Received: 10 June 2022 | Revised: 26 September 2022 | Accepted: 31 October 2022 | Published on: 5 December 2022

DOI: 10.1002/qj.4394

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Subseasonal predictability of onset, duration, and intensity of European heat extremes

Maria Pyrina[1] | **Daniela I. V. Domeisen**[1,2]

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

[2]Université de Lausanne, Lausanne, Switzerland

Correspondence
M. Pyrina, Institute for Atmospheric and Climate Science, ETH Zurich, Universitätstrasse 16, 8092 Zurich, Switzerland.
Email: maria.pyrina@env.ethz.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Numbers: PP00P2_170523, PP00P2_198896

Abstract

Successful weather forecasts on subseasonal time-scales can support societal preparedness and mitigate the impacts of extreme events. Heatwaves in particular can, in certain cases, be predicted on time-scales of several weeks in advance. Heatwave predictability is commonly assessed in terms of heatwave intensity. In addition to heatwave intensity, we assess the predictability of heatwave onset and duration, which are crucial components of early-warning systems and emergency preparedness plans. The forecast skill of heatwaves is investigated over the European region in the subseasonal forecasting system of the European Centre for Medium-Range Weather Forecasts (ECMWF). The heatwaves are first detected in ERA-Interim reanalysis data over the period 1998–2017 and then allocated into six clusters in the following regions: Black Sea (BSea), Russia (Ru), Western Europe (WEu), North Sea (NSea), Scandinavia (Sc), and Eastern Europe (EEu). The European regions with the highest predictability in heatwave onset and duration are the clusters Ru, Sc, and NSea. The WEu cluster has the lowest bias in heatwave intensity and is found to be the most predictable region in terms of the number of heatwave events with predictable intensity at lead week 2. Heatwave intensity is generally found to be the most predictable characteristic of European heatwaves, being predictable by the model ensemble mean up to lead times of 3 weeks. Furthermore, this analysis identifies the most predictable heatwaves, allowing for a further investigation of the physical mechanisms and heatwave characteristics leading to enhanced heatwave forecast skill over different European regions.

KEYWORDS

ECMWF, Europe, extreme events, heatwave duration, heatwave onset, heatwaves, subseasonal predictability

1 | INTRODUCTION

Extreme summer temperatures have catastrophic impacts on a range of different sectors. For instance, heatwaves are responsible for a substantial increase in mortality and morbidity with respect to normal seasonal temperatures (Fouillet *et al.*, 2008; Li *et al.*, 2015), accounting for a high percentage of weather-related deaths in

developing and high-income countries (Nissan *et al.*, 2017). For example, the heatwave event that occurred over Western Europe in 2003 was responsible for over 70,000 deaths (Trigo *et al.*, 2005; Fouillet *et al.*, 2006; Robine *et al.*, 2008). Moreover, heatwaves have disastrous impacts other than human health. Heatwaves impact plants and ecosystems (Breshears *et al.*, 2021) and sectors such as agriculture and infrastructure (Forzieri *et al.*, 2018; Brás *et al.*, 2021). Extreme heat exacerbated 500 wildfires over Russia in 2010 and Russian grain harvests suffered a loss of 30% after the heatwave event of the same year (Barriopedro *et al.*, 2011). In the summer of 2003, more than 30 nuclear power plant units in Europe had to reduce their production because of limitations in the possibilities to discharge cooling water (Koch and Vögele, 2009; Linnerud *et al.*, 2011).

Heatwaves will continue to be relevant in the future, with anthropogenic activity imposing changes on heatwave characteristics, that is, heatwave frequency and intensity (Meehl and Tebaldi, 2004; Perkins-Kirkpatrick and Gibson, 2017). It is certain now that human influence has warmed the climate at a rate that is unprecedented in the last 2000 years and that this change is already affecting many weather and climate extremes across the globe (IPCC, 2021). Specifically, heatwaves have become more frequent and intense since the 1950s, largely due to the increasing global mean temperature (Perkins-Kirkpatrick and Lewis, 2020). Consequently, heatwave prediction will remain of high societal relevance.

Early-warning systems can reduce the risk from temperature extremes effectively at seasonal and short-term (3–10 days) prediction time-scales over Europe, as well as over other regions of the globe (De Perez *et al.*, 2018; Merz *et al.*, 2020). Disaster early warnings on the subseasonal time-scale, which consider forecasts of more than 10 days and less than a season, were also found crucial for the preparedness of society in a wide range of sectors (White *et al.*, 2017; White *et al.*, 2021). However, early-warning systems require robust forecasts of heatwave characteristics, including not only the expected heatwave intensity but also heatwave onset and duration.

The prediction of heatwave features such as heatwave onset and duration would benefit impact prediction and early warnings. Heatwave intensity has a direct impact on human health by exacerbating underlying medical conditions (Martiello and Giacchi, 2010), but mortality and morbidity increase dramatically when extreme temperatures last longer than 2 days (Pantavou *et al.*, 2008). Heatwave duration can also determine the probability of an impacted system recovering, as longer-lasting heatwaves can lead to stronger ecosystem impacts (von Buttlar *et al.*, 2018). Moreover, during a heatwave, the drying out of vegetation can significantly increase the probability of occurrence of catastrophic bushfires (Pezza *et al.*, 2012). Heatwave timing is important for ecosystems as well. For example, changes in lake-water temperature can differently affect zooplankton phenology and abundance, depending on the timing of the heatwave onset (Huber *et al.*, 2010).

While seasonal forecasts have limited skill in predicting the timing and location of observed European heatwave events, they are able to indicate if a season is predisposed to the occurrence of heatwaves (Prodhomme *et al.*, 2022). The onset and end of European heatwaves, including all small- and large-scale events, were found to be less predictable than heatwave intensity on subseasonal time-scales (Lavaysse *et al.*, 2019). An average of about 35% of observed heatwave onsets or ends were forecast correctly with a five-day lead time during the period 1995–2015 (Lavaysse *et al.*, 2019). Predictability decreases further with increasing lead time; however, extended heatwave predictability exists for large-scale events and specific case studies (Domeisen *et al.*, 2022; Vitart and Robertson, 2018).

Case studies on heatwave predictability showed that heatwaves are globally among the most predictable extreme events on subseasonal time-scales, in particular compared with cold-air outbreaks, storms, or precipitation extremes (Domeisen *et al.*, 2022). The long-range predictability of the intensity of large-scale European heatwaves on subseasonal time-scales is well known for specific heatwave events, such as the events of 2003, 2010, and 2018 (Vitart and Robertson, 2018; Kueh and Lin, 2020; Domeisen *et al.*, 2022). Moreover, large-scale heatwave predictability over different European regions was previously assessed for heatwave occurrence, but in terms of single heatwave days without considering heatwave persistence, onset, and duration (Wulff and Domeisen, 2019). Even though it is known that large-scale heatwaves are in general more predictable than small-scale events (Lavaysse *et al.*, 2019), (a) which large-scale events were most and least predictable and (b) at which subseasonal lead times their onset and duration can be predicted are currently undetermined.

This study investigates the subseasonal predictability of large-scale European heatwaves in terms of intensity, onset, and duration, and provides detailed information on the predictability of each heatwave that occurred during the study period. For this analysis, we use the European Centre for Medium-Range Weather Forecasts (ECMWF) subseasonal hindcasts for the period 1998–2017 and explore heatwave predictability for lead times of 1–3 weeks. This study is presented in four sections. After an introduction, the data and methods are described in Section 2, while in Section 3 we present and discuss the results. Finally, Section 4 presents a summary and the conclusions of the study.

2 | DATA AND METHODOLOGY

2.1 | Hindcast and verification data

We retrieved the subseasonal hindcasts from ECMWF, available through the subseasonal to seasonal (S2S) prediction project (Vitart *et al.*, 2017). The ECMWF forecast system is initialized twice per week (Mondays and Thursdays) and provides hindcasts with 11 ensemble members integrated over 46 days. The hindcasts used here cover the period 1998–2017 and use the model version of the Integrated Forecasting System (IFS) cycle 45r1 (Haiden *et al.*, 2018a) with a horizontal spectral resolution of TL511 (equivalent to a 40-km spacing on a reduced Gaussian grid: Rémy *et al.*, 2019). The hindcasts cover a period of 20 years, and are commonly used to compute the climatology and bias-correct the operational forecasts. The most recent version of the ECMWF forecasting system (IFS Cycle 47r2) includes substantial upgrades compared with the older versions (Haiden *et al.*, 2021). For example, the increase in vertical resolution from 91 to 137 levels has led to statistically significant improvements in forecast evaluation metrics of about 0.5–2% throughout most of the free atmosphere. Even though there have been changes in 2-m temperature biases due to the changes in the modeling system, the biases are relatively robust in the different model versions in terms of geographical patterns and annual and diurnal variations (Haiden *et al.*, 2018b; Sandu *et al.*, 2020). For example, the main systematic 2-m temperature forecast error over Europe in summer is an underestimation of the diurnal cycle by about 1–2 K. Given the small differences in the surface temperature biases between the newest model version and the model version used in this study, we would not expect major changes to the results found by the current analysis.

The ECMWF hindcasts for the IFS cycle 45r1 used in this study are initialized from the ERA-Interim reanalysis data (Vitart *et al.*, 2019). The ERA-Interim reanalysis data are provided with a spectral T255 horizontal resolution that corresponds to approximately 79-km spacing on a reduced Gaussian grid (Dee *et al.*, 2011). In this study, we assess the predictability of European heatwaves in the ECMWF prediction system with respect to ERA-Interim reanalysis, which has been shown to be of good quality for 2-m temperature in the global mean and over Europe (Simmons *et al.*, 2014; 2017). In the remainder of this article, we will refer to the reference data as ERA-Interim or reanalysis data and to the ECMWF prediction system as the ECMWF model or the prediction system. The reanalysis and prediction system data considered in this study are daily summer (June, July, August) 2-m temperatures over 20 years. The daily mean 2-m temperature is retrieved directly from the output of the prediction system, whereas for the reanalysis data we calculate the daily mean 2-m temperature as the daily mean of 6-hr data.

2.2 | Methods

2.2.1 | Calculation of climatology and anomalies

The ERA-Interim daily 2-m temperature (t2m) anomalies are calculated by removing the daily t2m climatology with respect to the reference period 1998–2017. Given the relatively short historical period used, daily percentile values at a given location can fluctuate due to sampling variability rather than changes in seasonally varying climate. To minimize this effect, the climatology is calculated based on a 31-day running window centered on the daily t2m value under consideration. The prediction-system t2m anomalies are computed by removing the lead-time dependent climatology using as reference period the available hindcast years, 1998–2017, consistent with the climatology period in the reanalysis. Calculating anomalies with reference to a lead-time dependent climatology is expected to remove the systematic bias due to the forecast model drift (Manzanas, 2020) and provide with a good reference climatology for bias adjustment (Manrique-Suñén *et al.*, 2020). The daily ECMWF model climatology is calculated based on a 31-day window centered around the day of the hindcast start date in question. Due to the model hindcasts being initialized twice a week (Mondays and Thursdays), to compute the lead-time dependent climatology for a Monday run, that is, using a 7-day window, we would consider the date of that Monday plus the dates of the Thursdays immediately before and after, which represents a total of 660 integrations (3 start dates × 20 years × 11 members). In this study, we consider a 31-day window, representing a total of 1,980 integrations (9 start dates × 20 years × 11 members), which are used to calculate the daily t2m lead-time dependent climatology for the model. This large window brings the benefit of defining the climatology tails, that is, the more extreme values, better, which is important for consistent computations of hot days in the prediction system. For more information refer to the ECMWF documentation.[1]

2.2.2 | Hot days and heatwave events

After the daily climatology is removed from the ERA-Interim data, the 75th percentile of the daily mean

[1] https://www.ecmwf.int/en/forecasts/documentation-and-support/extended-range/re-forecast-medium-and-extended-forecast-range

t2m anomaly is calculated applying a 31-day running window centered on the day of interest taking into account the years of the period 1998–2017. The ERA-Interim 75th percentile of the daily t2m anomaly is calculated and applied separately at each grid point of the study region and is used as a threshold to define hot days in both ERA-Interim and the ECMWF model. The same percentile threshold is used for both data and model, in order to compare simulated hot days of similar magnitude with the reanalysis data. In this way we can define a predictability threshold for average intensity that is independent of lead time. This method has the advantage of identifying events with a predictable intensity, being realistically close to observations, even for long lead times. Moreover, using percentiles of the distribution of anomalies eliminates contributions to the forecast skill resulting from a successful reproduction of the seasonal cycle. Heatwave events are defined as periods with at least three consecutive hot days, with successive heatwaves having a minimum distance of three nonhot days.

Heatwave definitions generally use temperature thresholds higher than the 90th percentile (see table 2 in Perkins, 2015). We investigate the predictability of large-scale heatwave events, meaning that hot days occur for at least three days and simultaneously for a large number of grid points within a given region. The 90th percentile threshold is too strict for detecting such events for forecast lead times longer than 1–2 weeks. For those longer lead times, it is already known from heatwave case studies that the forecast heatwave magnitude is underestimated (Vitart and Robertson, 2018; Domeisen *et al.*, 2022). In addition to separating temperature anomalies above and bellow the 75th percentile, we proceed with an additional separation of the heatwave events into two heatwave classes within the upper quartile (75th–100th percentile range) in order to study the more extreme events (further details in Section 3.2).

2.2.3 | Spatial clustering

Agglomerative hierarchical clustering (Müllner, 2013) was applied to the ERA-Interim t2m data, to define the European regions that experience above-normal temperatures simultaneously. Prior to applying the clustering algorithm, the hot days of the period 1998–2017 were set equal to 1 and the nonhot days equal to 0, so that the algorithm takes into account only the time and region of hot-day occurrence and not the hot-day intensity. In this clustering method each element (data point) initially belongs to its own cluster. Based on a chosen measure of distance (here we choose the complete linkage method), the elements are grouped into clusters until finally the desired number of clusters remains. Following the study of Stefanon *et al.* (2012), which uses agglomerative clustering for the classification of European heatwave patterns, we choose $n = 6$ as the preferred number of clusters. This choice leads to the six heatwave clusters shown in Figure 1, indicated by black dots in each subpanel. The defined clusters have a spatial extent of roughly 3,000,000 km^2 (51 grid points of 2.5° latitude by 2.5° longitude, where 1° ≈ 100 km for the regions considered here), that is, considerably larger than the 151,000 km^2 suggested by Lyon *et al.* (2019) as the minimum spatial extent for heatwaves. The large heatwave area was chosen to emphasize large-scale heatwave events, being more likely to be tied to distinct variations in atmospheric circulation and prominent impacts. Moreover, as shown later in the definition of heatwave onset, a heatwave region is also allowed to be slightly smaller than 3,000,000 km^2 as long as a minimum of 70% of the cluster's grid points experience a heatwave.

2.2.4 | Onset dates of heatwaves in ERA-Interim

The onset dates of each cluster's heatwaves are based on the ERA-Interim data set. The cluster onset dates signify the onset of a hot extreme in a cluster (i.e., first hot day of at least three consecutive hot days) when at least 70% of the cluster's grid points experience a heatwave simultaneously. This definition of heatwave onset dates allows for a more flexible location of the hot extreme, but within the same cluster and still with a large heatwave spatial extent, having a minimum of approximately 2,000,000 km^2. The onset dates of ERA-Interim heatwaves are shown in Table S1 in the Supporting Information.

2.2.5 | ECMWF model lead time

In terms of temperature anomalies and occurrence of hot days, lead week 1 represents the temperature anomalies or occurrence of hot days exactly within the model week 1, which is 1–7 days off model initialization. In the same way, lead week 2 includes the hindcast days 8–14, and lead week 3 the hindcast days 15–21. Since the ECMWF model forecasts are initialized twice per week, there is the opportunity to sample a heatwave from the model forecast that captures the heatwave onset at the first half of, for example, week 1 or from the model forecast that captures the heatwave onset in the second half of the same lead week. Here, we select for lead week 1 heatwave events with onset occurring between days 1–4 after model initialization, for lead week 2 between days 8–11, and for lead week 3 between days 15–18. We have, however, calculated

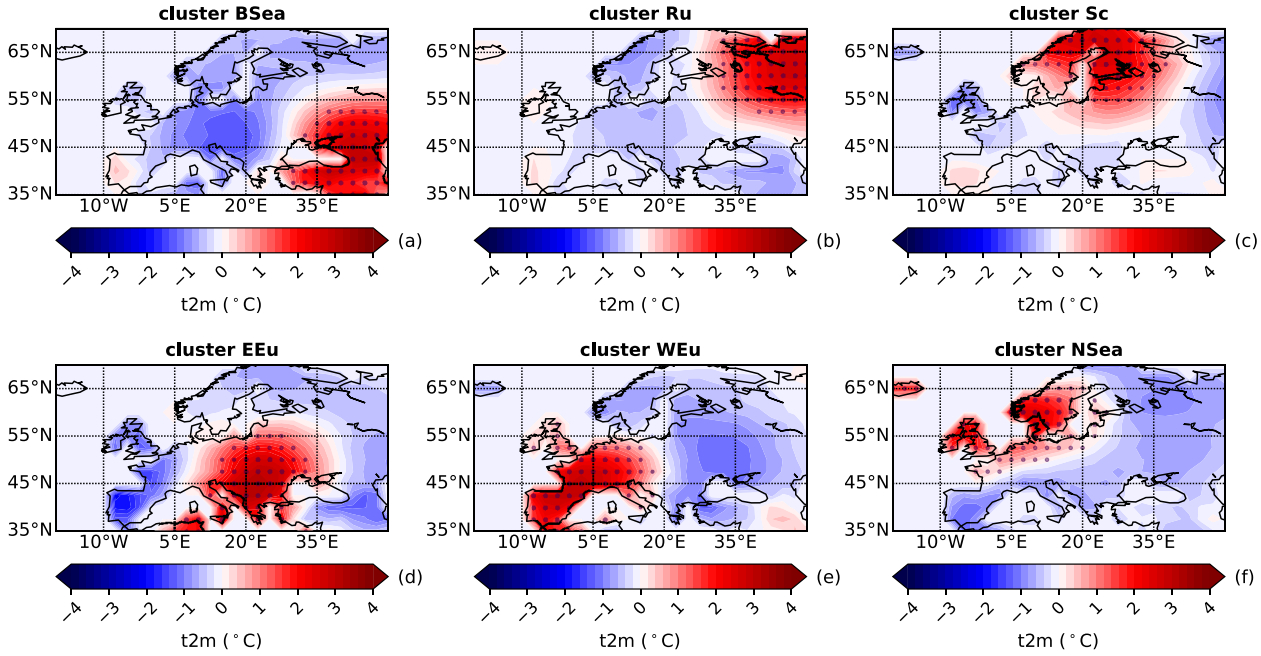

FIGURE 1 ERA-Interim summer t2m anomaly patterns obtained by hierarchical clustering for the wider European continent during 1998–2017. Daily mean summer t2m anomalies are shown for the following clusters: (a) Black Sea (BSea), (b) Russia (Ru), (c) Western Europe (WEu), (d) North Sea (NSea), (e) Scandinavia (Sc), and (f) Eastern Europe (EEu). The grid points shaded with black dots denote the area considered for each cluster [Colour figure can be viewed at wileyonlinelibrary.com]

the predictability of heatwave events with onset occurring in the second half of the respective lead week (not shown), and found that the model prediction skill falls in between the adjacent lead times. The ECMWF lead-time definition used here is also presented with a schematic in the Supporting Information (SI), Figure A1.

2.2.6 | Evaluation of heatwave intensity, duration, and onset in the ECMWF model

To evaluate the ECMWF predictability for heatwave onset, duration, and intensity separately, we choose onset, duration, and intensity definitions that are as independent from each other as possible.

The *onset* of a heatwave event is considered predicted by the ECMWF model if (a) the event starts a maximum of 2 days after the onset date shown by ERA-Interim, and (b) at least 6 out of 11 ensemble members (54%) predict the event's onset. The threshold of a maximum of 2 days after the ERA-Interim onset date is set in order to give some flexibility to the prediction system. Moreover, the threshold is set to +2 days and not ±2 days off the ERA-Interim onset date, as it is not possible for all heatwaves to select the 2 days before the ERA-Interim onset date for the lead time of week 1.

Normally, the average *intensity* of a heatwave (HW-AVI) is calculated by averaging the daily t2m for the heatwave duration. According to the definition of ECMWF model lead times we set here, at lead week 1 a heatwave that lasts five days would have its last day a minimum of 6 and at a maximum of 9 days away from model initialization. Therefore, the average intensity of events that last longer than 5 days would not be indicative of an average intensity evaluated at lead week 1. Moreover, the aim is to evaluate the model predictability of heatwave average intensity without taking into account the ability of the model to predict heatwave duration. For these reasons, the average intensity of heatwaves for either ERA-Interim or the ECMWF prediction system is calculated considering the onset date and the duration of the event shown by the ERA-Interim data (SI Table S1). Consequently, the forecast average intensity is independent of the forecast onset. In the case in which the heatwave event lasts for more than 5 days, the average intensity of the event for either ERA-Interim or the ECMWF prediction system is calculated by the average of t2m anomalies over the first 5 days of the event. If the event has a duration of 3 or 4 days, then the average intensity of the extreme event is expressed by the average of t2m anomalies over 3 or 4 days, respectively. The average intensity is integrated over all grid points of a particular cluster and is based on the model ensemble mean. Evaluating the heatwave intensity over the whole cluster rather than only over the grid points that are predicted to experience the heatwave provides an estimate of the predictability of the heat extreme over the full region.

Note that, by definition, we choose ERA-Interim heatwaves that are present during the full heatwave length in at least 70% of the cluster area.

The *duration* of heatwave events for each cluster is equal to the number of consecutive hot days with at least 70% of a cluster's grid points experiencing a hot day. If less than 70% of the cluster's grid points experience a hot day, then this day is defined as a nonhot day for the cluster. These definitions are used equally for both ERA-Interim and the ECMWF model. The duration in the ECMWF model is calculated based on the t2m anomalies of the ensemble mean. As described in Section 2.2.5, for the evaluation of heatwaves predicted at lead week 3 the onset date of heatwaves is within the first 4 days of lead week 3 (days 15–18). Therefore, at lead week 3 the prediction system can simulate a maximum heatwave duration of 28 days (46 days minus 18 days). According to the model ensemble mean, there were no events simulated with such a long duration. We evaluate how well the duration of the heatwaves shown in ERA-Interim data is predicted by the ECMWF model. In order to count the ECMWF simulated duration for the heatwaves found in ERA-Interim, (a) an ECMWF heatwave must have its onset on the indicated ERA-Interim onset date, or (b) an ECMWF heatwave may start a maximum of 2 days after the ERA-Interim onset date. If either (a) or (b) is not satisfied, then the ECMWF simulated duration for that particular heatwave is set to zero, indicating that the duration of that heatwave was not predictable by the ECMWF model.

Heatwave intensity and duration are evaluated according to the ensemble mean, whereas heatwave onset is evaluated according to the ensemble members. As described above, the duration is predictable only when the onset of the event in the ensemble mean is predicted a maximum of 2 days off the ERA-Interim onset date. Hence, the evaluation of duration indirectly provides information on how predictable the heatwave onset would be if its evaluation was based on the ensemble mean. For this reason, we have proceeded with an evaluation of heatwave onset that is based on the individual ensemble members and provides additional information on the overall performance of the ensemble members.

3 | RESULTS AND DISCUSSION

3.1 | Hot-day statistics in reanalysis and prediction system

To evaluate the ECMWF predictability of heatwaves that have spatial locations coherent with the heatwave locations indicated by the reanalysis data, we first identify in ERA-Interim the typical hot-day spatial patterns for the summers of 1998–2017 over the wider European continent (Figure 1). To check the stability of our classification we applied the same methodology to the full data period available for the ERA-Interim reanalysis data (1979–2018) and found equivalent hot-day patterns (not shown). Moreover, the hot-day patterns found in this study are very similar to the heatwave patterns obtained by Stefanon *et al.* (2012) for the Euro-Mediterranean region during the summers of 1950–2009 with the observational data set E-OBS (Klein Tank *et al.*, 2002). The grid points shaded with black dots in Figure 1 denote the representative regions of each cluster over which we evaluate the ECMWF heatwave predictability (51 grid points per cluster). van Straaten *et al.* (2020) showed that the longest t2m forecast horizons are obtained for hardly any spatial aggregation (1,158 clusters) or for full aggregation to the European scale. A comparison between the t2m forecast horizon for no spatial aggregation and for spatial aggregation with $n = 7$ (figure 5 of van Straaten *et al.*, 2020) indicates that the spatial clustering used in the current study is not expected to impact the forecast skill horizon of the t2m regional anomalies significantly.

As a first step, we compare the number of hot days found in the reanalysis (black dots) with the ECMWF forecasts at lead weeks 1–3 (Figure 2). A cluster hot day is considered to be a day with at least 70% of cluster grid points experiencing a hot day simultaneously, following the definition of cluster hot days given for heatwave onset dates and duration (Section 2.2.6). For this reason, we expect fewer hot days compared with a definition using the cluster's spatially averaged t2m anomalies and 75th percentile to define cluster hot days (SI, Figure A2). The clusters Ru and Sc have the highest number of cluster hot days in reanalysis and forecast data (≈180 hot days), while the number of hot days is lower by 50–80% for the other clusters. The Ru and Sc clusters are not under maritime influence, and are therefore expected to have generally higher climate variability with an increased number of hot days (Zschenderlein *et al.*, 2019).

The number of hot days and their distribution among the different clusters is well predicted by the ensemble mean of the forecast system at lead week 1 (left bars). The predicted hot-day number decreases significantly in lead weeks 2 and 3, indicating that for longer model lead times there are fewer hot days per grid point over Europe and only a small amount of cases where hot days occur simultaneously in at least 70% of the cluster. The decrease of hot days with lead time is more prominent in the EEu and WEu clusters. A similar result is also shown by Lavaysse *et al.* (2019) (see their figure 7), although a different method is used to transform the probabilistic ensemble prediction into a discrete prediction.

The extended-range prediction of European heatwaves tends to be related to the predictability of atmospheric

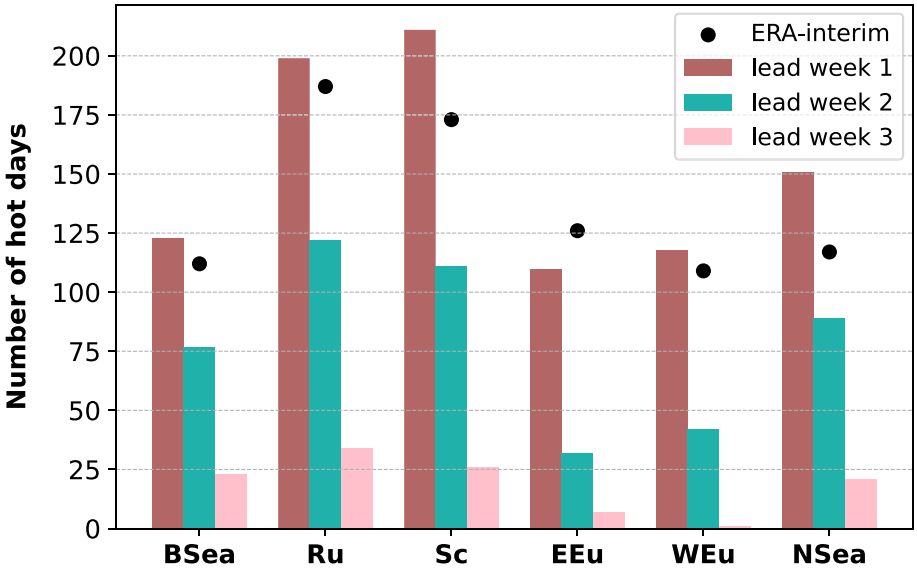

FIGURE 2 Number of hot days per cluster for the 1998–2017 summers. The number of hot days is given for ERA-Interim with a black dot and for the ECMWF prediction system for the lead times of week 1 (left bar), week 2 (middle bar), and week 3 (right bar) [Colour figure can be viewed at wileyonlinelibrary.com]

blocking or persistent weather regimes (Pfahl and Wernli, 2012; Sousa *et al.*, 2018). Whereas high near-surface temperatures over Scandinavia or Russia are often accompanied by omega-like blocking structures, heatwaves over the Mediterranean are connected to comparably flat ridges with lower magnitude (Zschenderlein *et al.*, 2019). These weaker disturbances of the zonal flow might be harder for the forecast model to predict at longer lead times, especially when the ridges relate to the occurrence of a single hot day.

To evaluate how well the magnitude of t2m anomalies for cluster hot and nonhot days can be simulated by the ECMWF ensemble mean, the mean daily summer t2m anomalies are averaged over the areas shaded with black dots for each cluster (Figure 1) using the cosine of latitude to weight the data. The ECMWF ensemble mean t2m distributions are narrower compared with the ERA-Interim t2m distribution for all clusters (Figure 3). The interquartile range becomes even narrower for longer lead times, exemplifying the lower t2m variability predicted by the forecast system's ensemble mean. Focusing on the hot extremes (upper whisker) of the ECMWF t2m anomaly distributions, it is also evident that the predicted magnitude of hot t2m anomalies for all clusters is lower at lead weeks 2 and 3 compared with lead week 1. For longer lead times, the ensemble mean is expected to reproduce lower t2m variability compared with the individual ensemble members. However, model evaluation according to the ensemble mean is a standard approach to reduce climate noise in model predictions (Kharin *et al.*, 2001).

The ERA-Interim 75th percentile of the cluster t2m anomalies is the threshold above which the t2m cluster values correspond to a hot day and is indicated by the upper edge of the ERA-Interim box (e.g., approximately equal to 1.5°C for the BSea cluster: Figure 3). The t2m anomaly values simulated by the ECMWF model that are above this threshold also correspond to a hot day, as the threshold used for reanalysis is also used for model data in this study. For this reason, the 75th percentile value shown by the upper edge of the ERA-Interim box is an indication of the lower limit of predictability of hot-day intensity for the model predictions. The ECMWF 75th percentile of hot t2m anomalies at lead week 1 is lower than in the reanalysis, indicating that there are fewer predictable hot days. The cold bias of the ECMWF 75th percentile at lead week 1 is approximately equal to 0.3, 0.2, 0.1, 0.3, 0.2, and 0.2°C for the BSea, Ru, Sc, EEu, WEu, and NSea clusters, respectively, and increases for longer lead times.

3.2 | Predictability of heatwave characteristics

3.2.1 | Predictability of heatwave intensity

To evaluate the ECMWF model predictability of heatwave average intensity (HW-AVI), the simulated average intensity of cluster heatwave events is compared with the reanalysis data (Figure 4). Figure 4 shows the distribution of HW-AVI for the heatwaves detected in ERA-Interim, computed for ERA-Interim data (left box) and for the ECMWF ensemble mean for lead times of weeks 1, 2, and 3. At lead week 1, the prediction system reproduces the width of the HW-AVI distribution shown by ERA-Interim for all clusters except for the Ru cluster, where the model's HW-AVI distribution is clearly narrower compared with the reanalysis. The ECMWF HW-AVI distribution becomes wider than in the reanalysis in lead

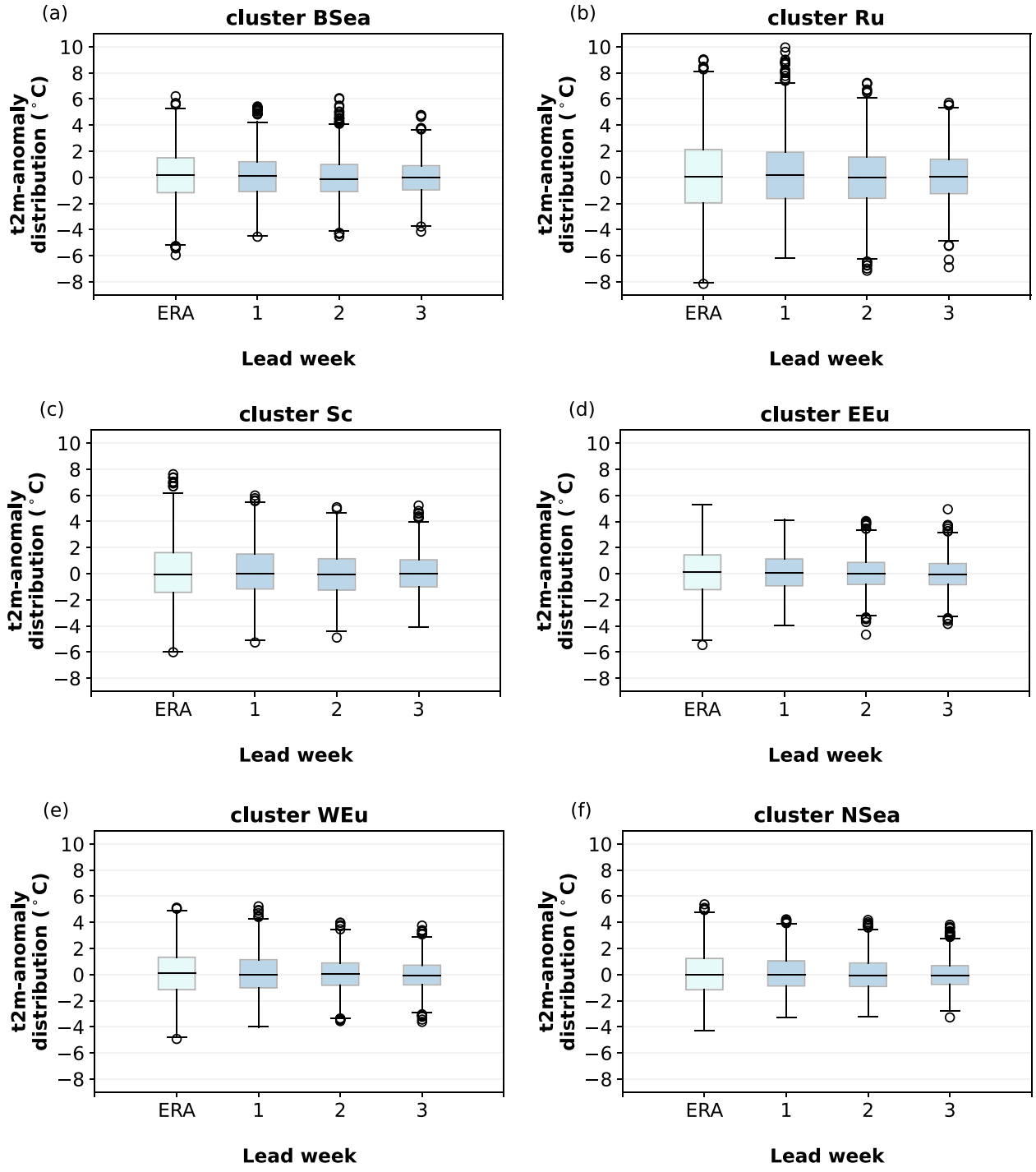

FIGURE 3 Distribution of daily t2m-anomalies for the 1998–2017 summers. The anomalies are given for ERA-Interim (left box) and the ECMWF prediction system at lead times of weeks 1, 2, and 3. The distribution's interquartile range (IQR: i.e., the range between the quartiles Q1 (25th percentile) and Q3 (75th percentile)) is shown by the edges of the boxes. The whiskers represent the minimum (Q1−1.5*IQR) and maximum (Q3+1.5*IQR) values of the distribution, being the 0.35th and 99.65th percentiles, respectively. The values below and above the whiskers show the outliers of the distribution. The distribution's median is given by a black line [Colour figure can be viewed at wileyonlinelibrary.com]

weeks 2 and 3, with negative average intensities indicating that the prediction system produces t2m anomalies below climatology. The 75th percentile of the t2m anomalies in the reanalysis spatially averaged over each full cluster (upper edge of ERA-Interim box in Figure 3 and shown in Figure 4 with a dashed black line) indicates

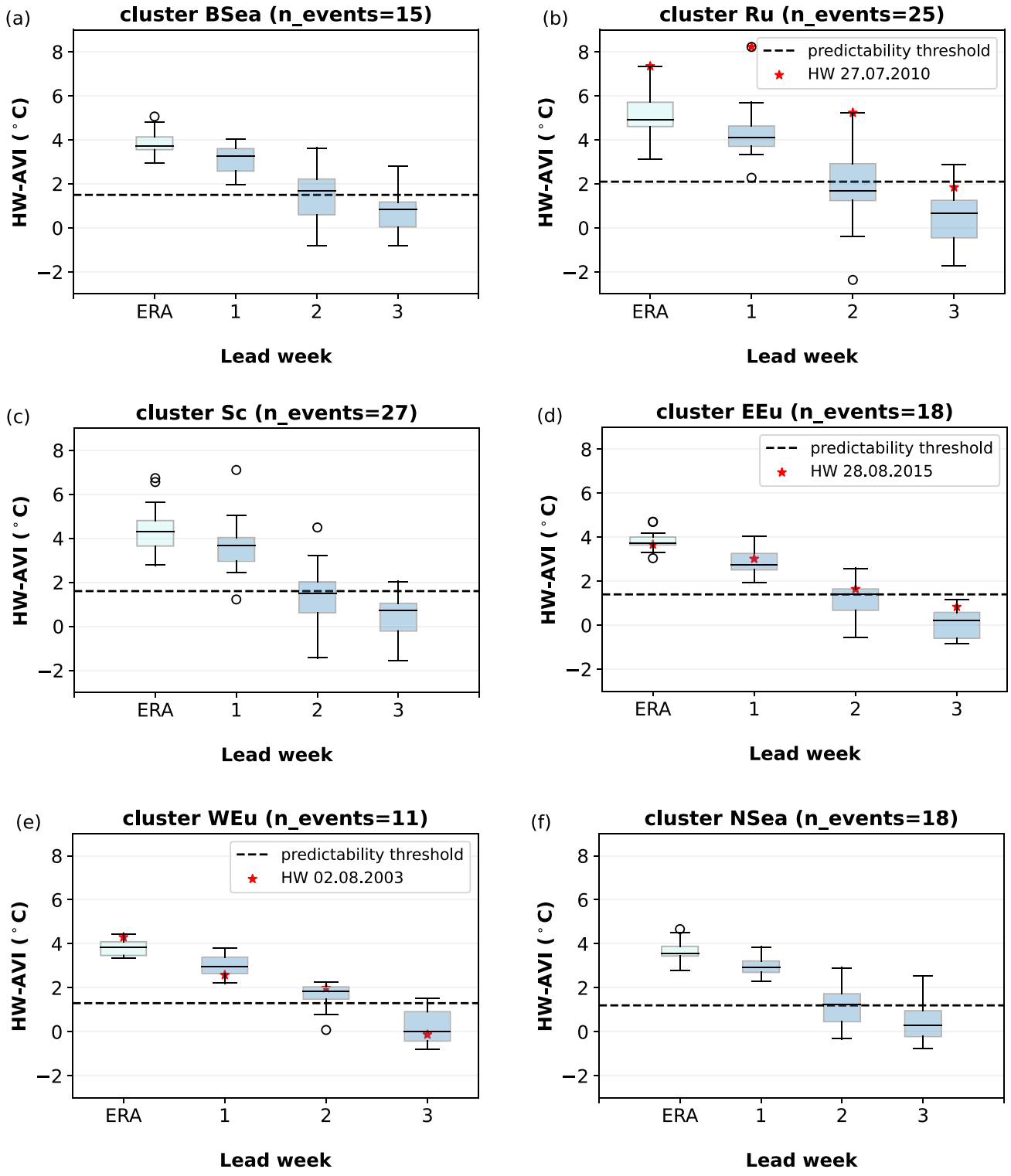

FIGURE 4 Distribution of the average intensity of the heatwaves (HW-AVI) found in ERA-interim during the summers 1998-2017, shown for ERA-interim data (left box) and for the ECMWF ensemble mean for lead times of week-1, week-2, and week-3. Details on the boxes and whiskers are given in Figure 3. The reanalysis 75th percentile calculated for the spatially averaged t2m anomalies of the full cluster is shown with a dashed black line. This line indicates the lower limit of HW-AWI predictability defined in this study. The stars in (b, d, e) denote the average heatwave intensity of the most intense (b) 2010 heatwave in the Ru cluster (start date July 27, 2010), (d) 2015 heatwave in the EEu cluster (start date August 28, 2015), and (e) 2003 heatwave in the WEu cluster (start date August 02, 2010) [Colour figure can be viewed at wileyonlinelibrary.com]

which HW-AVI values correspond to days predicted with a cluster intensity that corresponds to a heatwave intensity, as defined here. Therefore, in this study we set the dashed black line as an indication of the lower limit of HW-AVI predictability, as explained in detail in the previous subsection.

The predictability of HW-AVI is high for lead week 1 for all clusters, with a general model cold bias of around 1°C in the maximum, minimum, and median values of event distribution. The Ru cluster exhibits a stronger cold bias in the maximum values (≈2°C) and a small warm bias in the minimum values of the event distribution (≈0.3°C). For lead week 2, the HW-AVI of approximately 50% of the events lies above the lower limit of predictability in all clusters, except for the WEu cluster. The WEu cluster has the highest HW-AWI predictability, with 75% of its events crossing the predictability line (8 events). As expected from the low number of cluster hot days (Figure 2), there are few events with HW-AVI above the predictability threshold at lead week 3.

Three record-breaking heatwaves struck Europe during our study period: in 2003, 2010, and 2015 (Russo *et al.*, 2015). The HW-AVI of these heatwaves is shown with a red star for the most intense 2003 heatwave identified in the WEu cluster (start date August 2, 2003), the most intense 2010 heatwave of the Ru cluster (start date July 27, 2010), and the most intense 2015 heatwave of the EEu cluster (start date August 28, 2015). As these heatwaves had a large spatial extent and an extended duration, they can also be found in other clusters for close-by start dates (SI, Table S1). For example, the 2015 heatwave began in late June in Western Europe and then spread towards Eastern Europe (Sippel *et al.*, 2016). We find that the 2003 and 2015 events have a predictable HW-AVI for lead weeks 1 and 2 in the ECMWF forecast system, agreeing with other studies (Rodwell and Doblas-Reyes, 2006; Magnusson *et al.*, 2015; Ardilouze *et al.*, 2017). The 2010 event is predicted as an exceptionally warm event at lead times of weeks 1 and 2, also previously shown by Vitart and Robertson (2018). Moreover, these authors showed that the most intense week of the Russian heatwave event, corresponding to the dates August 1–7, was predictable by the ECMWF forecast system in terms of heatwave intensity even at lead week 3. Here, even though we evaluate the 5 days prior to August 1, Figure 4b indicates that the HW-AVI value at lead week 3 is very close to the HW-AVI predictability limit.

To understand whether the heatwave events indicated by ERA-Interim as high HW-AVI events are better predicted in terms of HW-AVI and heatwave onset compared with low HW-AVI events, we separate the ERA-Interim events into these two event classes. The first class corresponds to 50% of the highest ERA-Interim HW-AVI events and the second class to 50% of the lowest ERA-Interim HW-AVI events. The predictability of these events by the ECMWF forecast system is assessed by computing the HW-AVI relative bias (Figure 5). The forecast model at lead week 1 tends to predict better the heatwave intensities of the high HW-AVI events of clusters BSea (Figure 5a), EEu (Figure 5d), and Sc (Figure 5c), and the low HW-AVI events of clusters Ru (Figure 5b), WEu (Figure 5e), and NSea (Figure 5f). The same tendency applies for lead weeks 2 and 3 for all clusters, except for the Sc cluster, which has a lower HW-AVI bias for the low HW-AVI events at lead week 2.

The HW-AVI biases might occur due to a misrepresentation by the forecast model of the exact atmospheric circulation, soil moisture, and processes such as land–atmosphere coupling (Fischer *et al.*, 2007a; Wehrli *et al.*, 2019). For example, land–atmosphere coupling was found to increase mean and maximum temperatures averaged for anomalously warm summers, amplifying daily temperature extremes during the hottest summer days (Fischer *et al.*, 2007a). Further investigation is needed in order to understand why the ECMWF model predicts better the average intensity of either high or low HW-AVI events in specific clusters.

Even though the events of 2015 and 2010 belong to the category that is overall less well predicted in their respective clusters, both of these events were predicted with very low relative biases in their respective category. The same is not true for the WEu 2003 heatwave, which was predicted with a relative bias at the high end of the bias distribution at lead weeks 1 and 3 (−35 and −105%, respectively), and with a bias approximately equal to the median of the bias distribution at lead week 2 (−50%). A reason for that might be the high soil moisture bias over the WEu cluster (see figure 2j in Dutra *et al.*, 2021), in combination with the fact that soil moisture anomalies had a substantial impact on the intensity of the 2003 heatwave (Fischer *et al.*, 2007b). However, even though land-surface hydrology played a crucial role for the seasonal prediction of this event, to forecast the 2003 event successfully with the ECMWF forecasting system S3 (Anderson *et al.*, 2007), revised formulations of the radiative and convective model parameterizations were also required (Weisheimer *et al.*, 2011).

3.2.2 | Predictability of heatwave onset

The predictability of heatwave onset at lead week 1 (see definition in Section 2.2.6) is better for high HW-AVI events than for low HW-AVI events for all clusters (Figure 6). The higher predictability of the onset of high HW-AVI events is related to the ECMWF model's ability to predict the occurrence of hot days better compared

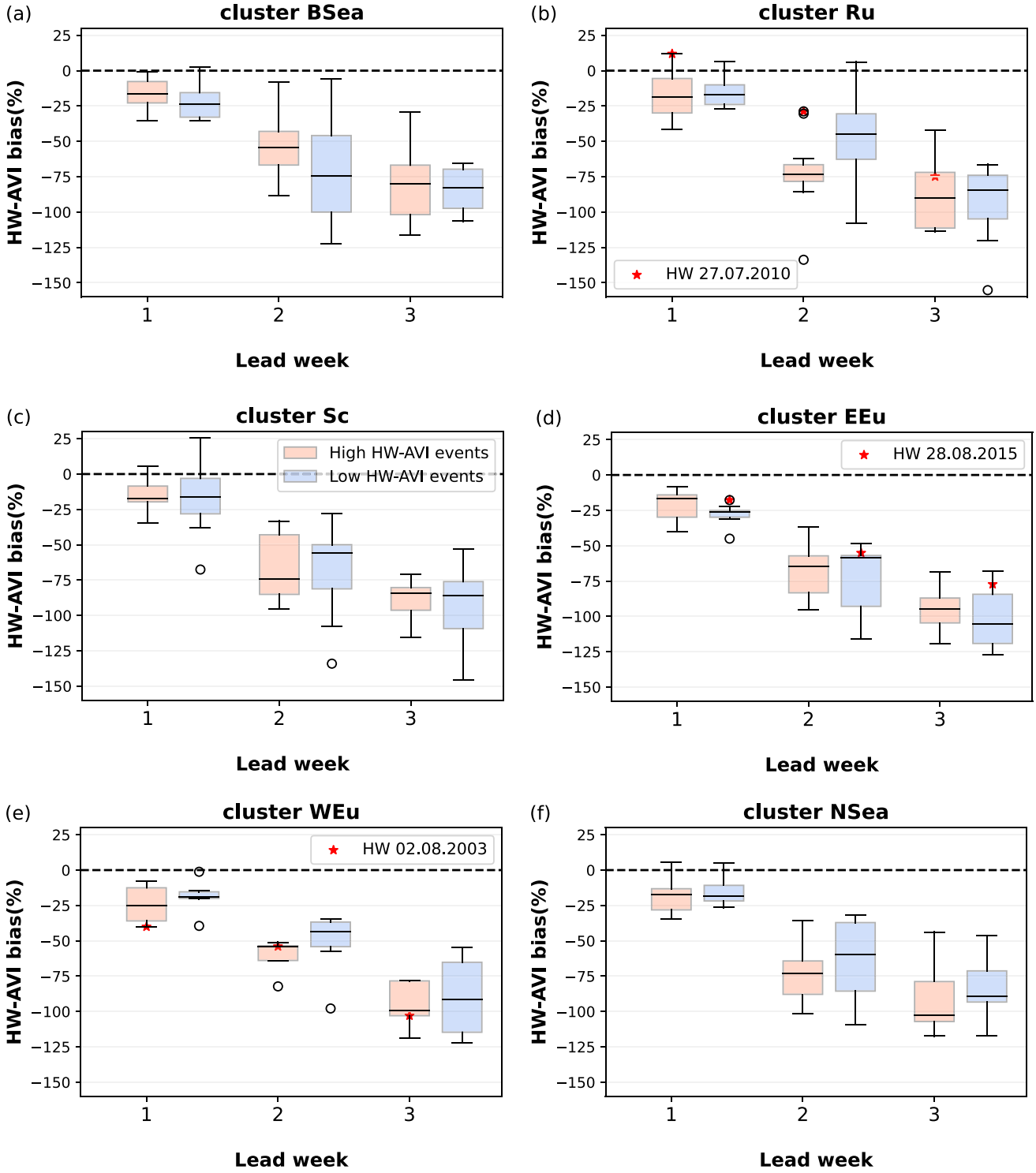

FIGURE 5 Distribution of the HW-AVI bias in percentage (ECMWF model minus reference data, divided by reference data) for the heatwaves found in ERA-interim during the summers 1998-2017, for the lead times of week-1, week-2, and week-3. The bias is given separately for the high HW-AVI (left box) and the low HW-AVI (right box) ERA-interim events. The median of the distributions is indicated by a black line. The events with HW-AVI bias above the zero line have their HW-AVI overestimated by the model. The stars in (b, d, e) denote the average heatwave intensity bias of the most intense (b) 2010 heatwave in the Ru cluster (start date July 27, 2010), (d) 2015 heatwave in the EEu cluster (start date August 28, 2015), and (e) 2003 heatwave inthe WEu cluster (start date August 02, 2010) [Colour figure can be viewed at wileyonlinelibrary.com]

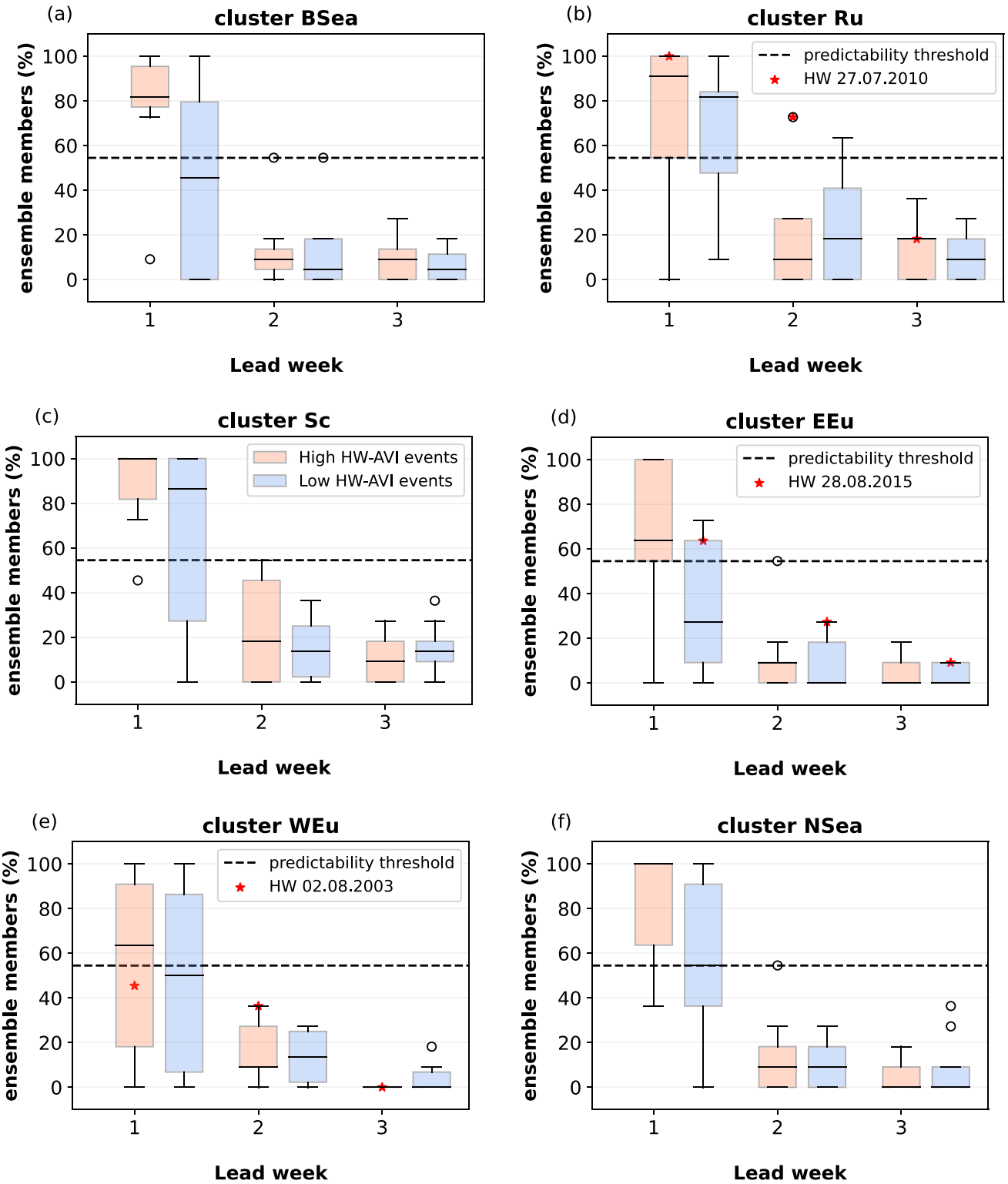

FIGURE 6 Distribution of percentage of ECMWF ensemble members predicting the onset of ERA-interim high HW-AVI events (left box) and low HW-AVI events (right box) at lead times of week-1, week-2, and week-3. The model extreme events may have an onset at the latest 2 days after the ERA start date. The dashed line denotes the lowest limit of onset predictability defined in this study, being at least 6 out of 11 ensemble members predicting the heatwave onset. The stars in (b, d, e) denote the percentage of ensemble members predicting the onset of the most intense (b) 2010 heatwave in the Ru cluster (start date July 27, 2010), (d) 2015 heatwave in the EEu cluster (start date August 28, 2015), and (e) 2003 heatwave in the WEu cluster (start date August 02, 2010) [Colour figure can be viewed at wileyonlinelibrary.com]

with average t2m days. Higher predictability for the occurrence of extreme hot days compared with the occurrence of days with average temperature was previously shown by Wulff and Domeisen (2019). Upper-level ridges or blocks are commonly located above the regions experiencing the heatwave and are found to be directly related to heatwave onset (Pfahl and Wernli, 2012; Sousa *et al.*, 2018; Zschenderlein *et al.*, 2019). Consequently, the percentage of ensemble members predicting the heatwave onset could be related to the strength and location of the simulated ridge or block that is responsible for the heatwave.

In general, the onset of heatwaves is predicted comparably worse in the EEu and WEu clusters, possibly due to the sensitivity of heatwave-related dynamical processes to small-scale microphysical and convective processes important for heatwave predictions on subseasonal time-scales (Zschenderlein *et al.*, 2020). For example, the influence of diabatic heating is larger for heatwave-related anticyclones in northern Europe and western Russia and smaller in southern Europe (Zschenderlein *et al.*, 2020).

Regarding the predictability of heatwave onset at lead weeks 2 and 3, there is no clear distinction between high and low HW-AVI events, as the number of events with predicted onset is very low for these lead times (0–5 events in total, depending on the cluster). Varying the threshold set for a predictable heatwave onset by the model for lead weeks 2 and 3 from +2 days to the range between ±2 days and ±5 days does not noticeably change the probability of the model capturing heatwave onset with more than five ensemble members (SI Figures A3, A4). This is expected, as the number of cluster hot days is low for these lead times (Figure 2).

3.2.3 | Predictability of heatwave duration

The ability of the ECMWF model to simulate the duration of ERA-Interim heatwaves is shown in Figure 7. As described in Section 2.2, an event has a predictable duration only when the onset of the event in the ensemble mean is predicted a maximum of 2 days after the ERA-Interim onset date. More than half of the ERA-Interim heatwaves have a predictable duration at lead week 1, for all clusters. Specifically, 60% (9/15) of events have a predictable duration in the cluster BSea, 80% (20/25) in Ru, 78% (21/27) in Sc, 60% (11/18) in EEu, 73% (8/11) in WEu, and 83% (15/18) in NSea. Moreover, the duration distributions between model and reanalysis at lead week 1 are similar, with the values of maximum duration ranging from 10–18 days depending on the cluster. However, the model cannot reproduce the long-duration events of the WEu cluster. Specifically, there is no event with a duration longer than 9 days simulated by the model in the WEu cluster at lead week 1.

The events of 2003 in WEu, 2010 in Ru, and 2015 in EEu, were predicted at lead week 1 with durations equal to 6 days (13 days in ERA), 11 days (17 days in ERA), and 5 days (4 days in ERA), respectively. Among these three extreme events, only the 2010 Ru event has a predictable duration at lead week 2, which is equal to 7 days. Generally, the duration of only a few events is simulated by the model ensemble mean for lead week 2 and of only one event in the NSea cluster for lead week 3. The evolution of heatwaves, and by extension their duration, is affected by land–atmosphere coupling through both local and remote effects (Fischer *et al.*, 2007a). As defined in this study, a predictable duration is related to the skill of the model to predict high t2m over 70% of the spatial cluster. Higher boreal summer t2m prediction skill was demonstrated by an accurate soil-moisture initialization (Seo *et al.*, 2019). However, the improvement by soil-moisture initialization is limited to specific regions where the large-scale wavetrain-like teleconnection patterns driving the heatwaves wane in the model forecasts.

The predictability of the duration is dependent on the way duration is assessed. As an alternative to the definition of duration in this study, the duration can be evaluated without requiring the ensemble mean to produce a hot day, but only requiring a certain percentage of ensemble members. In this case the duration, as well as the HW-AVI of the event, could be calculated either according to the ensemble mean of the ensemble members that produce a hot day or probabilistically. Here we choose to evaluate duration according to the ensemble mean, to be consistent with the evaluation of HW-AVI, which is commonly shown as the mean over all ensemble members (Vitart and Robertson, 2018; Wulff and Domeisen, 2019).

As seen previously in the evaluation of onset, where predictability is assessed probabilistically, there are extremely few cases at lead weeks 2 and 3 with at least 6 out of 11 ensemble members predicting the onset of a heatwave. It might be that, for these lead times, using fewer ensemble members (30%) might already be indicative of capturing the onset of a heatwave (Lavaysse *et al.*, 2019). However, even if we consider onsets as predictable when 30% of the ensemble members capture them, that still only accounts for 1–2 more predictable events in the EEu and WEu clusters and 2–4 more predictable events in the Ru and Sc clusters for lead week 2 (Figure 6). Regarding lead week 1, preselecting the ensemble members that produce a hot day is expected to increase the prediction skill of the model for HW-AVI and possibly increase the number of HWs with predictable duration.

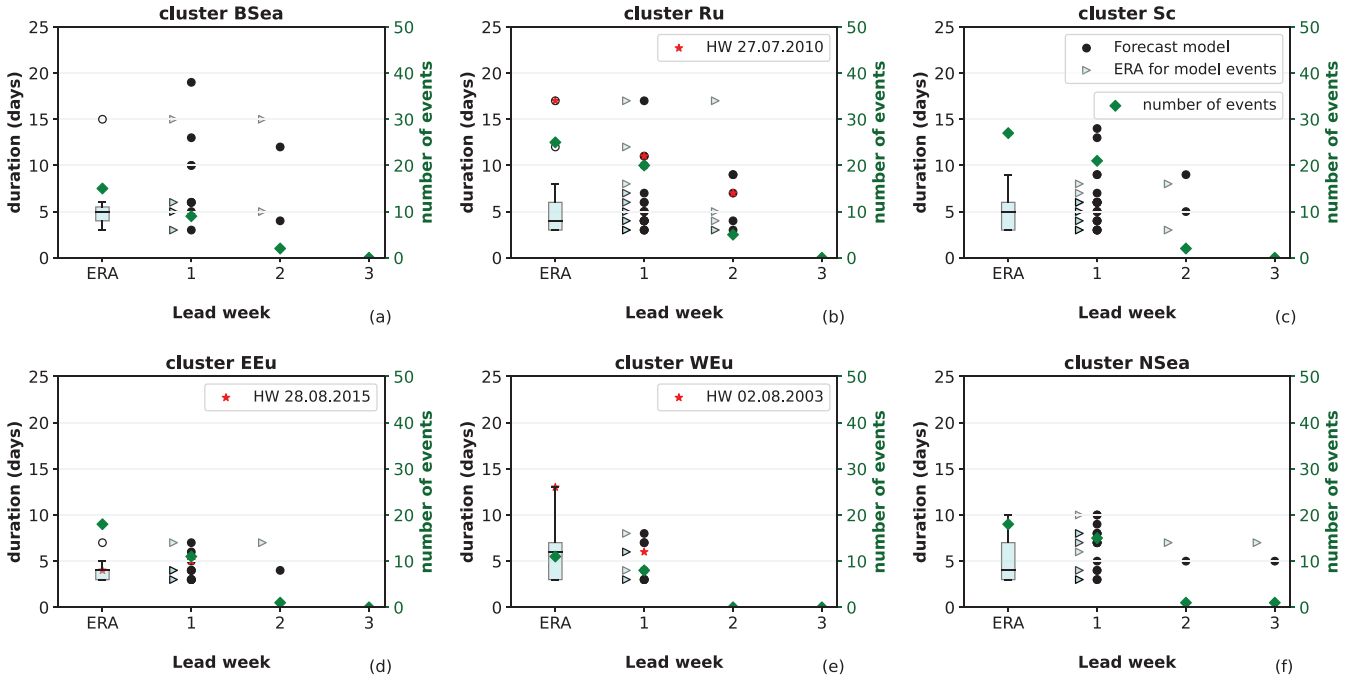

FIGURE 7 Duration of ERA-interim heatwaves for all events in ERA-interim (boxplot) and for the part of the ERA-interim events with simulated duration by the ECMWF ensemble mean for lead times of week-1, week-2 and week-3 (right triangles). The duration of the ECMWF events at different lead times is shown by the black dots. For each lead time and each cluster, the number of right triangles is equal to the number of black dots, as they correspond to the same heatwaves (those with predictable duration). The diamond (right-hand *y*-axis) denotes the total number of ERA-interim extreme events and the number of ERA-interim events simulated by the ECMWF model at lead times of week-1, week-2, and week-3. The stars in (b, d, e) denote the ERA-interim and ECMWF heatwave duration for the most intense (b) 2010 heatwave in the Ru cluster (start date July 27, 2010), (d) 2015 heatwave in the EEu cluster (start date August 28, 2015), and (e) 2003 heatwave in the WEu cluster (start date August 02, 2010) [Colour figure can be viewed at wileyonlinelibrary.com]

3.2.4 | Most predictable European heatwaves

The ERA-Interim number of heatwave events per cluster, together with the heatwave onset dates, HW-AVI, and heatwave duration, is given in the SI Table S1. The most predictable events in terms of HW-AVI are marked with an asterisk when their HW-AVI was predicted as far as 2 weeks lead time and with a red I when their HW-AVI was predicted as far as 3 weeks lead time. Moreover, the events marked with a green L13 have their HW-AVI predicted at lead weeks 1 and 3. There are several events in each cluster with predicted HW-AVI at lead times of more than a week, starting with a minimum of 50% of events in the Sc cluster and reaching a maximum of 82% of events in the WEu cluster. These events include the most extreme 2003, 2010, and 2015 heatwaves identified in the WEu, Ru, and EEu clusters, respectively. However, not all identified 2010 and 2015 heatwave events were predictable at long lead times (e.g., the Ru event with start date July 6, 2010), highlighting the need to understand what drives inter-event differences in predictability within the same cluster.

The events with the highest predictability, meaning that all their characteristics (HW-AVI, onset, duration) were predicted as far as 2 weeks lead time, are few. These events are highlighted by grey shading in Table S1. The most predictable events took place during years 2010 and 2016 in BSea, 2001, 2010, 2014, and 2017 in Ru, 2013 and 2014 in Sc, 2012 in EEu, and 2007 in NSea. In contrast, there is a higher number of events with poorly predicted HW-AVI at lead week 2. Note that HW-AVI is the most predictable heatwave characteristic. The least predictable events include, for example, all events of 2006 and 2016 in the Ru cluster and the event of 2017 in WEu. The 2017 WEu heatwave was characterised as the earliest European summer mega-heatwave (Sánchez-Benítez *et al.*, 2018).

4 | SUMMARY AND CONCLUSIONS

The ECMWF forecast system is used to evaluate the subseasonal predictability of large-scale heatwaves over six European regions during the period 1998–2017. The predictability of three heatwave characteristics, that is, the onset, duration, and average intensity, is assessed for lead times of 1–3 weeks.

The forecast model's tendency is to predict with low relative bias the average intensity of the high-intensity heatwaves that occurred in the Black Sea, East Europe, and Scandinavia clusters and the low-intensity heatwaves of the Russia, West Europe, and North Sea clusters. However, the onset of high-intensity events is consistently better predicted than the onset of low-intensity events for all clusters. The European regions experiencing the highest predictability of onset and duration are the clusters Russia, Scandinavia, and North Sea. The West Europe cluster is the most predictable in terms of the number of heatwave events that at lead week 2 have predictable average intensity, as defined in this study.

The evaluation of heatwave average intensity does not depend on the prediction of heatwave duration in this study. Therefore, average intensity is found to be the most predictable characteristic of European heatwaves. Heatwave average intensity is predictable by the model ensemble mean up to lead times of 3 weeks, even in the case when the onset (probabilistic evaluation) and duration (evaluation based on ensemble mean) of the particular heatwave are not considered to be predicted successfully. Taking into account only events with predictable duration, we find that the forecast system follows the duration distribution well for all clusters at lead week 1. However, the prediction of duration at lead weeks 2 and 3 is much more challenging. The same finding applies for the predictability of heatwave onset.

We expect that the subseasonal predictability of heatwaves found in the current study is similar for other subseasonal forecasting systems. For example, an investigation of the predictability of heatwave occurrence over Europe for the period 1999–2010 showed that four forecasting systems, including the ECMWF prediction model used in the current study, had small differences in forecast skill at subseasonal lead times (Wulff and Domeisen, 2019).

The midlatitude circulation is difficult to predict, being dominated by unpredictable weather noise. However, we have shown that subseasonal forecasts from the ECMWF forecast system can predict the intensity, onset, and duration of several heatwaves, including the extreme heatwaves of 2003, 2010, and 2015. Given the importance of predicting such strong extreme events, it is crucial that the processes leading to their predictability are investigated and consequently modelled correctly in state-of-the-art subseasonal forecasting models. This study identified the most and least predictable heatwaves, which sets the basis for a further investigation of the causes of differences in predictability for different heatwave events in Europe.

AUTHOR CONTRIBUTIONS

Maria Pyrina: conceptualization; data curation; formal analysis; investigation; methodology; software; validation; visualization; writing – original draft; writing – review and editing. **Daniela I. V. Domeisen:** conceptualization; funding acquisition; project administration; resources; supervision; writing – review and editing.

ACKNOWLEDGEMENTS

This work was conducted as part of the projects PP00P2_170523 and PP00P2_198896 funded by the Swiss National Science Foundation.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The ERA-Interim reanalysis was obtained from the ECMWF server at https://apps.ecmwf.int/datasets/data/interim-full-daily. The ECMWF S2S hindcast simulations are publicly accessible at https://apps.ecmwf.int/datasets/data/s2s.

ORCID

Maria Pyrina https://orcid.org/0000-0002-4890-0732

REFERENCES

Anderson, D., Stockdale, T., Balmaseda, M., Ferranti, L., Vitart, F., Molteni, F., Doblas-Reyes, F., Mogensen, K. and Vidard, A. (2007) Development of the ECMWF seasonal forecast system 3. ECMWF Technical Memoranda. Reading: ECMWF. 503.

Ardilouze, C., Batté, L. and Déqué, M. (2017) Subseasonal-to-seasonal (S2S) forecasts with CNRM-CM: a case study on the July 2015 west-European heat wave. *Advances in Science and Research*, 14, 115–121.

Barriopedro, D., Fischer, E.M., Luterbacher, J., Trigo, R.M. and García-Herrera, R. (2011) The hot summer of 2010: redrawing the temperature record map of Europe. *Science*, 332, 220–224.

Brás, T.A., Seixas, J., Carvalhais, N. and Jägermeyr, J. (2021) Severity of drought and heatwave crop losses tripled over the last five decades in Europe. *Environmental Research Letters*, 16, 065012.

Breshears, D.D., Fontaine, J.B., Ruthrof, K.X., Field, J.P., Feng, X., Burger, J.R., Law, D.J., Kala, J. and Hardy, G.E.S.J. (2021) Underappreciated plant vulnerabilities to heat waves. *New Phytologist*, 231, 32–39.

von Buttlar, J., Zscheischler, J., Rammig, A., Sippel, S., Reichstein, M., Knohl, A., Jung, M., Menzer, O., Arain, M.A., Buchmann, N., Cescatti, A., Gianelle, D., Kiely, G., Law, B.E., Magliulo, V., Margolis, H., McCaughey, H., Merbold, L., Migliavacca, M., Montagnani, L., Oechel, W., Pavelka, M., Peichl, M., Rambal, S., Raschi, A., Scott, R.L., Vaccari, F.P., van Gorsel, E., Varlagin, A. and Wohlfahrt, G. (2018) Impacts of droughts and extreme-temperature events on gross primary production and ecosystem respiration: a systematic assessment across ecosystems and climate zones. *Biogeosciences*, 15, 1293–1318.

De Perez, E.C., Van Aalst, M., Bischiniotis, K., Mason, S., Nissan, H., Pappenberger, F., Stephens, E., Zsoter, E. and Van Den Hurk, B. (2018) Global predictability of temperature extremes. *Environmental Research Letters*, 13, 054017.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., 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., Holm, E.V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, 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.

Domeisen, D.I., White, C.J., Afargan-Gerstman, H., Muñoz, G., Janiga, M.A., Vitart, F., Wulff, C.O., Antoine, S., Ardilouze, C., Batté, L. and Bloomfield, H.C. (2022) Advances in the subseasonal prediction of extreme events: Relevant case studies across the globe. *Bulletin of the American Meteorological Society*, 103, E1473–E1501.

Dutra, E., Johannsen, F. and Magnusson, L. (2021) Late spring and summer subseasonal forecasts in the northern hemisphere midlatitudes: biases and skill in the ECMWF model. *Monthly Weather Review*, 149, 2659–2671.

Fischer, E.M., Seneviratne, S.I., Lüthi, D. and Schär, C. (2007a) Contribution of land–atmosphere coupling to recent European summer heat waves. *Geophysical Research Letters*, 34, L06707.

Fischer, E.M., Seneviratne, S.I., Vidale, P.L., Lüthi, D. and Schär, C. (2007b) Soil moisture–atmosphere interactions during the 2003 European summer heat wave. *Journal of Climate*, 20, 5081–5099.

Forzieri, G., Bianchi, A., e Silva, F.B., Herrera, M.A.M., Leblois, A., Lavalle, C., Aerts, J.C.H. and Feyen, L. (2018) Escalating impacts of climate extremes on critical infrastructures in Europe. *Global Environmental Change*, 48, 97–107.

Fouillet, A., Rey, G., Laurent, F., Pavillon, G., Bellec, S., Guihenneuc-Jouyaux, C., Clavel, J., Jougla, E. and Hémon, D. (2006) Excess mortality related to the August 2003 heat wave in France. *International Archives of Occupational and Environmental Health*, 80, 16–24.

Fouillet, A., Rey, G., Wagner, V., Laaidi, K., Empereur-Bissonnet, P., Le Tertre, A., Frayssinet, P., Bessemoulin, P., Laurent, F., De Crouy-Chanel, P. and Jougla, E. (2008) Has the impact of heat waves on mortality changed in France since the European heat wave of summer 2003? A study of the 2006 heat wave. *International Journal of Epidemiology*, 37, 309–317.

Haiden, T., Janousek, M., Bidlot, J., Buizza, R., Ferranti, L., Prates, F. and Vitart, F. (2018a) Evaluation of ECMWF forecasts, including the 2018 upgrade. ECMWF Technical Memoranda. Reading: ECMWF. 831.

Haiden, T., Janousek, M., Vitart, F., Ben-Bouallegue, Z., Ferranti, L. and Prates, F. (2021) Evaluation of ECMWF forecasts, including the 2021 upgrade. ECMWF Technical Memoranda. Reading: ECMWF. 884.

Haiden, T., Sandu, I., Balsamo, G., Arduini, G. and Beljaars, A. (2018b) Addressing biases in near-surface forecasts. *ECMWF Newsletter*, 157, 20–25.

Huber, V., Adrian, R. and Gerten, D. (2010) A matter of timing: heat wave impact on crustacean zooplankton. *Freshwater Biology*, 55, 1769–1779.

IPCC (2021) Summary for policymakers. In: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R. and Zhou, B. (Eds.) *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 3–32). Cambridge, Newyork: Cambridge university press.

Kharin, V., Zwiers, F. and Gagnon, N. (2001) Skill of seasonal hindcasts as a function of the ensemble size. *Climate Dynamics*, 17, 835–843.

Klein Tank, A., Wijngaard, J., Können, G., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C. and Heino, R. (2002) Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessment. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 22, 1441–1453.

Koch, H. and Vögele, S. (2009) Dynamic modelling of water demand, water availability and adaptation strategies for power plants to global change. *Ecological Economics*, 68, 2031–2039.

Kueh, M.-T. and Lin, C.-Y. (2020) The 2018 summer heatwaves over northwestern Europe and its extended-range prediction. *Scientific Reports*, 10, 1–18.

Lavaysse, C., Naumann, G., Alfieri, L., Salamon, P. and Vogt, J. (2019) Predictability of the European heat and cold waves. *Climate Dynamics*, 52, 2481–2495.

Li, M., Gu, S., Bi, P., Yang, J. and Liu, Q. (2015) Heat waves and morbidity: current knowledge and further direction—a comprehensive literature review. *International Journal of Environmental Research and Public Health*, 12, 5256–5283.

Linnerud, K., Mideksa, T.K. and Eskeland, G.S. (2011) The impact of climate change on nuclear power supply. *The Energy Journal*, 32, 149–168.

Lyon, B., Barnston, A.G., Coffel, E. and Horton, R.M. (2019) Projected increase in the spatial extent of contiguous US summer heat waves and associated attributes. *Environmental Research Letters*, 14, 114029.

Magnusson, L., Thorpe, A., Buizza, R., Rabier, F. and Nicolau, J. (2015) Predicting this year's European heat wave. *ECMWF Newsletter*, 145, 4–5.

Manrique-Suñén, A., Gonzalez-Reviriego, N., Torralba, V., Cortesi, N. and Doblas-Reyes, F.J. (2020) Choices in the verification of S2S forecasts and their implications for climate services. *Monthly Weather Review*, 148, 3995–4008.

Manzanas, R. (2020) Assessment of model drifts in seasonal forecasting: sensitivity to ensemble size and implications for bias correction. *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001751.

Martiello, M.A. and Giacchi, M.V. (2010) High temperatures and health outcomes: a review of the literature. *Scandinavian Journal of Public Health*, 38, 826–837.

Meehl, G.A. and Tebaldi, C. (2004) More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, 305, 994–997.

Merz, B., Kuhlicke, C., Kunz, M., Pittore, M., Babeyko, A., Bresch, D.N., Domeisen, D.I., Feser, F., Koszalka, I., Kreibich, H. and Pantillon, F. (2020) Impact forecasting to support emergency

management of natural hazards. *Reviews of Geophysics*, 58, e2020RG000704.

Müllner, D. (2013) fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python. *Journal of Statistical Software*, 53, 1–18.

Nissan, H., Burkart, K., Coughlan de Perez, E., Van Aalst, M. and Mason, S. (2017) Defining and predicting heat waves in Bangladesh. *Journal of Applied Meteorology and Climatology*, 56, 2653–2670.

Pantavou, K., Theoharatos, G., Nikolopoulos, G., Katavoutas, G. and Asimakopoulos, D. (2008) Evaluation of thermal discomfort in Athens territory and its effect on the daily number of recorded patients at hospitals' emergency rooms. *International Journal of Biometeorology*, 52, 773–778.

Perkins, S.E. (2015) A review on the scientific understanding of heatwaves—their measurement, driving mechanisms, and changes at the global scale. *Atmospheric Research*, 164, 242–267.

Perkins-Kirkpatrick, S. and Gibson, P. (2017) Changes in regional heatwave characteristics as a function of increasing global temperature. *Scientific Reports*, 7, 1–12.

Perkins-Kirkpatrick, S. and Lewis, S. (2020) Increasing trends in regional heatwaves. *Nature Communications*, 11, 1–8.

Pezza, A.B., Van Rensch, P. and Cai, W. (2012) Severe heat waves in southern Australia: synoptic climatology and large scale connections. *Climate Dynamics*, 38, 209–224.

Pfahl, S. and Wernli, H. (2012) Quantifying the relevance of atmospheric blocking for co-located temperature extremes in the northern hemisphere on (sub-) daily time-scales. *Geophysical Research Letters*, 39, L12807.

Prodhomme, C., Materia, S., Ardilouze, C., White, R.H., Batté, L., Guemas, V., Fragkoulidis, G. and García-Serrano, J. (2022) Seasonal prediction of European summer heatwaves. *Climate Dynamics*, 58, 2149–2166.

Rémy, S., Kipling, Z., Flemming, J., Boucher, O., Nabat, P., Michou, M., Bozzo, A., Ades, M., Huijnen, V., Benedetti, A. and Engelen, R. (2019) Description and evaluation of the tropospheric aerosol scheme in the European Centre for medium-range weather forecasts (ECMWF) integrated forecasting system (IFS-AER, cycle 45R1). *Geoscientific Model Development*, 12, 4627–4659.

Robine, J.-M., Cheung, S.L.K., Le Roy, S., Van Oyen, H., Griffiths, C., Michel, J.-P. and Herrmann, F.R. (2008) Death toll exceeded 70,000 in Europe during the summer of 2003. *Comptes Rendus Biologies*, 331, 171–178.

Rodwell, M.J. and Doblas-Reyes, F.J. (2006) Medium-range, monthly, and seasonal prediction for Europe and the use of forecast information. *Journal of Climate*, 19, 6025–6046.

Russo, S., Sillmann, J. and Fischer, E.M. (2015) Top ten European heatwaves since 1950 and their occurrence in the coming decades. *Environmental Research Letters*, 10, 124003.

Sánchez-Benítez, A., García-Herrera, R., Barriopedro, D., Sousa, P.M. and Trigo, R.M. (2018) June 2017: the earliest European summer mega-heatwave of reanalysis period. *Geophysical Research Letters*, 45, 1955–1962.

Sandu, I., Haiden, T., Balsamo, G., Schmederer, P., Arduini, G., Day, J., Beljaars, A., Ben-Bouallegue, Z., Boussetta, S., Leutbecher, M., Magnusson, L. and de Rosnay, P. (2020) *Addressing near-surface forecast biases: outcomes of the ECMWF project 'Understanding Uncertainties in Surface Atmosphere Exchange'(USURF). ECMWF Technical Memoranda. Reading: ECMWF. 875.*

Seo, E., Lee, M.-I., Jeong, J.-H., Koster, R.D., Schubert, S.D., Kim, H.-M., Kim, D., Kang, H.-S., Kim, H.-K., MacLachlan, C. and Scaife, A.A. (2019) Impact of soil-moisture initialization on boreal summer subseasonal forecasts: mid-latitude surface air temperature and heat wave events. *Climate Dynamics*, 52, 1695–1709.

Simmons, A., Berrisford, P., Dee, D., Hersbach, H., Hirahara, S. and Thépaut, J.-N. (2017) A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets. *Quarterly Journal of the Royal Meteorological Society*, 143, 101–119.

Simmons, A., Poli, P., Dee, D., Berrisford, P., Hersbach, H., Kobayashi, S. and Peubey, C. (2014) Estimating low-frequency variability and trends in atmospheric temperature using Era-Interim. *Quarterly Journal of the Royal Meteorological Society*, 140, 329–353.

Sippel, S., Otto, F.E., Flach, M. and van Oldenborgh, G.J. (2016) 11. The role of anthropogenic warming in 2015 central European heat waves. *Bulletin of the American Meteorological Society*, 97, S51–S56.

Sousa, P.M., Trigo, R.M., Barriopedro, D., Soares, P.M. and Santos, J.A. (2018) European temperature responses to blocking and ridge regional patterns. *Climate Dynamics*, 50, 457–477.

Stefanon, M., D'Andrea, F. and Drobinski, P. (2012) Heatwave classification over Europe and the Mediterranean region. *Environmental Research Letters*, 7, 014023.

van Straaten, C., Whan, K., Coumou, D., van den Hurk, B. and Schmeits, M. (2020) The influence of aggregation and statistical post-processing on the subseasonal predictability of European temperatures. *Quarterly Journal of the Royal Meteorological Society*, 146, 2654–2670.

Trigo, R.M., García-Herrera, R., Díaz, J., Trigo, I.F. and Valente, M.A. (2005) How exceptional was the early August 2003 heatwave in France?. *Geophysical Research Letters*, 32, L10701.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M. and Hendon, H. (2017) The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, 98, 163–173.

Vitart, F., Balmaseda, M., Ferranti, L., Benedetti, A., Sarojini, B., Tietsche, S., Yao, J., Janousek, M., Balsamo, G., Bechtold, P. and Leutbecher, M. (2019) Extended-range prediction. ECMWF Technical Memoranda. Reading: ECMWF. 854.

Vitart, F. and Robertson, A.W. (2018) The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Climate and Atmospheric Science*, 1, 1–7.

Wehrli, K., Guillod, B.P., Hauser, M., Leclair, M. and Seneviratne, S.I. (2019) Identifying key driving processes of major recent heat waves. *Journal of Geophysical Research: Atmospheres*, 124, 11746–11765.

Weisheimer, A., Doblas-Reyes, F.J., Jung, T. and Palmer, T. (2011) On the predictability of the extreme summer 2003 over Europe. *Geophysical Research Letters*, 38, L05704.

White, C.J., Carlsen, H., Robertson, A.W., Klein, R.J., Lazo, J.K., Kumar, A., Vitart, F., Coughlan de Perez, E., Ray, A.J., Murray, V. and Bharwani, S. (2017) Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications*, 24, 315–325.

White, C.J., Domeisen, D.I., Acharya, N., Adefisan, E.A., Anderson, M.L., Aura, S., Balogun, A.A., Bertram, D., Bluhm, S., Brayshaw,

D.J. and Browell, J. (2021) Advances in the application and utility of subseasonal-to-seasonal predictions. *Bulletin of the American Meteorological Society*, 103, E1448–E1472.

Wulff, C.O. and Domeisen, D.I. (2019) Higher subseasonal predictability of extreme hot European summer temperatures as compared with average summers. *Geophysical Research Letters*, 46, 11520–11529.

Zschenderlein, P., Fink, A.H., Pfahl, S. and Wernli, H. (2019) Processes determining heat waves across different European climates. *Quarterly Journal of the Royal Meteorological Society*, 145, 2973–2989.

Zschenderlein, P., Pfahl, S., Wernli, H. and Fink, A.H. (2020) A Lagrangian analysis of upper-tropospheric anticyclones associated with heat waves in Europe. *Weather and Climate Dynamics*, 1, 191–206.

SUPPORTING INFORMATION

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

How to cite this article: Pyrina, M. & Domeisen, D.I.V. (2023) Subseasonal predictability of onset, duration, and intensity of European heat extremes. *Quarterly Journal of the Royal Meteorological Society*, 149(750), 84–101. Available from: https://doi.org/10.1002/qj.4394