use 1500 model years that correspond to the years 1981-2010 historical climate. A 'heatwave day' occurs when the local daily maximum temperature exceeds the 95th JJA percentile of the whole period. We use the three-day-running mean in order to obtain robust signals. In total, we obtain more than 50,000 heatwave days. For an exact definition of a heatwave day, see the section 'Methods'. Following previous studies on heatwave classification[24,29], in the first step, we apply hierarchical agglomerative clustering on the heatwave days in order to identify predominant heatwave patterns. The clustering algorithm starts by assigning each data point to its own cluster (agglomeration). Then it merges all the clusters using a defined similarity measure and builds a hierarchy between clusters, which is based on how similar they are to one another[44]. We use cosine similarity as our similarity measure. The optimal number of clusters is determined using the elbow method. It picks the number where the added information by creating one more cluster sharply drops[45]. This point is determined by calculating the knee of the curve. For a detailed description, see 'Methods'. Subsequently, the obtained spatial patterns of heatwaves are analysed in terms of the influence of soil moisture and precipitation conditions in spring and summer on heatwave formation and the influence of heatwave occurrences on dry conditions in the following fall/winter.

RESULTS

Typical European heatwave patterns

We focus our investigations on the European domain of the CRCM5-LE[39], as we are interested in regional heatwaves. We obtain a total of nine significant spatial patterns from CRCM5-LE for the years 1981–2010. Figure 1 shows the identified spatial patterns, which we order from West to East: Iberian Peninsula (IP), Western Europe 2 (WE2), Western Europe 1 (WE1), Britain and Ireland (BI), South-Eastern Europe (SEE), Greece and South Italy (GSI), Scandinavia (SCA), Central-Eastern Europe (CEE) and North-Eastern Europe (NEE).

The pattern significance is assessed via bootstrapping, which we apply according to the existing literature on heatwave clustering[24,29]. For bootstrapping, we divide the data set into a validation and training set 100 times so that one-hundredth of the data is assigned to the validation set and the rest to the training data set. We perform clustering using the training data set and then assign clustering classes to the validation data according to the nearest distance to data points within the training data set.

The obtained labels are compared to the ones originating from clustering the whole data set. A stability score is calculated for each cluster. It corresponds to the number of correctly assigned events vs the total number of validation events per spatial pattern. The stability scores are compared to the ones from a Monte-Carlo pseudo-experiments, where we assign the validation data points 1000 times to one of the clusters in a random way. This allows us to estimate the probability density function of the null hypothesis that the clustering does not entail information. In Fig. 2, the mean stability scores per cluster derived from bootstrapping are compared with the ones from the Monte-Carlo pseudo experiments. The nine patterns are significant on the 99 % level according to a two-sided t-test; the least stable spatial patterns with a stability score below 0.9 are WE1, SEE and CEE.

The visual inspection of the spatial patterns confirms their meaningfulness since natural geographical boundaries like mountains serve as delimiting boundaries, as is the case for IP, WE2 and SEE. In order to characterise the heatwave patterns, we examine the mean maximum temperature and the mean calendar day of the first heatwave occurrence (see Fig. 1). We find three spatially related groups when looking at the mean first calendar day of the heatwave in a year. The earliest events happen in the BI pattern with the mean first calendar day of the event of 25th June, followed by Northern patterns of SCA, CEE, and NEE at the beginning of July. The mean first calendar day of heatwave is the latest in the Southern and Central European patterns of IP, WE2, WE1, SEE, and GSI, where the mean first calendar day of heatwave occurs in the second half of July. The mean maximum temperature is higher for the patterns with fewer events - e.g., WE2 and WE1. From that, we can derive that events that belong to those patterns have their hot spots over the same area, while, e.g., in the case of GSI, the maxima of the respective events match to a lesser percentage.

Next, we visually compare the patterns to observed historical heatwaves in Europe. We find that many patterns obtained from the analysis on CRCM5-LE reproduce historical events, even though those have not been part of the analysis. For example, the WE1 is similar to the French heatwave in the summer of 2003[1]. The record-breaking heatwave in the summer of 1976 in Britain can be matched with the BI pattern[1]. CEE pattern reproduces the heatwave of 1994 in Eastern Germany and Poland[1]. Finally, the events of 2007 in the Balkans and Greece and 2010 in Russia can be matched to SEE and NEE, respectively[1].

Additional validation is performed by comparing the spatial patterns from CRCM5-LE to the ones derived from the clustering of heatwaves derived from the observational data set E-OBS. The E-OBS' spatial patterns can be found in Fig. 3. To compare both clustering results, we calculate the cosine similarity between the spatial patterns obtained from CRCM5-LE and those from E-OBS and match them by the maximum value. The measure is chosen to stay consistent with the distance measure used for clustering. Cosine similarity corresponds to one when the input vectors are identical and to zero if they are orthogonal. The results are shown in Table 1. The patterns IP, BI, SCA and CEE, are in excellent correspondence, as can be seen visually and from the pattern cosine similarity. Furthermore, WE1 and WE2 combine to one common pattern in E-OBS - the WE, as indicated by the high similarity value. Therefore, we calculate the sum of the patterns by adding the values pixelwise. Similar behaviour can be seen in SEE and GSI, which divide into a Southern and a Northern part. In contrast, the patterns originating from E-OBS divide into West (Italy) and East (Balkans and Greece). Finally, the two North-Eastern patterns in E-OBS combine into the NEE pattern of CRCM5-LE. Supplementary Table 1 shows the pattern similarity values derived from the ERA-Interim-driven model run of CRCM5 (CRCM5/ERA) and CRCM5-LE. The patterns can be found in Supplementary Figure 1. The results are comparable to those for E-OBS. They confirm that the dominating spatial heatwave
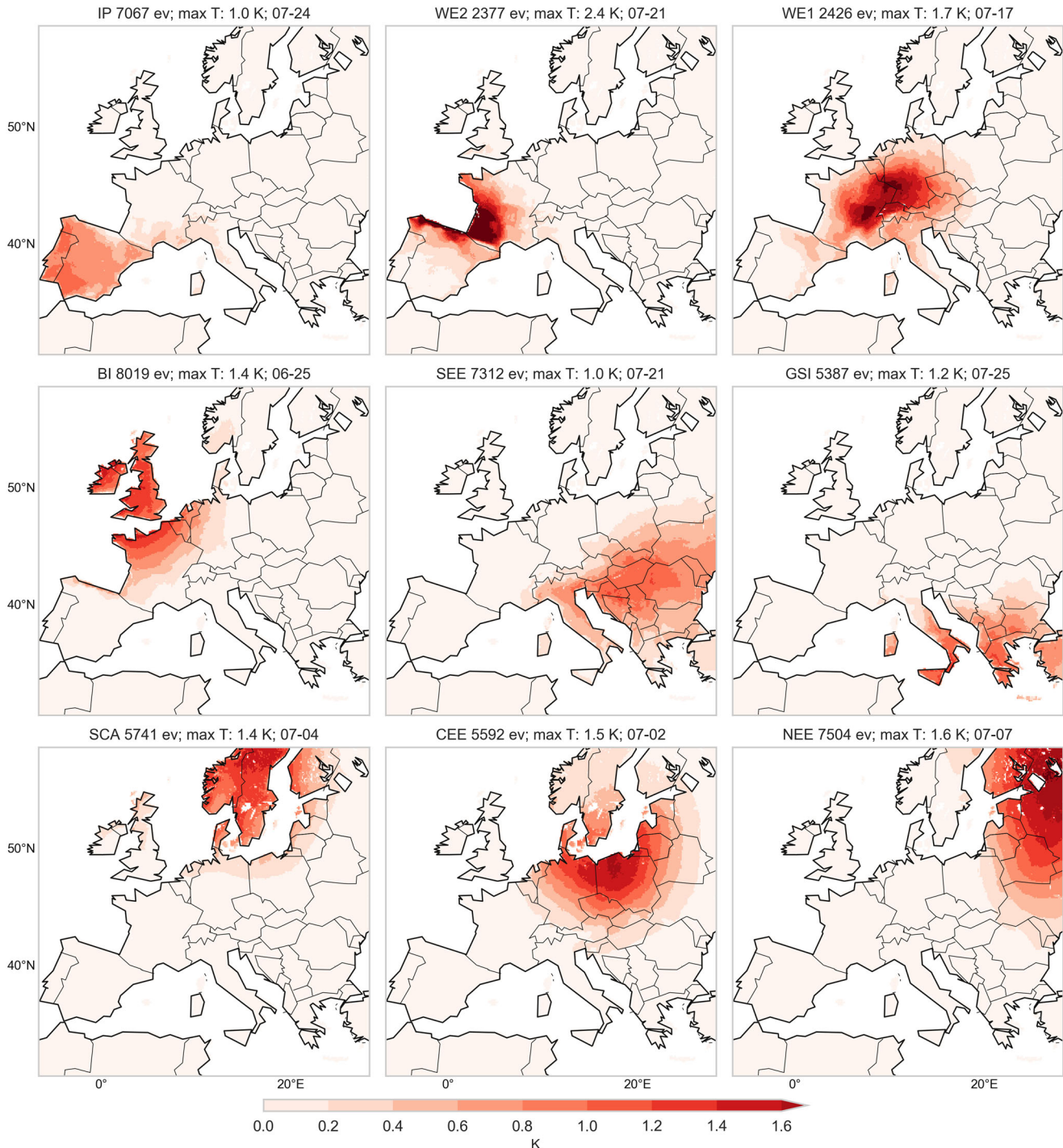
Fig. 1 Nine typical heatwave patterns over Europe derived from CRCM5-LE. Patterns obtained by hierarchical clustering of 1981-2010 Canadian Regional Climate Model 5 Large Ensemble (CRCM5-LE). In the title from left to right: pattern abbreviation, number of events belonging to the pattern (ev), mean maximum temperature in K, mean calendar day of the first heatwave occurrence. From left to right, from top to bottom: IP: Iberian Peninsula, WE2: Western Europe 2, WE1: Western Europe 1, BI: Britain and Ireland, SEE: South-Eastern Europe, GSI: Greece/Southern Italy, SCA: Scandinavia, CEE: Central/Eastern Europe, NEE: North-Eastern Europe.

patterns from CRCM5-LE are similar in the area they cover with those found when clustering observational data or the reanalysis-driven run of CRCM5. Given the difference in the number of events used as input for the analysis (1059 events from E-OBS vs 51.044 events from the 50 members of CRCM5-LE), we argue that the patterns originating from CRCM5-LE allow reliable statistical interpretability and robustness and are therefore used for further analysis.

Additionally, we test the robustness of our results in terms of domain choice. As we cannot pick a larger domain, we compare the resulting patterns for a smaller domain. We cut off ten boundary pixels on each side, thereby reducing the 280 × 280 grid to 260 × 260. The resulting heatwave patterns are similar in form and shape, as in Fig. 1, but without the grid cells at the domain's border. Therefore, we conclude that the resulting heatwave patterns do not depend on the domain choice. Moreover, we
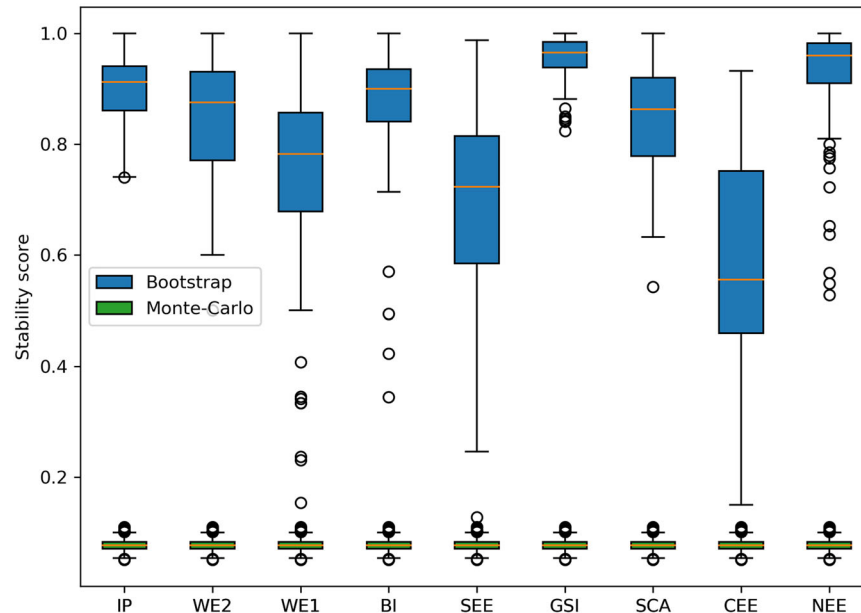
Fig. 2   Mean stability score per heatwave pattern. The stability score of bootstrap samples is compared to Monte-Carlo pseudo-experiments. The median in orange, end of the box, indicates the first and third quartiles. The boundaries of the whiskers indicate the 1.5 interquartile range.
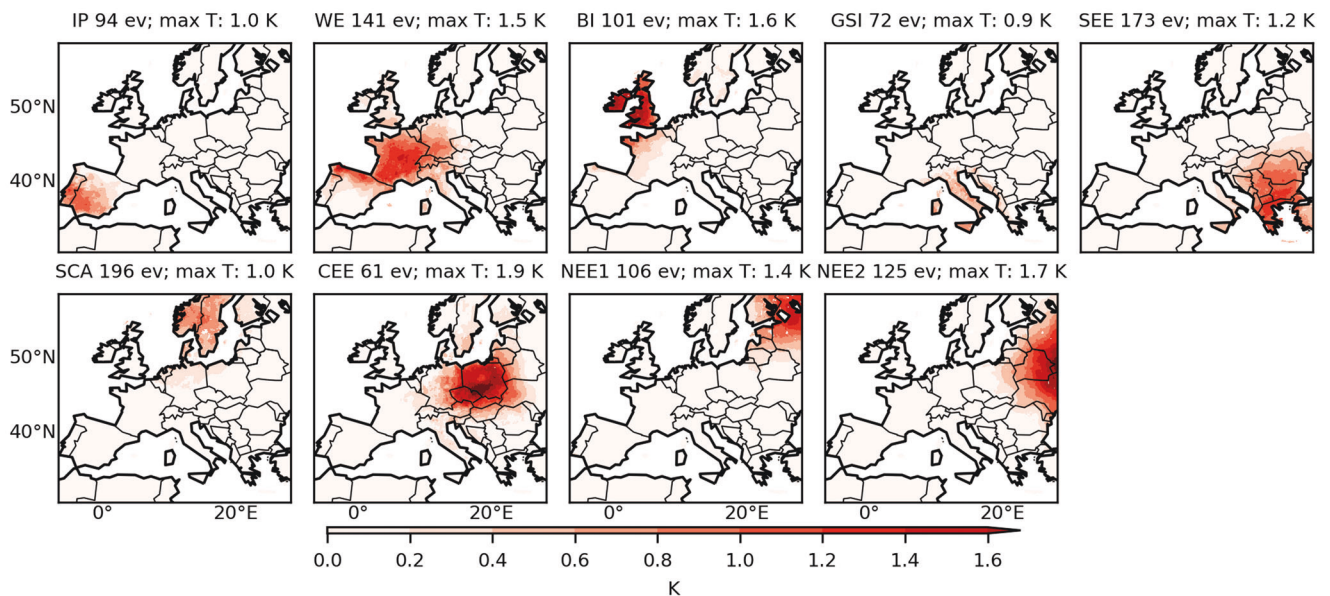

Fig. 3   Nine typical heatwave patterns over Europe derived from E-OBS. In the title from left to right: pattern abbreviation, number of events belonging to the pattern (ev), mean maximum temperature in K. Pattern names same are the same as in Fig. 1, except NEE1: North-Eastern Europe 1, NEE2: North-Eastern Europe 2.

compare the patterns when including sea grid cells. We calculate the spatial patterns with sea grid cells for the ERA-Interim driven run of CRCM5. The results are shown in Supplementary Fig. 3. We see that new patterns emerge over the sea areas that are not impacting land clusters. In 7 out of 9 cases, land patterns stay very similar—they cover the same area and, in some cases, add the sea areas along the coast (e.g., WE and BI). The CEE pattern is no longer present; however, it is constituted only out of 24 events without the sea grid cells and is, therefore, unstable. Moreover, IP splits up into IP1 and IP2. Therefore, we conclude that heatwave patterns over land are mainly unrelated to sea heatwaves, and we omit sea grid cells in further analysis.

Seasonal connection to soil moisture and precipitation

Heatwaves and droughts are related phenomena that influence the formation of one another, as the hydrological cycle is inseparably connected to the heat-related processes in the atmosphere. Therefore, we inspect soil moisture anomalies and anomalies in seasonal precipitation before the heatwave occurrence (JFMA), after the heatwave (OND) and during the heatwave (MJJAS) in dependence on the number of heatwave days in every spatial pattern per summer.

The quantile regression method is applied to investigate the relationship between the number of heatwave days per summer season and the soil moisture or seasonal precipitation (for more

Table 1. E-OBS patterns assigned to CRCM5-LE patterns by the maximum value of cosine similarity.

CRCM5-LE pattern	EOBS pattern	Similarity
IP	IP	0.86
BI	BI	0.85
SCA	SCA	0.92
CEE	CEE	0.83
SEE	SEE	0.74
GSI	SEE	0.74
NEE	NEE1	0.87
WE1	WE	0.82
WE2	WE	0.68
WE1 + WE2	WE	0.93
GSI + SEE	GSI + SEE	0.92
NEE	NEE1 + NEE2	0.95

information, see 'Methods' section). A scatter plot of soil moisture and precipitation versus the number of heatwave days for IP in JFMA is shown in Supplementary Fig. 2. We use the range of 10-90th percentile, which allows us to investigate if there is a link between the variables for the upper and hence more extreme quantiles. We expect that the relationship between soil moisture/precipitation and the number of heatwave days differs for upper quantiles. For each pattern, we plot the soil moisture anomalies for the 25 years (2% of most extreme events) with the highest number of heatwave days to obtain a visual validation for the correlations. The soil moisture anomalies in the upper portion of the soil column (0cm-10cm depth) are used instead of deeper soil moisture levels due to data availability. Additionally, we repeat the analysis using a model run where ERA-Interim is used as a boundary condition instead of CanESM2 to compare and validate the findings.

Figure 4a, b shows the quantile regression slopes of the number of heatwave days N_{HW} in relation to soil moisture $mrsos_{JFMA}$ and precipitation anomaly pr_{JFMA} in the preceding winter/spring season (JFMA). Statistically significant slopes with a 90% confidence level for a two-sided t-test are identified with black edge. We find gradually increasing negative slopes for an increasing number of heatwave days for North European patterns of BI and NEE and Southern-European patterns GSI, SEE, IP and WE2. In the case of the SEE pattern, there is a stronger influence of precipitation deficit on the number of heatwave days than when compared to other patterns. By contrast, we find no significant relationship between the soil moisture anomaly and precipitation deficit in winter/spring and the number of heatwave days in the Central European WE1 and CEE patterns, as well as in the Northern European pattern SCA. Our results suggest that there is a predictive power of soil moisture in the preceding winter/spring (JFMA) for heatwave occurrence in summer for South and North Europe. Moreover, our results suggest that there is predictive power of seasonal precipitation anomalies in winter/spring (JFMA) in SEE for heatwave occurrence in summer. In Supplementary Fig. 4, we show the results of the quantile regression analysis for the ERA-Interim-driven run of CRCM5. The results confirm the negative relationship between the number of heatwave days and mrsos anomaly in winter only for NEE, BI, SEE and IP, although none of the slopes is significant.

Figure 4c displays the spatial distribution of mrsos anomalies for the patterns IP, WE2, SEE, GSI, NEE and BI for the 25 years with the highest number of heatwave days. We choose the patterns that show a significant relationship in the quantile analysis. Significant anomalies are indicated with the black edge. We find that the extreme heatwave years in IP and SEE patterns are connected not only to local soil moisture and precipitation deficit in the pattern area but also in other parts of South Europe. We find that extreme heatwaves in the GSI pattern are connected to continental and Northern Europe soil moisture anomalies. Following Fig. 4a, the anomalies are insignificant for the Northern patterns (NEE, BI). The obtained results confirm findings concerning the positive influence of the dry winter season on hot summers[16,19,46]. We cannot find a significant dependency between soil moisture deficits in the South and heatwaves in the North of Europe in contrast to what is suggested by previous studies[16,24]. Results for the remaining patterns are displayed in Supplementary Fig. 5.

A deficit in soil moisture in the season during the heatwave (MJJAS) is present for all identified patterns in Fig. 5. We see significantly decreasing slopes in Fig. 5a of N_{HW} versus $mrsos_{MJJAS}$ for all patterns. We find the same, although mostly non-significant, relationship when performing the analysis on the ERA-Interim dataset (see Supplementary Fig. 4b). These results are in accordance with a previous study that also found increasing negative slopes for the quantile analysis of soil moisture in relation to the percentage of heatwave days in Central and Eastern Europe[26]. We extend these results by finding this relationship also in Northern Europe. The results are similar in the case of seasonal precipitation as the dependent variable. We see that for the BI pattern, soil moisture has a bigger influence than for other patterns; for the IP pattern, precipitation anomalies are a more robust predictor compared to other patterns.

In Fig. 5c, the spatial patterns are displayed. A significant soil moisture deficit in SCA, CEE, and NEE for the 25 years with the highest number of heatwave days is also connected to a significant soil moisture increase in Western and Central Europe. A contrasting pattern is visible in the IP region: for the 25 years with the highest number of heatwave days, negative soil moisture anomalies are observed in South-Western Europe and positive in North-Eastern Europe. The dipolar structure is a well-known phenomenon: it has been shown in previous studies that positive phases of the North Atlantic Oscillation are connected to negative SPI averages in Southern Europe and positive averages in Northern Europe[47].

Extremely long periods of heatwaves pose substantial stress on the soil moisture visible in the following season (OND), as seen in Fig. 6. While for quantiles below 0.2, coefficients equal zero, the slopes turn negative for higher values for patterns SCA, NEE, SEE, CEE and IP. Out of those, significant soil moisture anomalies in the following season are present only for patterns SCA, NEE and SEE. This is confirmed by the analysis of the ERA-Interim-driven run (Supplementary Fig. 4c) apart from the highest quantile. These results serve as an indication of the memory of soil moisture in Northern Europe, as well as in the South-Eastern parts and suggest that there is a predictive power of the number of heatwave days per summer on dry anomalies in soil moisture in subsequent fall/winter (OND). The slopes of the quantile regression for precipitation anomalies are shown in Fig. 6b. None of the slope coefficients is negative; this suggests that hot summers do not lead to dry fall/winter in Europe. In contrast, we see a positive correlation between the upper quantiles of the number of hot days and seasonal precipitation in BI and NEE. The observed quantile regression coefficients are the lowest compared to the other seasons. For SCA, NEE and SEE, the negative soil moisture anomalies are also visible in the following winter season (see Fig. 6c). Results for the remaining patterns are displayed in Supplementary Fig. 5.

DISCUSSION

Using cluster analysis, we identified (1) nine distinct patterns of European heatwaves, which we validated by comparing with E-OBS and applying bootstrapping. The spatial patterns offer not only the possibility to investigate regional heatwave characteristics, e.g., BI
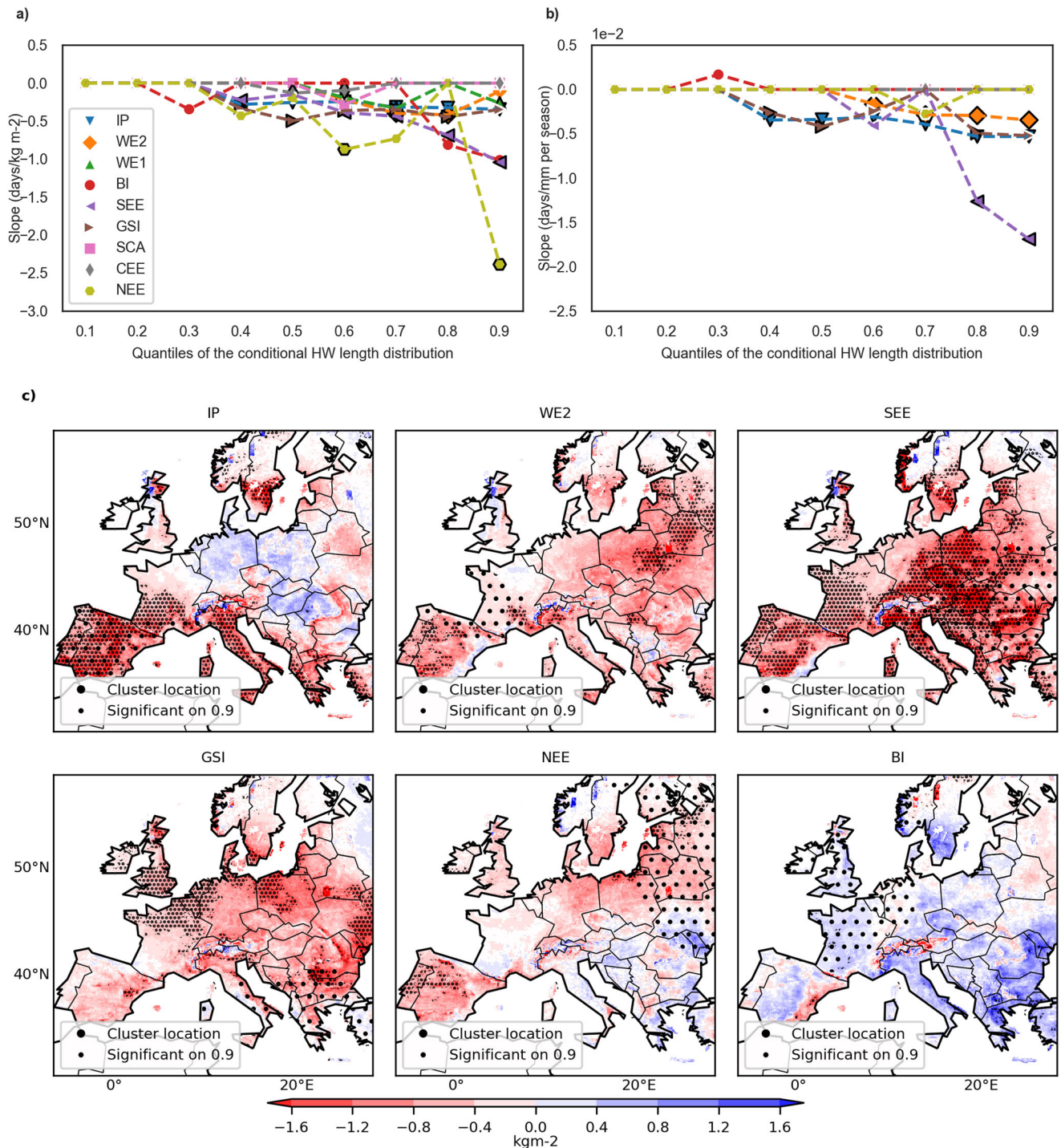
Fig. 4 Quantile regression slopes for N_{HW} in relation to moisture-related variables in winter before. N_{HW} versus $mrsos_{JFMA}$ (**a**) and pr_{JFMA} (**b**). Statistically significant slopes with a 90% confidence level with a two-sided t-test are identified with black edge. $mrsos_{JFMA}$ for the 25 years with the highest number of heatwave days in chosen patterns with significant precipitation anomalies (**c**).

earliest heatwaves to latest in Southern parts of Europe, but also offer to understand further the seasonal influence of large-scale soil moisture anomalies and precipitation anomalies on the number of heatwave days in the chosen patterns and vice versa. We show that (2) soil moisture deficit in the preceding winter/spring (JFMA) can serve as a predictor for heatwaves in Southern (GSI, SEE, IP, WE2) and Northern (BI, NEE) Europe. Moreover, (3) all patterns show a significant negative relationship between soil moisture in the summer season (MJJAS) and the number of heatwave days. (4)

The analysis of soil moisture anomalies in the following season (OND) shows a significant negative relationship for SCA, SEE and NEE. This shows that long heatwave events lead to a substantial soil moisture deficit preserved until the following season. For now, the obtained findings apply only to present-day climate; it has to be further investigated whether future climate change impacts these relationships.

In this study we perform a clustering analysis of heatwaves using a SMILE of a high-resolution RCM. Through the employment
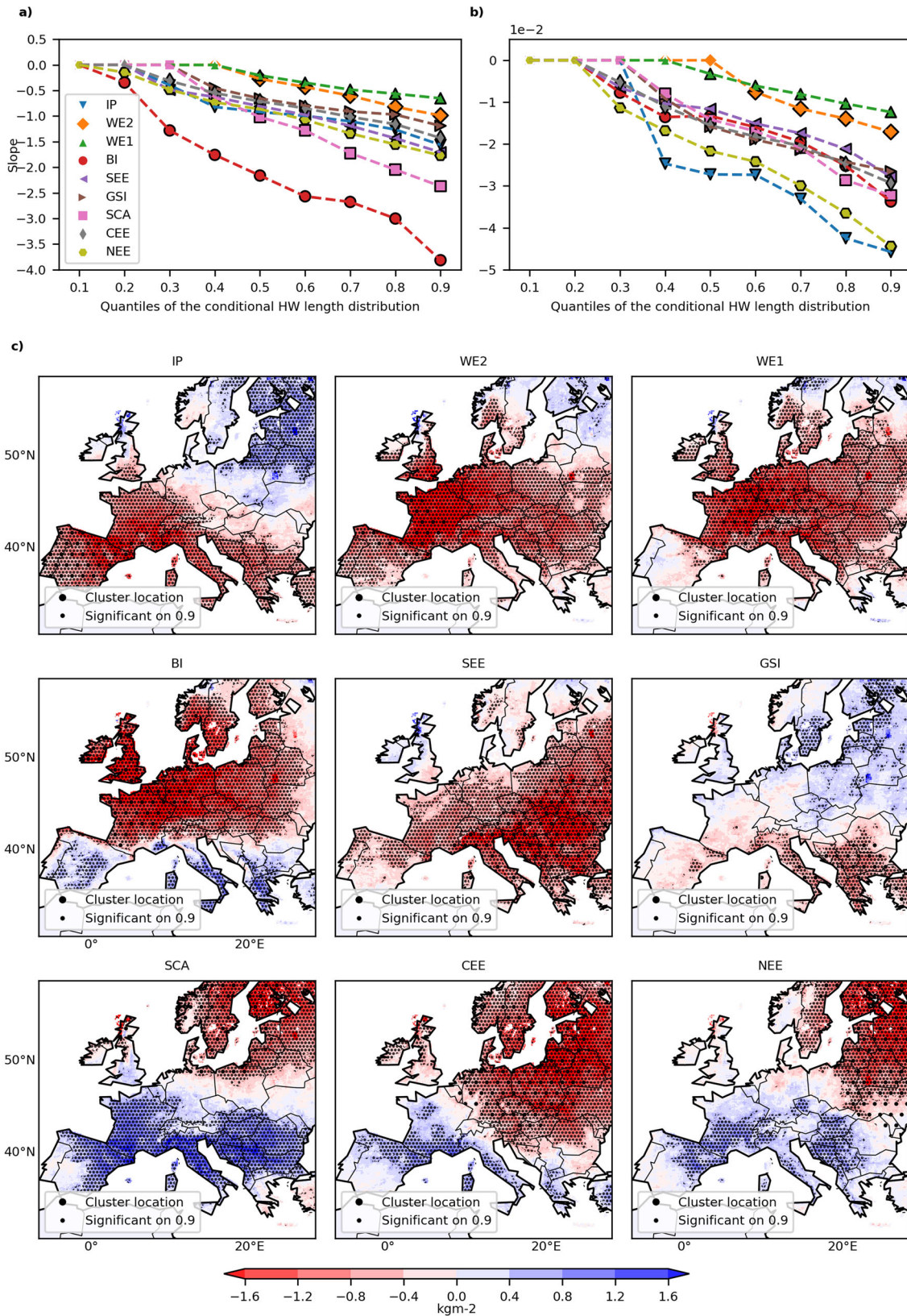
Fig. 5 Quantile regression slopes for N_{HW} in relation to moisture-related variables in summer. N_{HW} versus $mrsos_{MJAAS}$ (a) and pr_{MJJAS} (b). Statistically significant slopes with a 90% confidence level with a two-sided t-test are identified with the black edge. $mrsos_{MJJAS}$ for 25 years with the highest number of heatwave days in (c).
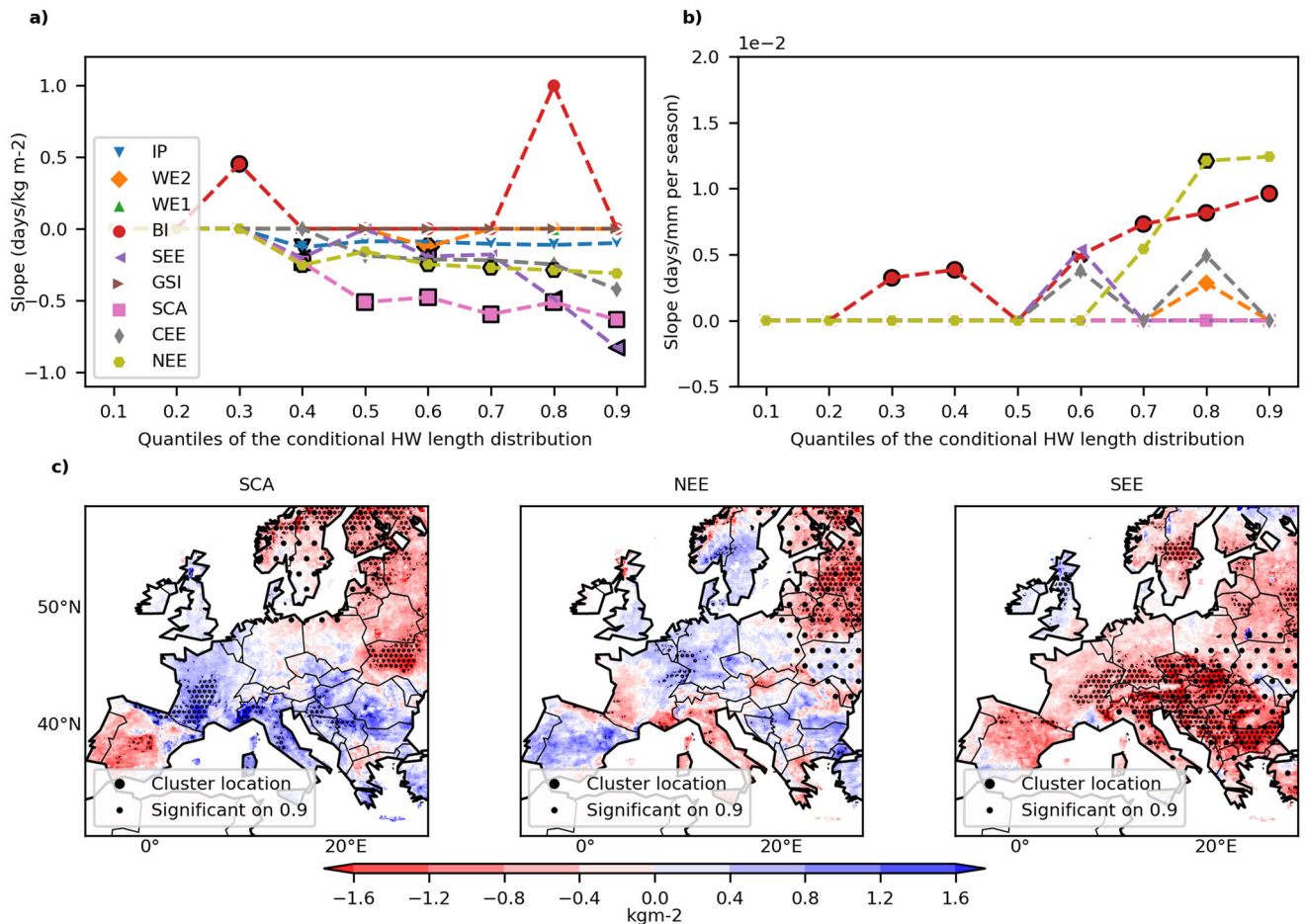
a)

b)

c)

Fig. 6 Quantile regression slopes for N_{HW} in relation to moisture-related variables in fall after. N_{HW} versus $mrsos_{OND}$ (**a**) and pr_{OND} (**b**). Statistically significant slopes with a 90% confidence level with a two-sided t-test are identified with black edging. $mrsos_{OND}$ for the 25 years with the highest number of heatwave days in chosen patterns with significant precipitation anomalies (**c**).

of the CRCM5-LE, we assess the natural variability of heatwaves and derive stable patterns of heatwaves. Regional Large Ensembles have proven useful in research on extreme events[15,32,33]. Nevertheless, it is known that the models are prone to biases regarding the modelling of land-atmosphere interactions that contribute to the formation of heat waves[48]. We find similar patterns when clustering using the E-OBS dataset or ERA-Interim driven run of CRCM5.

The classification into nine distinct heatwave patterns in Europe is a unique finding of this study. The study by Stefanon[24] finds six heatwave patterns, however, based on 78 heatwave events that consist of 643 heatwave days in contrast to more than 50,000 heatwave days used in our case. When we compare those patterns to the ones found by our analysis, we can assign them in the following way: 'Russian' pattern compares to NEE; 'Western Europe' pattern to WE1 and WE2; 'Eastern Europe' pattern to CEE, SEE, GSI; 'Iberian' pattern to IP; 'North Sea' pattern to BI and SCA; 'Scandinavian' pattern to none, however, its area is in huge parts outside of our domain. Therefore, we find similar heatwave patterns in both studies. Also, previous studies mostly used the percentage of rain days as a soil moisture proxy for the analysis instead of direct soil moisture, as in our case[16,17].

The coupling between spring soil moisture availability in Southern Europe (GSI, IP, WE2, SEE) and heatwave occurrence in summer has been described in previous studies[24,49]. The southern regions of Europe have a dry climate, where evaporation is soil moisture limited[50]. The described link between reduced soil moisture leading to fewer clouds and more solar radiation and,

therefore, more heatwaves is valid for that region. It appears to be one of the main driving mechanisms for heatwave formation. The link for NEE is less significant, however, present. It has to be investigated further how far other factors, such as snow cover, influence the link. The missing coupling between spring soil moisture and precipitation and the occurrence of a heatwave in SCA, WE1 and CEE can be explained by the fact that the vegetation system in those regions is rarely water-limited[51]. Therefore, even if there is a comparably dry spring, the soil still has enough moisture for evaporation and the formation of clouds.

We see a coupling between the heatwave occurrence in the Eastern and Northern parts of the domain (SCA, NEE, and SEE) and autumn drought occurrence. Those regions have a temperate climate and relatively high mean soil moisture values. This allows for a higher variability of soil moisture when compared to more Southern regions. Therefore, it takes the soil until the next season to recover after a prolonged heatwave. This relationship is missing for the Southern regions (IP, GSI, WE2). Those regions experience low mean precipitation in summer and, therefore, a lower expected and possible variability of soil moisture.

These results can, in most cases, be confirmed when performing the analysis on the ERA-Interim-driven run of CRCM5. The results mainly differ for upper quantiles (0.9). It can be explained by the fact that we analyse only 30 years; therefore, the upper quantile includes only three values, leading to high slope value uncertainty.

For future research, we suggest performing the analyses on deeper soil moisture levels, as those are known to show higher persistence[52]. Further, we suggest analysing for interdependencies

between heatwaves and precipitation/soil moisture deficits across different areas, as we see, e.g., in Fig. 6 for BI pattern, soil moisture over Central and Eastern Europe shows significant anomalies for the 25 years with the highest number of heatwave days. Also, a further heatwave-pattern-based investigation in terms of the effects of other variables, such as latent and sensible heat fluxes, would be of interest for future studies on the interrelation of heatwaves and droughts.

We suggest using the obtained patterns for heatwave analysis and predictability instead of pixelwise or even country-wise approaches in future studies. The obtained patterns allow a meaningful complexity reduction by finding spatially coherent regions instead of arbitrary grouping, e.g., by country. For agricultural research and the general public, the study's outcomes can enhance the predictability of heatwave events in Southern (GSI, SEE, IP, WE2) and Northern (BI, NEE) Europe on a seasonal scale.

Moreover, applying the described framework offers great potential for investigating other extreme events like droughts.

METHODS

Data sets

The central part of our analysis is based on the daily maximum temperature and monthly soil moisture and precipitation data from the single-model initial condition large ensemble (SMILE) consisting of 50 members, the Canadian Regional Climate Model 5 Large Ensemble (CRCM5-LE). The data was produced within the scope of the ClimEx Project (Ref. [39], www.climex-project.org). Dynamical downscaling via CRCM5-LE is applied to the data originating from the 50-member initial condition Canadian Earth System Model 2 (CanESM2)[53]. The data is provided at a resolution of 0.11° (12 km) and is produced for the years 1950–2099 for a European and an Eastern North America domain. Historical greenhouse gas concentrations are used for the years 1950–2005; starting from 2006, the RCP8.5[54] forcing scenario is used. We use the data from all 50 members for the years 1981–2010, translating to 1500 model years, which are analysed for heatwave events. A comparison of the CRCM5-LE to the E-OBS dataset has been performed in a previous study[39] and showed a temperature bias between −2 and +2 °C, while warm deviations mainly happen over highlands. For the validation of the obtained patterns, we use daily gridded observational data set E-OBS[55] for the years 1981-2010, as well as one model run of the CRCM5, which was driven by the global atmospheric reanalysis data set ERA-Interim via boundary conditions[39,56].

Heatwave definition

Literature gives evidence for a wide range of similar heatwave definitions, which are adapted to the specific study goals[24,57–59]. In this study, we define heatwaves for land areas in continental Europe (EUR-11 domain) as prolonged periods of above-average temperatures in an extended area during the period 1981–2010. These heatwaves consist of at least three consecutive hot days, where hot days are characterised by a positive anomaly of daily maximum temperature (tasmax) to the local 95th JJA (1981–2010) percentile, allowing for comparability across the domain. In order to obtain robust signals, we use the 3-day-running mean to derive these anomalies. Negative anomalies are set to zero to focus on hot extremes[24]. Two heatwaves are separated by a minimum of three days below threshold[57]. We remove heatwave days consisting of patterns smaller than 9×9 grid cells. An additional filter is introduced to eliminate spatially small events covering an area of less than 1% of the land area (500 grid cells). Positive anomalies only occur during the months May-October in our data sets. The analysis is based on heatwave days fulfilling the above-mentioned criteria and amounts in the case of CRCM5-LE to around 50,000 heatwave days used as input for the clustering analysis.

Clustering analysis

In literature, clustering has frequently been used to analyse and classify weather patterns in the mid-latitudes[60–62]. This study uses the obtained heatwave days as input to the agglomerative hierarchical clustering algorithm[24,29]. Distance between two vectors, $\boldsymbol{r}$ and $\boldsymbol{q}$, is defined here as follows:

$$d(\boldsymbol{r}, \boldsymbol{q}) = 1 - cs(\boldsymbol{r}, \boldsymbol{q}) \tag{1}$$

$$cs(\boldsymbol{r}, \boldsymbol{q}) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} r_{i,j} q_{i,j}}{\left(\sum_{i=1}^{N} \sum_{j=1}^{M} r_{i,j}\right)^{1/2} \left(\sum_{i=1}^{N} \sum_{j=1}^{M} p_{i,j}\right)^{1/2}} \tag{2}$$

$cs(\boldsymbol{r}, \boldsymbol{q})$ refers to the cosine similarity measure between two vectors[63]. It is defined as 1 for parallel vectors and as 0 for orthogonal. For the clustering algorithm, the average linkage is used[63].

The optimal number of clusters is determined by applying the elbow method[64]: we compute the distortion score as the sum of squared distances to the assigned centre for every possible number of clusters and pick the number of clusters that corresponds to the knee of the curve[45,64].

Due to a large number of events, the obtained data set has, in absolute numbers, a higher number of a-typical events, which have a big distance to all other events of the data set. Filtering by preliminary clustering to 32 clusters is introduced to remove these events. Events belonging to so-called minority clusters with a small number of events (<0.1% of the data) are removed from the data set[65]. In total less than 1% of heatwave events are removed that way. Repeated clustering is performed on the resulting data set. We derive 12 clusters as the optimal number from the elbow method. The obtained clusters are cross-validated by 100 times dividing the data set into a verification period that amounts to 1/100 of the data set and the remaining 99/100 used as input to clustering. The nearest distance to the training clusters then determines the labels for the verification period. Those are then compared to the ones originating from the clustering on the whole data set. Finally, a stability score is computed per cluster that amounts to the number of correctly assigned validation events to the total number of events per cluster. The results are then compared to a Monte-Carlo pseudo-experiment, where the labels are assigned in a purely random way 1000 times. Three out of twelve clusters do not pass the described validation; nine are significant on the 99%-level according to a two-sided t-test.

Quantile regression

In order to evaluate the impact of heatwave length on soil moisture and seasonal precipitation deficit, we use quantile regression as suggested by similar studies[17,26]. Quantile regression is a method that goes beyond standard linear regression, as it can be used when the linearity and independence of variables are not given. It estimates the conditional median of the target variable[66]. Here, we use a linear model for the conditional quantiles.

For the quantile regression, we use the following variables derived from the CRCM5-LE data set:

1. Number of heatwave days per pattern N_{HW} per summer season of May, June, July, August, and September (MJJAS)
2. Mean soil moisture anomalies in the upper portion of the soil column (top 10 cm) averaged for the following three seasons in the pattern region: January, February, March, April (JFMA) $mrsos_{JFMA}$; May, June, July, August, September (MJJAS) $mrsos_{MJJAS}$ and October, November, December (OND) $mrsos_{OND}$
3. Summed precipitation anomalies for the same seasons as in (2): pr_{JFMA}, pr_{MJJAS} and pr_{OND} in the pattern region.

The pattern region for $mrsos$ and pr is defined as the area of 100 land pixels around the maximum of the spatial pattern. Given the

spatial resolution of 12 km, which amounts to approximately 14400 km^2.

CODE AVAILABILITY

The source codes for the analysis of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Russo, S., Sillmann, J. & Fischer, E. M. Top ten European heatwaves since 1950 and their occurrence in the coming decades. *Environ. Res. Lett.* **10**, 124003 (2015).
2. Sedlmeier, K., Feldmann, H. & Schädler, G. Compound summer temperature and precipitation extremes over Central Europe. *Theor. Appl. Climatol.* **131**, 1493–1501 (2018).
3. Orth, R., O, S., Zscheischler, J., Mahecha, M. D. & Reichstein, M. Contrasting bio-physical and societal impacts of hydro-meteorological extremes. *Environ. Res. Lett.* **17**, 014044 (2022).
4. Robine, J.-M. et al. Death toll exceeded 70,000 in Europe during the summer of 2003. *C. R. Biol.* **331**, 171–178 (2008).
5. Kjellström, T., Maître, N., Saget, C., Otto, M. & Karimova, T. *Working on a Warmer Planet: The Effect of Heat Stress on Productivity And Decent Work* (International Labour Organization, 2019).
6. Zink, M. et al. The German drought monitor. *Environ. Res. Lett.* **11**, 074002 (2016).
7. Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M. & García-Herrera, R. The hot summer of 2010: redrawing the temperature record map of Europe. *Science* **332**, 220–224 (2011).
8. Perkins-Kirkpatrick, S. & Lewis, S. Increasing trends in regional heatwaves. *Nat. Commun.* **11**, 1–8 (2020).
9. García-León, D. et al. Current and projected regional economic impacts of heatwaves in Europe. *Nat. Commun.* **12**, 1–10 (2021).
10. Ballester, J., Rodó, X. & Giorgi, F. Future changes in Central Europe heat waves expected to mostly follow summer mean warming. *Clim. Dyn.* **35**, 1191–1205 (2010).
11. Rossi, G., Cancelliere, A. & Giuliano, G. Case study: multicriteria assessment of drought mitigation measures. *J. Water Resour. Plan. Manag.* **131**, 449–457 (2005).
12. Lowe, D., Ebi, K. L. & Forsberg, B. Heatwave early warning systems and adaptation advice to reduce human health consequences of heatwaves. *Int. J. Environ. Res. Public Health* **8**, 4623–4648 (2011).
13. Hao, Z., Singh, V. P. & Xia, Y. Seasonal drought prediction: advances, challenges, and future prospects. *Rev. Geophys.* **56**, 108–141 (2018).
14. Wang, Y. et al. A random forest model to predict heatstroke occurrence for heatwave in China. *Sci. Total Environ.* **650**, 3048–3053 (2019).
15. Felsche, E. & Ludwig, R. Applying machine learning for drought prediction in a perfect model framework using data from a large ensemble of climate simulations. *Nat. Hazards Earth Syst. Sci.* **21**, 3679–3691 (2021).
16. Vautard, R. et al. Summertime European heat and drought waves induced by wintertime Mediterranean rainfall deficit. *Geophys. Res. Lett.* **34**, https://doi.org/10.1029/2006GL028001, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2006GL028001 (2007).
17. Quesada, B., Vautard, R., Yiou, P., Hirschi, M. & Seneviratne, S. I. Asymmetric European summer heat predictability from wet and dry southern winters and springs. *Nat. Clim. Change* **2**, 736–741 (2012).
18. Kornhuber, K., Petoukhov, V., Petri, S., Rahmstorf, S. & Coumou, D. Evidence for wave resonance as a key mechanism for generating high-amplitude quasi-stationary waves in boreal summer. *Clim. Dyn.* **49**, 1961–1979 (2017).
19. Seneviratne, S. I. et al. Investigating soil moisture–climate interactions in a changing climate: a review. *Earth Sci. Rev.* **99**, 125–161 (2010).
20. Miralles, D. G., Gentine, P., Seneviratne, S. I. & Teuling, A. J. Land–atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges. *Ann. N. Y. Acad. Sci.* **1436**, 19 (2019).
21. Schumacher, D. L. et al. Amplification of mega-heatwaves through heat torrents fuelled by upwind drought. *Nat. Geosci.* **12**, 712–717 (2019).
22. O, S. et al. The role of climate and vegetation in regulating drought-heat extremes. *J. Clim.* **1**, 1–21 (2022).
23. Horton, D. E. et al. Contribution of changes in atmospheric circulation patterns to extreme temperature trends. *Nature* **522**, 465–469 (2015).
24. Stefanon, M., D'Andrea, F. & Drobinski, P. Heatwave classification over Europe and the Mediterranean region. *Environ. Res. Lett.* **7**, 014023 (2012).
25. Della-Marta, P. M. et al. Summer heat waves over Western Europe 1880–2003, their relationship to large-scale forcings and predictability. *Clim. Dyn.* **29**, 251–275 (2007).
26. Hirschi, M. et al. Observational evidence for soil-moisture impact on hot extremes in Southeastern Europe. *Nat. Geosci.* **4**, 17–21 (2011).
27. Whan, K. et al. Impact of soil moisture on extreme maximum temperatures in Europe. *Weather Clim. Extremes* **9**, 57–67 (2015).
28. Sousa, P. M. et al. Distinct influences of large-scale circulation and regional feedbacks in two exceptional 2019 European heatwaves. *Commun. Earth Environ.* **1**, 1–13 (2020).
29. Wang, P., Tang, J., Shuyu, W., Xinning, D. & Juan, F. Regional heatwaves in China: a cluster analysis. *Clim. Dyn.* **50**, 1901–1917 (2018).
30. Katsafados, P., Papadopoulos, A., Varlas, G., Papadopoulou, E. & Mavromatidis, E. Seasonal predictability of the 2010 Russian heat wave. *Nat. Hazards Earth Syst. Sci.* **14**, 1531–1542 (2014).
31. Black, E. et al. Factors contributing to the summer 2003 European heatwave. *Weather* **59**, 217–223 (2004).
32. Böhnisch, A., Mittermeier, M., Leduc, M. & Ludwig, R. Hot spots and climate trends of meteorological droughts in Europe-Assessing the percent of normal index in a single-model initial-condition large ensemble. *Front. Water* **3** https://doi.org/10.3389/frwa.2021.716621 (2021).
33. Poschlod, B., Ludwig, R. & Sillmann, J. Ten-year return levels of sub-daily extreme precipitation over Europe. *Earth Syst. Sci. Data* **13**, 983–1003 (2021).
34. Perkins-Kirkpatrick, S. E., Fischer, E. M., Angélil, O. & Gibson, P. B. The influence of internal climate variability on heatwave frequency trends. *Environ. Res. Lett.* **12**, 044005 (2017).
35. Jeong, D. I., Yu, B. & Cannon, A. J. Links between atmospheric blocking and North American winter cold spells in two generations of Canadian earth system model large ensembles. *Clim. Dyn.* **57**, 2217–2231 (2021).
36. Deser, C. et al. Insights from earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Change* **10**, 277–286 (2020).
37. Laprise, R. et al. Climate projections over CORDEX Africa domain using the fifth-generation Canadian regional climate model (CRCM5). *Clim. Dyn.* **41**, 3219–3246 (2013).
38. Lucas-Picher, P., Laprise, R. & Winger, K. Evidence of added value in north american regional climate model hindcast simulations using ever-increasing horizontal resolutions. *Clim. Dyn.* **48**, 2611–2633 (2017).
39. Leduc, M. et al. The ClimEx Project: a 50-member ensemble of climate change projections at 12-km resolution over Europe and Northeastern North America with the Canadian Regional Climate Model (CRCM5). *J. Appl. Meteorol. Clim.* **58**, 663–693 (2019).
40. von Trentini, F., Aalbers, E. E., Fischer, E. M. & Ludwig, R. Comparing interannual variability in three regional single-model initial-condition large ensembles (SMILEs) over Europe. *Earth Syst. Dyn.* **11**, 1013–1031 (2020).
41. von Trentini, F., Leduc, M. & Ludwig, R. Assessing natural variability in RCM signals: comparison of a multi model EURO-CORDEX ensemble with a 50-member single model large ensemble. *Clim. Dyn.* **53**, 1963–1979 (2019).
42. Poschlod, B., Zscheischler, J., Sillmann, J., Wood, R. R. & Ludwig, R. Climate change effects on hydrometeorological compound events over Southern Norway. *Weather Clim. Extremes* **28**, 100253 (2020).
43. Mittermeier, M., Braun, M., Hofstätter, M., Wang, Y. & Ludwig, R. Detecting climate change effects on Vb cyclones in a 50-member single-model ensemble using machine learning. *Geophys. Res. Lett.* **46**, 14653–14661 (2019).
44. Lukasová, A. Hierarchical agglomerative clustering procedure. *Pattern Recognit.* **11**, 365–381 (1979).
45. Kodinariya, T. M. & Makwana, P. R. Review on determining number of cluster in k-means clustering. *Int. J.* **1**, 90–95 (2013).
46. Wang, G., Dolman, A. & Alessandri, A. A summer climate regime over Europe modulated by the north Atlantic oscillation. *Hydrol. Earth Syst. Sci.* **15**, 57–64 (2011).
47. López-Moreno, J. I. & Vicente-Serrano, S. M. Positive and negative phases of the wintertime North Atlantic oscillation and drought occurrence over Europe: a multitemporal-scale approach. *J. Clim.* **21**, 1220–1243 (2008).
48. Sippel, S. et al. Refining multi-model projections of temperature extremes by evaluation against land–atmosphere coupling diagnostics. *Earth Syst. Dyn.* **8**, 387–403 (2017).

49. Zampieri, M. et al. Hot European summers and the role of soil moisture in the propagation of Mediterranean drought. *J. Clim.* **22**, 4747–4758 (2009).

50. Teuling, A. et al. A regional perspective on trends in continental evaporation. *Geophys. Res. Lett.* **36**, L02404, https://doi.org/10.1029/2008GL036584 (2009).

51. Koster, R. D. et al. Regions of strong coupling between soil moisture and precipitation. *Science* **305**, 1138–1140 (2004).

52. Xu, Z.-g, Wu, Z.-y, He, H., Guo, X. & Zhang, Y.-l Comparison of soil moisture at different depths for drought monitoring based on improved soil moisture anomaly percentage index. *Water Sci. Eng.* **14**, 171–183 (2021).

53. Martynov, A. et al. Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation. *Clim. Dynam.* **41**, 2973–3005 (2013).

54. IPCC. *Climate Change 2013 - The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the IPCC*. Assessment report (Intergovernmental Panel on Climate Change): Working Group (Cambridge University Press, Cambridge, 2013).

55. Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M. & Jones, P. D. An ensemble version of the e-obs temperature and precipitation data sets. *J. Geophys. Res.* **123**, 9391–9409 (2018).

56. Dee, D. P. et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).

57. Keellings, D. & Moradkhani, H. Spatiotemporal evolution of heat wave severity and coverage across the United States. *Geophys. Res. Lett.* **47**, e2020GL087097 (2020).

58. Clemesha, R. E. S., Guirguis, K., Gershunov, A., Small, I. J. & Tardy, A. California heat waves: their spatial evolution, variation, and coastal modulation by low clouds. *Clim. Dyn.* **50**, 4285–4301 (2018).

59. Lhotka, O. & Kyselý, J. Spatial and temporal characteristics of heat waves over central Europe in an ensemble of regional climate model simulations. *Clim. Dyn.* **45**, 2351–2366 (2015).

60. Smyth, P., Ghil, M., Ide, K., Roden, J. & Fraser, A. Detecting atmospheric regimes using cross-validated clustering. In *KDD*, 61–66 (1997).

61. Machado, J. & Lopes, A. M. Rare and extreme events: the case of Covid-19 pandemic. *Nonlinear Dyn.* **100**, 2953–2972 (2020).

62. Hannachi, A., Straus, D. M., Franzke, C. L., Corti, S. & Woollings, T. Low-frequency nonlinearity and regime behavior in the northern hemisphere extratropical atmosphere. *Rev. Geophys.* **55**, 199–234 (2017).

63. Cheng, X. & Wallace, J. M. Cluster analysis of the northern hemisphere wintertime 500-hpa height field: Spatial patterns. *J. Atmos. Sci.* **50**, 2674–2696 (1993).

64. Jung, Y., Park, H., Du, D.-Z. & Drake, B. L. A decision criterion for the optimal number of clusters in hierarchical clustering. *J. Glob. Optim.* **25**, 91–111 (2003).

65. Li, Y. et al. Cluster-based data filtering for manufacturing big data systems. *J. Quality Technol.* 54:3, 290–302 (2021).

66. Koenker, R. & Hallock, K. F. Quantile regression. *J. Econ. Perspect.* **15**, 143–156 (2001).

AUTHOR CONTRIBUTIONS

E.F. designed the research under the supervision of R.L. E.F. and A.B. performed the data analysis and led the writing of the paper. E.F. prepared all figures. All authors contributed to the interpretation of the findings and revision of the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41612-023-00330-5.

Correspondence and requests for materials should be addressed to Elizaveta Felsche.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.