Seasonal prediction of European summer heatwaves

Chloé Prodhomme[1,2,3] · Stefano Materia[4] · Constantin Ardilouze[1] · Rachel H. White[3,5] · Lauriane Batté[1] · Virginie Guemas[1] · Georgios Fragkoulidis[6] · Javier García-Serrano[2]

Abstract

Under the influence of global warming, heatwaves are becoming a major threat in many parts of the world, affecting human health and mortality, food security, forest fires, biodiversity, energy consumption, as well as the production and transportation networks. Seasonal forecasting is a promising tool to help mitigate these impacts on society. Previous studies have highlighted some predictive capacity of seasonal forecast systems for specific strong heatwaves such as those of 2003 and 2010. To our knowledge, this study is thus the first of its kind to systematically assess the prediction skill of heatwaves over Europe in a state-of-the-art seasonal forecast system. One major prerequisite to do so is to appropriately define heatwaves. Existing heatwave indices, built to measure heatwave duration and severity, are often designed for specific impacts and thus have limited robustness for an analysis of heatwave variability. In this study, we investigate the seasonal prediction skill of European summer heatwaves in the ECMWF System 5 operational forecast system by means of several dedicated metrics, as well as its added-value compared to a simple statistical model based on the linear trend. We are able to show, for the first time, that seasonal forecasts initialized in early May can provide potentially useful information of *summer heatwave propensity*, which is the tendency of a season to be predisposed to the occurrence of heatwaves.

Keywords Seasonal forecast · Heatwave

✉ Chloé Prodhomme
chloe.prodhomme@meteo.fr

1 CNRM, Université de Toulouse, Météo France, CNRS, UMR3589, Toulouse, France

2 Group of Meteorology, Universitat de Barcelona (UB), Barcelona, Spain

3 Barcelona Supercomputing Center (BSC), Barcelona, Spain

4 CSP Division, Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy

5 Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, Canada

6 Institute for Atmospheric Physics, Johannes Gutenberg University, Mainz, Germany

1 Introduction

Heatwave impacts are various, considerable and often underestimated by the general public (Perkins 2015). A heatwave can result in disruptions to transport and energy networks, crop loss (Barriopedro et al. 2011), air-quality degradation, forest fires (Karoly 2009), accelerated glacier melting (Hughes 2008), constitute a major threat for biodiversity (Welbergen et al. 2008), and cause increased human mortality (Ballester et al. 2011). As an illustration, the 2003 and 2010 heatwaves caused around 70,000 heat-related deaths in Europe (Robine et al. 2008) and around 55,000 deaths in Russia (Otto et al. 2012), respectively. During the summer of 2019, the months of June and July were the hottest on record for the European continent, leading to numerous fatalities and large disruptions in the European transport network (WMO 2019). Due to global warming, heatwave frequency and magnitude are expected to increase substantially in the coming decades (Russo et al. 2014; Schoetter et al. 2015; Vogel et al. 2020). Giving insights into the risk of occurrence of extreme temperature a few months in advance will be crucial for climate services dedicated to adaptation

and risk management (Thomson et al. 2006; García-Herrera et al. 2010; Ceglar et al. 2018).

The mechanisms leading to the generation of a heatwave are not yet fully understood (Perkins 2015). Heatwaves are almost always associated with a synoptic-scale quasi-stationary high-pressure system (i.e. anticyclone) that persists for several days (Duchez et al. 2016; Brunner et al. 2018; Sousa et al. 2018). The occurrence of such persistent anticyclones is strongly influenced by the chaotic variability of the atmosphere and it is therefore hardly predictable several weeks to months ahead. However, several studies have shown that European heatwaves are also associated with slowly-evolving fields, such as North Atlantic sea surface temperatures (SSTs) (Cassou et al. 2005; Duchez et al. 2016) and various low-frequency phenomena, such as the El Niño–Southern Oscillation (ENSO) (Zhu et al. 2015; Wulff et al. 2017), the North Atlantic Oscillation (NAO) (Kenyon and Hegerl 2008), and the Atlantic Multi-decadal Oscillation (AMO) (Della-Marta et al. 2007). These variability modes are predictable to some extent, but their teleconnections and contributions to heatwaves over Europe are not yet fully understood. The impact of such modes of variability on the sources and propagation of Rossby waves likely plays an important role in this regard (e.g. Wolf et al. 2020). In addition, the land-atmosphere interaction also contributes to the evolution of heatwaves. Dry soils amplify heatwaves through a positive feedback: soil moisture depletion by warm and dry atmospheric conditions leads to a reduction of evaporative cooling which further increases air temperatures (Miralles et al. 2014; Ardilouze et al. 2019). Quesada et al. (2012) showed that late spring wet soils generally inhibit the development of summer heatwaves, while the impact of dry soils is more uncertain. Nonetheless, Fischer et al. (2007) and Whan et al. (2015) found that late-spring dry conditions could enhance heatwave development. Vegetation phenology has also been found to play a role on heatwaves magnitude (Lorenz et al. 2013). Finally, it has been shown that a better representation of soil–atmosphere interactions improves the seasonal prediction skill of summer temperatures (Bunzel et al. 2018).

Heatwaves are meteorological events; therefore, they cannot be predicted with exact timing and location beyond approximately 2 weeks in advance. However, given that large-scale teleconnections and soil moisture play a role in heatwave development, as well as the warming trend associated with climate change (Russo et al. 2014), the occurrence of European heatwaves is potentially partially predictable a few weeks-to-months ahead. Pepler et al. (2015) and Bhend et al. (2017) found that the seasonal predictive skill for extreme temperature indices was closely related to the skill of the seasonal mean temperature. The soil moisture initialization increases the seasonal prediction skill for extreme temperature indices (Prodhomme et al. 2016b;

Ardilouze et al. 2017). Recently, Wulff and Domeisen (2019) showed that heatwaves were more predictable at the sub-seasonal timescale than near-average and cold conditions. In addition, a few studies have investigated the retrospective predictability at seasonal time-scales of specific heatwaves. The 2003 European heatwave has been shown to be predictable at the seasonal time-scale, although the sources of predictability remain unclear (Weisheimer et al. (2011), Prodhomme et al. (2016b), among others). For the 2010 Russian heatwave, Dole et al. (2011) and Katsafados et al. (2014) have shown limited prediction skill, although, it was suggested by Prodhomme et al. (2016b), the 2010 heatwaves was predictable with a proper soil initialization. A comprehensive assessment of European heatwave predictive skill by operational seasonal forecasts has not yet been carried out.

In recent years, several climate indices have been used to measure heatwave duration and severity [see Perkins (2015) for a review]. These indices are often designed to target specific impacts (e.g., forest fires, mortality, or crop loss) and thus have limited usefulness for a comprehensive analysis of heatwave variability (Otto et al. 2012). Heatwave indices are often based on fixed thresholds, such as, for example, a relative threshold of 5 °C above climatology (Frich et al. 2002) or an absolute threshold of 30 °C for Tmax (maximum temperature over a day) (Ceglar et al. 2016). Fixed thresholds are not suitable for comparing regions with different climatological mean temperature and/or variance, nor for measuring the spatial extent of a heatwave encompassing climatologically heterogeneous regions, as discussed for Africa by Barbier et al. (2018) and Batté et al. (2018). In addition, they are not easily applied to model data, as models usually have biases in both climatology and variability. Indices based on percentiles have also been used (Fischer and Schär 2010; Pepler et al. 2015; Bhend et al. 2017). Most of the existing indices do not allow simultaneous measurement of heatwave duration and amplitude while accounting for the local climate. To overcome these limitations, Russo et al. (2014) designed the Heat Wave Magnitude Index (HWMI). This index computes for one season the most severe heatwave (at least 3 days above the 90th percentile), with respect to local climate (see Sect. 2 for more details). We extend this index to produce a metric of seasonal heatwave propensity.

In this paper, we present a novel methodology for heatwave detection at the seasonal time scale and, based on several complementary heatwave indices, we systematically assess the prediction skill of the ECMWF operational seasonal forecast system 5 (SEAS5) when it comes to European heatwaves. Section 2 presents the data used and our methodology. Section 3 describes the skill achieved for heatwave prediction by SEAS5 and Sect. 4 offers a summary and outlook of the work.

2 Data and methods

2.1 Seasonal prediction system and observational data

This study employs ensemble forecast and hindcast data from SEAS5 initialized on May 1st. The hindcasts span the period from 1981 to 2016 and consist of 25 ensemble members. This system became operational in 2017 (Johnson et al. 2019). The hindcast and the forecast use exactly the same model version, however some differences occur in the initialisation, therefore we decided to exclude the operational forecast for score calculation at grid point level and for regional averages (Correlation and Brier Skill Score). However, to obtain a more robust estimation of skill for the heatwave detection method, with more events, we also include the operational forecasts of years 2017, 2018 and 2019, using 25 out of the 51 available members to be consistent with the hindcasts. The sensitivity of this choice is discussed in Sect. 3. All the analyses that follow are applied to the full summer season, from May 15th to August 31st (hereafter denoted, 15MJJA). We remove the two first weeks in order to exclude the predictability of the deterministic range and thus focus on the added value of the seasonal prediction system thereafter. We retain the second half of May, since heatwaves also occur during that period. We do not expect the model to be able to predict the exact timing of heatwaves, therefore removing the entire first month would arbitrarily limit the predictability of the model. The impact of this choice is discussed in Sect. 3.

To verify the performance of the SEAS5 forecasts, we use the ERA5 reanalysis (Hersbach et al. 2020) and a purely observational gridded temperature dataset, which is derived from a dense array of European weather stations, namely the E-OBS dataset (Cornes et al. 2018) (see supplementary material).

2.2 Heatwave definition

We first define heatwaves based on three indicators at a local, i.e. grid-point level. We then extend this local heatwave identification across contiguous grid points to produce a heatwave index that incorporates information of both event magnitude and spatial extent.

2.2.1 Local warm extremes

We use three indicators, described below, to define the seasonal heatwave indices at a grid point level. The steps for the construction of these indices is illustrated on Fig. 1.

– N90:

1. Estimation of the 90th percentile: As illustrated on Fig. 1a, we estimate the climatological daily 90th percentile of the maximum daily 2m air temperature (Tmax) at each grid point. The resulting 90th percentile as a function of the calendar day from May 15th to August is smoothed by applying a polynomial regression following Mahlstein et al. (2015), rather than the running mean applied by Russo et al. (2014).

2. **N90**: $N90$ is then defined as the number of days exceeding the smoothed 90th percentile within the considered season (15MJJA).

$$N90(i,j,year) = \sum_{15/05/year < t < 31/08/year} \mathbb{1}_{(Tmax(i,j,t) > P_{90}(i,j,t))}$$

where $Tmax(i,j,t)$ is the maximum 2m temperature for a given day t, longitude i and latitude j. $P_{90}(i,j,t)$ is the smoothed climatological 90th percentile for this date, latitude and longitude.

– HWMI

3. Sub-heatwave definition: A heatwave is defined when at least 3 consecutive days exceed the smoothed 90th percentile. As in Russo et al. (2014), heatwaves lasting longer than 3 days are separated into consecutive sub-heatwaves; these sub-heatwaves are all 3 days in length, with the exception of the last sub-heatwave which may be shorter (e.g. a heatwave lasting 7 days would be separated into 3 sub-heatwaves ((days 1–3 and 4–6, and one of 1 day (day 7). For each sub-heatwave the corresponding magnitude M is given by:

$$M(subHW(i,j,T))$$
$$=$$
$$\sum_{t=T}^{T+3} \mathbb{1}_{(Tmax(i,j,t) > P_{90}(i,j,t))} \cdot (Tmax(i,j,t) - P_{90}(i,j,t))$$

where T is the first day of the sub-heatwave. Each sub-heatwave is shown with a different color in Fig. 1b.

4. Scaled Sub-heatwave: For each year and each grid-point, we find the maximum M in the whole season (red star in Fig. 1b), hereafter $MaxM$, defined as:

$$MaxM(i,j,year) = \max_{\{T \in year\}} M(subHW(i,j,T))$$

With the maximum $MaxM$ for each year, An empirical cumulative density function (ECDF) is fitted to the annual maximum of the sub-heatwave unscaled magnitudes, as illustrated in Fig. 1c. The reason why
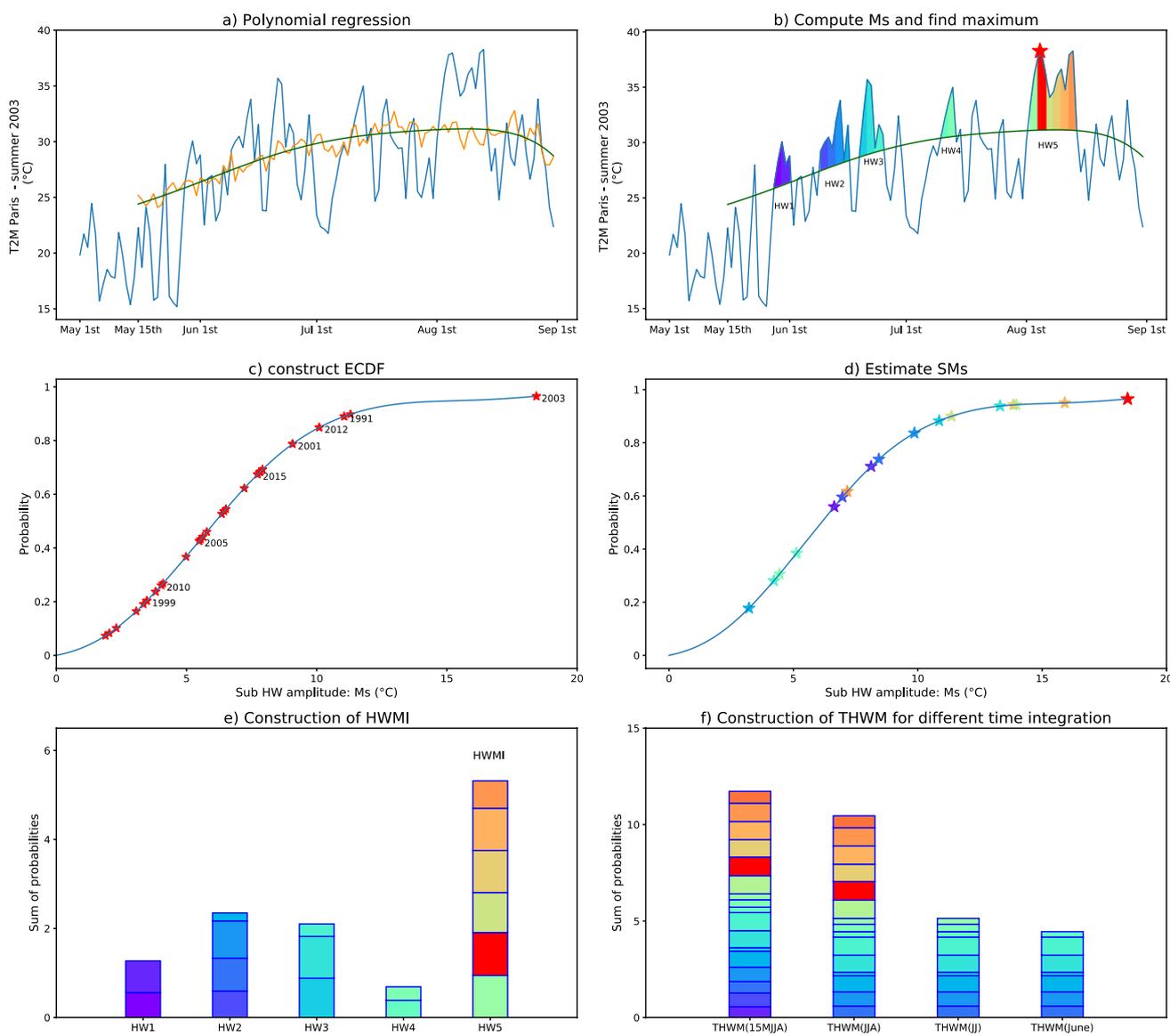
Fig. 1 Construction of the HWMI and THWM indices for a grid point close to Paris during summer 2003 based on ERA5. For simplicity, the process is illustrated without the cross validation. **a** The blue curve shows the daily time series of Tmax, the orange curve shows the climatological 90th percentile computed over the 1981–2016 period, and the green line shows the smoothed climatological 90th percentile, estimated with a 6th order polynomial regression. **b** The blue and green lines are the same as in **a**. Each sub-heatwave is indicated with a different colour and colored areas correspond to the sub-heatwave magnitude Ms estimate. The red star and area mark the maximum sub-heatwave of the season. **c** Construction of the ECDF using the maximum sub-heatwave of each year. **d** Estimation of SMs using the ECDF of **c**, with colors corresponding to the sub-heatwaves of **b**. **e** Estimation of HWMI: for each heatwave (HW) marked on **b**, we sum the SMs of all sub-heatwaves belonging to it. The maximum value is the HWMI (corresponding to HW5 for this example). **f** Estimation of the THWM for different periods: 15MJJA, JJA, JJ and June. For each period, the SMs of all the sub-heatwaves in the period are summed

only the annual maximum MS's are used is because the index means to exclude heatwaves that are less notable: heatwaves are in fact extreme events, therefore they are not expected to occur frequently. In a more recent article that makes use of the HWMI index, Russo et al. (2019) point out that, in order to

study heatwaves on a statistical point of view, the decadal maximum has to be taken. For each M, we find its scaled magnitude (between 0 and 1) amongst all sub-heatwaves using the ECDF, as shown in Fig. 1d:

$$ECDF(i,j,M) = \frac{\sum_{year=1}^{N_{year}} \mathbb{1}_{MaxM(i,j,year) \leq M}}{N_{year}}$$

$$SM(subHW(i,j,T)) = ECDF(M(subHW(i,j,T)))$$

where N_{year} is the total number of years. This step scales all sub-heatwaves relative to the local climate variability. It offers an alternative to classical bias correction methods by defining derived quantities which are fully comparable between model and observations in terms of mean and variances. If $SM(subHW) = 1$, this sub-heatwave corresponds to the warmest 3-day period (sub-heatwave) of the considered season (15MJJA) over all years for the grid point in question.

5. Heatwave magnitude: The magnitude of a heatwave is defined as the sum of the SM of the consecutive sub-heatwaves that the heatwave (HW) is comprised of:

$$M_{HW}(HW) = \sum_{\{subHW \in HW\}} SM(subHW)$$

6. HWMI: A given year can have multiple heatwaves, and thus multiple values of $M_{HW}(HW)$. The $HWMI$ is defined as the maximum M_{HW} for a given season (i.e. we obtain one value for each year, each ensemble member and each grid point in our dataset). It represents the strongest heatwave for a given season based on both duration and intensity. The construction of M_{HW} and HWMI is illustrated on Fig. 1e.

$$HWMI(i,j,year) = \max_{HW \in year} (M_{HW}(i,j,HW))$$

– THWM: We define *THWM* for a grid-point as the sum of SMs values over a given season for one grid-point (Fig. 1f). This gives a seasonal heatwave index, incorporating the number of heatwaves, their duration and their magnitude.

$$THWM(i,j,year) = \sum_{HW \in year} M_{HW}(i,j,HW)$$

For the hindcast period (1981–2016), steps 1 and 4 above are done in cross-validation for both SEAS5 and re-analysis/observational datasets, meaning that the climatology and the ECDF are estimated excluding the year we are considering. For the forecast period (2017–2019) the climatology and ECDF are estimated using the hindcast period as it would be done in operational prediction.

2.3 Studied regions

Spatial aggregations of the HWMI and THWM indices are also employed in order to investigate the predictive skill for different regions. Table 1 lists the selected regions, which are also illustrated in Fig. S1:

2.3.1 Heatwave detection

In order to identify severe heatwaves, i.e. long-lasting and large-impact heatwaves that cover a large spatial area, we use a heatwave detection algorithm on the scaled sub-heatwave field of the entire European domain, *SM*. This algorithm is based on a seeding algorithm similar to Barbier et al. (2018). A heatwave event is defined as encompassing all the sub-heatwave points separated by less than 2° spatially and 1 day temporally. All sub-heatwave points are classified as part of one heatwave. Based on this detection algorithm, we compute for each heatwave the following quantities:

– total Magnitude: area-weighted, using the cosinus of the latitude, and temporal sum of all *SMs* over land belonging to the heatwave.

$$TM(HW) = \frac{\sum_{(i,j,t) \in HW} SM(subHW(i,j,t)) \cos(lat(j))}{\sum_{(i,j) \in HW} \cos(lat(j))}$$

– Magnitude (1D-timeseries): area-weighted sum of all scaled sub-heatwaves over land belonging to the heatwave.

$$HWM(t) = \frac{\sum_{(i,j) \in HW} SM(subHW(i,j,t)) \cos(lat(j))}{\sum_{(i,j) \in HW} \cos(lat(j))}$$

– Heatwave pattern (2D field): at each grid point, the sum over time of all scaled sub-heatwaves belonging to the heatwave:

$$HWF(i,j) = \sum_{T \in HW} SM(subHW(i,j,T))$$

– Start/end date: date of the first/last point belonging to the heatwave.

2.3.2 Severe heatwave

The algorithm described in the previous section groups sub-heatwaves into a single heatwave; any number of

Table 1 Definitions of the different European regions

Central Europe	5° E 25° E–44° N 55° N
Mediterranean	10° W 25° E–36° N 45° N
Western Europe	10° W 5° E–40° N 60° W
Northern Europe	10° W 30° E–55° N 70° N
Eastern Europe	30° E 60° E–40° N 65° N
Europe	15° W 60° E–35° N 70° N

sub-heatwaves can be grouped. The resulting heatwaves can therefore cover a large spectrum of duration and spatial extent. We do not expect a seasonal forecast system to be able to predict short and localized events, but we hope that it could give information about long-lasting and large-extent ones. Therefore, we define a *severe heatwave* as a heatwave stronger to either of the *N* strongest heatwaves that occurred in the whole re-forecast period, excluding the year to be predicted (in terms of total magnitude, as defined in Sect. 2.3.1). For the operational forecasts (2017–2019) the entire hindcast period is used to estimate the threshold. This threshold is defined separately for model and observations to account for the different characteristics of simulated and observed heatwaves. For a given severity threshold *N*, due to cross-validation, it could happen than more than *N* severe heatwaves are detected. Afterwards, we refer to *N* as the *Heatwave Severity Threshold, HST*.

2.4 Verification

We assess the model performance by comparing the indices described above in our hindcasts/forecasts with both re-analysis data and temperature observations (E-OBS). To this end, we use various widely-used forecast skill scores. At the grid point level, and for regional averages, we use the temporal Pearson correlation coefficient (Corr). We computed the Brier Skill Score (BSS; Brier 1950) using two different references: the climatology and a reference accounting for the warming trend (see Appendix for more details). The correlation coefficient indicates how close the ensemble mean is to the truth while the BSS indicates to what extent the distribution of the members is representative of the likelihood of occurrence of a heatwave.

The detection of severe heatwaves at European scale is verified with several metrics based on the contingency table for event detection verification: accuracy, hit rate, threat score (TS) and equitable threat score (ETS) (Jolliffe and Stephenson 2012). All the scores are described in detail in Appendix 1.

2.4.1 Heatwave propensity

We do not expect seasonal forecasts to be able to accurately predict the onset, duration and magnitude of a particular heatwave event, in other words with the same timing and location that the observed event (this is verified in the Sect. 3 on Fig. 9). Seasonal forecasts might be able to indicate if a season is predisposed to the occurrence of heatwaves. This is referred afterwards as *heatwave propensity*. For a given grid point, heatwave propensity could be estimated using the three local indices defined above: N90, HWMI and THWM. The heatwave propensity of a given region could be obtained by spatially averaging these indices over

this region. Another method to obtain the heatwave propensity for the whole European domain, is to apply the detection algorithm described above and estimate the number of severe heatwaves produced by the dynamical forecast system. The skill of heatwave propensity in SEAS5 for these different metrics is described in the Sect. 3.

2.5 Statistical prediction based on linear trend

A large part of temperature predictability over Europe is associated with the global warming trend (Doblas-Reyes et al. 2006). To assess whether the model heatwave prediction skill is related solely to the warming trend, we use a simple statistical prediction model, assuming a linear trend, constructed in cross-validation:

$$Pred_{stat}(var, year) = \alpha_{year}(var).year + \beta_{year}(var),$$

where *var* is the variable we want to predict (i.e., N90, HWMI or THWM), *year* the year we want to predict, $\alpha_{year}(var)$ and $\beta_{year}(var)$ are the regression coefficients estimated for the variable *var* from all the years except the year *year* (cross-validation), using the least squares method.

For the detection of severe heatwaves, this statistical prediction necessarily produces a more severe heatwave every year, therefore it simply locates the $HST = N$ strongest heatwaves during the *N* last years of the period of interest.

This model is particularly simple, but it constitutes a proper benchmark prediction for our purposes. Any predictive skill beyond the one of this benchmark can be attributed to the ability of SEAS5 to predict the interannual variability in heatwave propensity and characteristics over Europe (i.e., the crucial "noise" around the decadal warming trend). It is beyond the scope of this paper to construct a more sophisticated statistical prediction model for this purpose.

3 Results

We first verify the ability of SEAS5 to forecast/hindcast observed heatwave metrics from 1981–2016 at a grid-point level. We then consider larger regions (given in Table 1) through a simple spatial averaging of the grid-point metrics. Finally, we use our heatwave detection algorithm to identify observed severe heatwaves, taking into account amplitude, duration and spatial scale, and assess the model ability to predict the amplitude, location and timing of these events.

3.1 Analysis of local warm extremes

In order to provide a first assessment of seasonal heatwave prediction skill, Fig. 2a–d shows the correlation index (*Corr*) of the SEAS5 ensemble mean for mean temperature, number
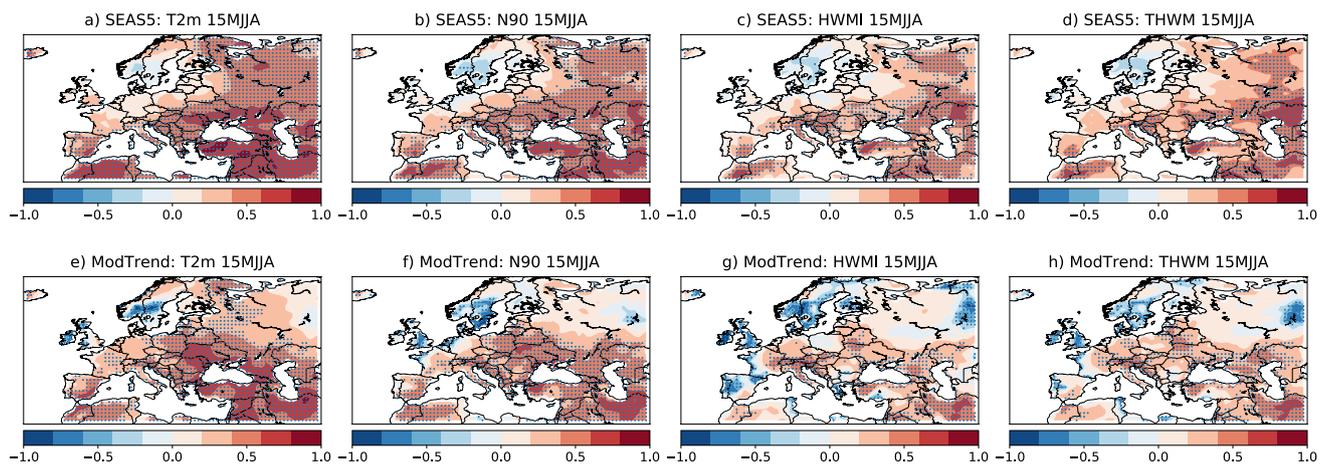
Fig. 2 Grid point correlation (*Corr*) between SEAS5 and ERA5 for the **a** average 2m-temperature (15MJJA), **b** number of days above daily N90 in 15MJJA, **c** HWMI computed for 15MJJA and **d** THWM computed for 15MJJA. **e–h** Same as **a–d** but for the statistical prediction model based on linear trend (ModTrend). For all panels stippling indicates correlation significant at the 95% confidence level

of days exceeding the 90th percentile (N90), HWMI and THWM verified against ERA5. For all metrics, SEAS5 exhibits positive correlation over most of the European domain that is statistically significant at the 95% confidence level especially in southern and eastern Europe. The skill for the heatwave indices (three rightmost columns) is slightly lower and less significant than that for mean 2m-temperature, particularly over western Europe. This is consistent with previous studies analysing the skill of extreme indices at seasonal time scale (Pepler et al. 2015; Bhend et al. 2017). The correlation achieved for N90 is larger than for HWMI and THWM, thus implying that SEAS5 is able to predict the total number of hot days, but it is less skillful in predicting the heatwave intensity. The level of skill measured with the E-OBS dataset is very similar (see supplementary material Fig. S2), therefore in the rest of the paper the verification is performed using only ERA5.

To ascertain how much of the prediction skill of SEAS5 can be attributed to simply reproducing the observed warming trend (i.e. more recent years are hotter, with more heatwave activity), we compare the skill, as determined by correlation, of the SEAS5 model with that from a simple prediction model based on a linear trend (ModTrend, see Sect. 2). Figure 2e–h, shows the correlation obtained with ModTrend for the mean 2m-temperature and warm extreme indices. For all indices and generally over the entire region, SEAS5 exhibits higher correlations than ModTrend, except over Central Europe. The prediction skill of SEAS5 is indeed to some extent related to the warming trend, but there is useful additional information from the dynamical model. Such a gain in predictability by the dynamical model appears particularly pronounced over Russia.

We further analyse the skill of SEAS5 in predicting summer mean temperature and heatwave metrics using the Brier

Skill Score (BSS) computed with the climatology as a reference. Figure 3a–c shows BSS of T2M compared to ERA5 for three categories: lower quintile (cold summers, inferior to 20%), inter-quintile range (normal summers, between 20 and 80%) and upper quintile (warm summers, superior to 80%). SEAS5 shows little probabilistic prediction skill above the statistical prediction (*BSS* > 0). The SEAS5 shows little prediction skill (*BSS* > 0) for normal summer (Fig. 3b) and higher for extreme summer, especially for warm extremes (right column). Figure 3d–l shows the BSS for the same three categories for heatwave indices (N90, HWMI and THWM). For each of the three indices, the SEAS5 shows low skill for weak and moderate heatwaves (left column and central column), and higher skill for strong heatwaves, especially over Eastern Europe and Mediterranean region (right column). Figure S3, shows the same figure but for BSS computed with a reference accounting for the warming trend, this figures shows similar results, with a higher predictability of warm extremes, beyond the warming trend, especially over Eastern Europe.

3.2 Regional heatwave analysis

In the above section, we show that SEAS5 has some skill in predicting local heatwave metrics over much of Europe; however, heatwaves are typically widespread events, affecting many neighbouring grid points simultaneously. One simple way to consider heatwave metrics over a larger area is to spatially average them. We calculate seasonal-mean area-averages of HWMI and THWM for the different regions shown in Table 1 and Fig. S1. Figures 4 and 5 show these time series for HWMI and THWM, respectively; the black line shows ERA-5 values, while red dots and blue stars show the SEAS5 ensemble members and ensemble mean,
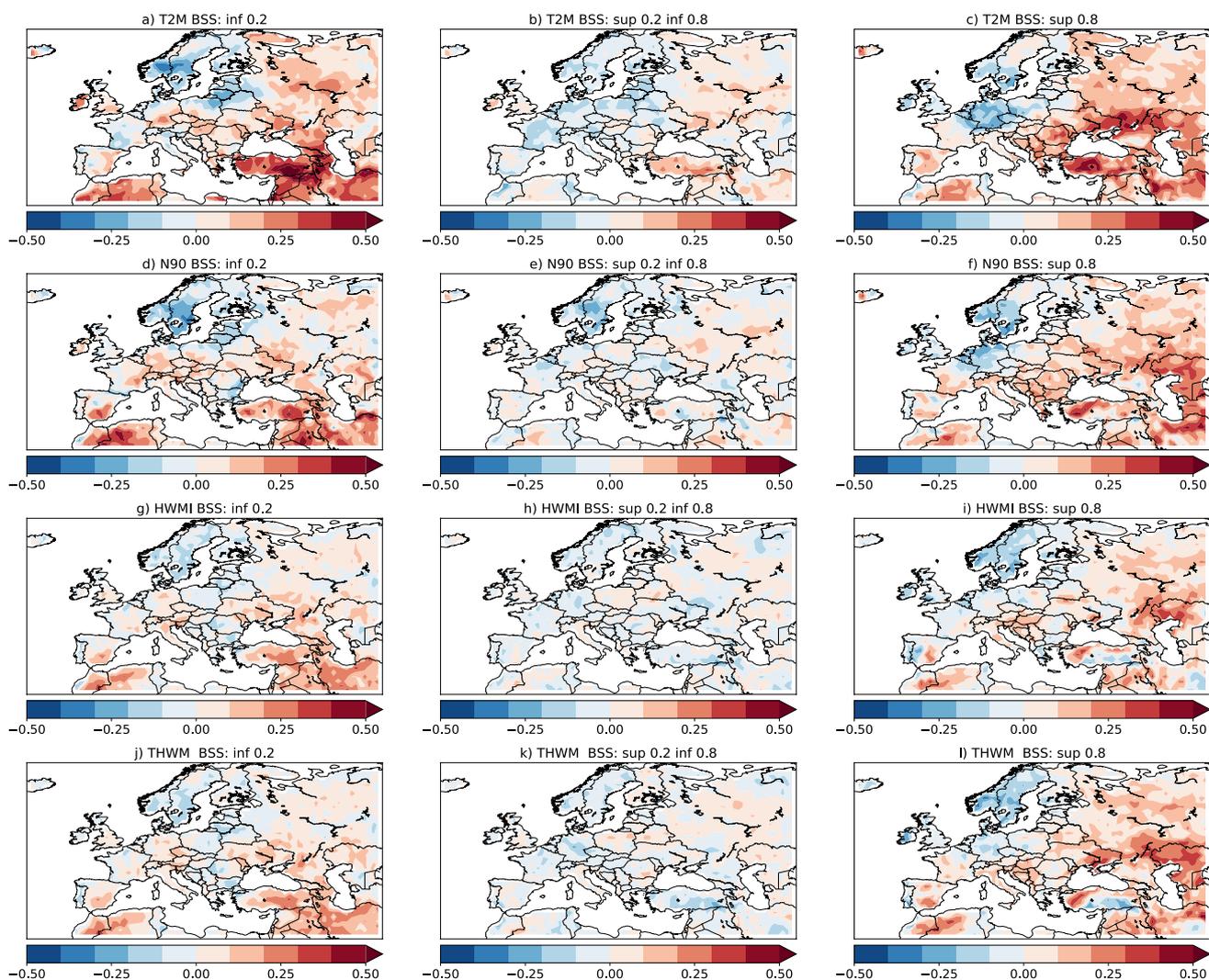
Fig. 3 Brier Skill Score for SEAS5 heatwave prediction using **a** average 2m-temperature (15MJJA) below the first quintile, **b** average 2m-temperature (15MJJA) in the inter-quintile range, **c** average 2m-temperature (15MJJA) above the third quintile. **d**–**f** Same as **a**–**c** but for N90. **g**–**i** Same as **a**–**c** but for HWMI. **j**–**l** Same as **a**–**c** but for THWMI

respectively. The corresponding correlation coefficients are shown in the upper left of each panel. SEAS5 achieves significant correlation, and therefore potentially useful skill, for all considered regions, except Northern and Western Europe, for both HWMI and THWM. A positive trend associated with warming is apparent in all regions and it is also well-reproduced by the SEAS5 ensemble mean. As a confirmation of what is shown in the correlation maps (Fig. 2), the skill is slightly higher for THWM than HWMI.

The black dashed line in Figs. 4 and 5 shows the results obtained from the Modtrend simple model. For both the HWMI (Fig. 4) and THWM (Fig. 5) indices, the dynamical model prediction (SEAS5) tends to outperform the simple linear model, except over Central Europe, consistent with Fig. 2. The increase in correlation coefficient for SEAS5 relative to that from the linear model is significant at the

95% confidence level (see Sect. 1 for more details) for both HWMI and THWM indices for Eastern Europe.

SEAS5 performs slightly better than a simple statistical forecast in capturing characteristics of heatwaves, as described by the HWMI and THWM indices. To better understand up to which time horizon heatwaves could be predicted, we compute the THWM correlation with ERA5 for different time periods within the summer season (Fig. 6). All skill scores are for forecasts initialized at the beginning of May, e.g. for the time period starting on July 1st we ignore the first 2 months of forecast. As examples, the circle denotes the correlation for THWM calculated from June 1st for 30 days, i.e. June 1st-30th; the black square is for THWM calculated from June 1st for 90 days, i.e. June 1st–Aug 29th. Correlation values, not significant at the 95% confidence level and correlations inferior to
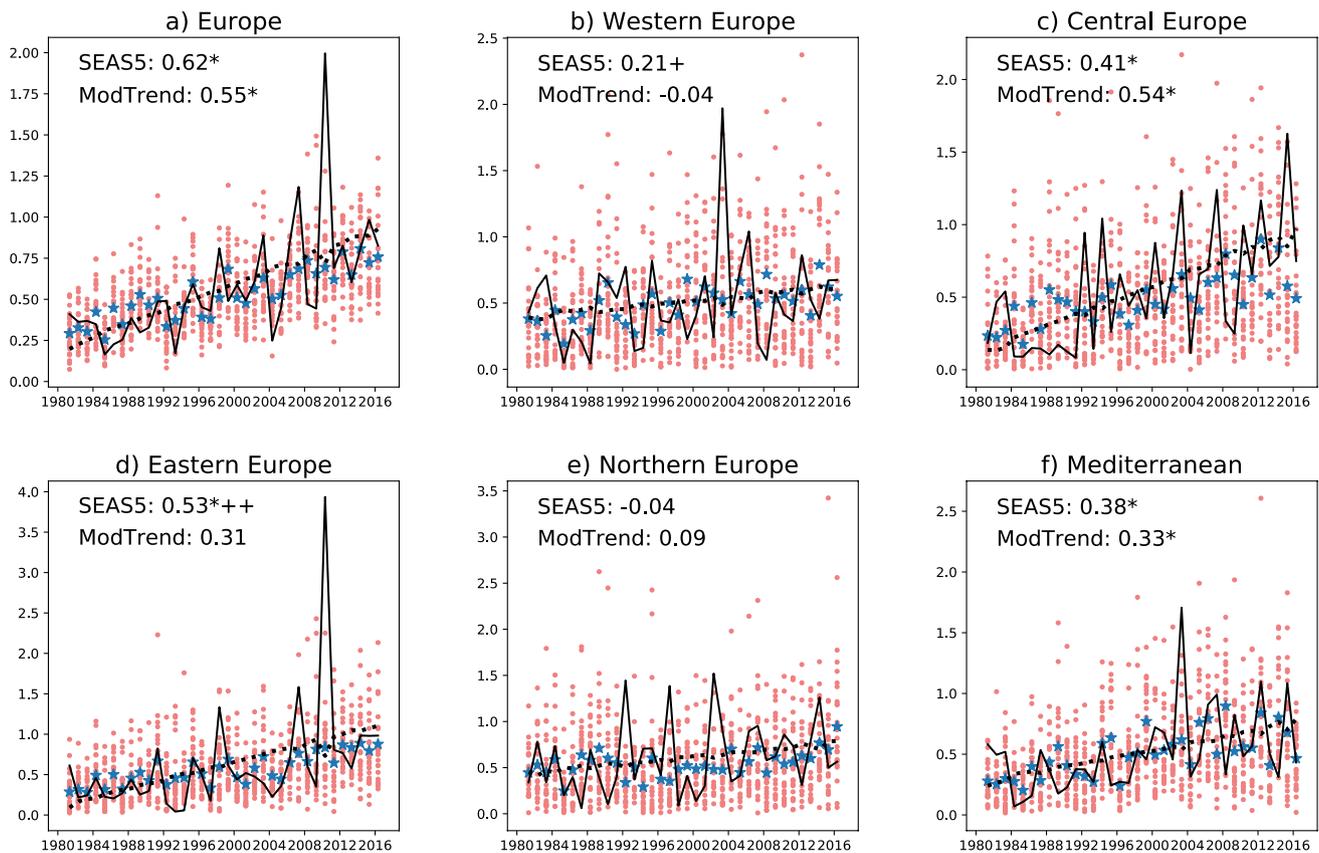
Fig. 4 a HWMI averaged over the whole Europe for 15MJJA. The black line shows ERA5, red dots show the different ensemble members and blue stars the ensemble mean. The black dashed line shows the results of the statistical prediction model constructed using the linear trend with cross-validation. The correlation between ensemble mean and observations is shown on top left of the figure for SEAS5 and for the statistical model. One star indicates correlation significant at the 95% confidence level and + and ++ show that SEAS5 has correlation higher than the statistical model at the 90% and 95% confidence interval, respectively. **b–f** Same as **a** but for the Western Europe, Central Europe, Eastern Europe, Northern Europe and the Mediterranean region, respectively

that of the simple statistical model are masked. Stippling indicates correlation significantly higher at the 90% confidence level than the ones achieved by the statistical model. In Northern, Western and Central Europe, there is almost no predictive skill for summer heatwaves (consistent with Figs. 4e, 5e). For Central Europe, SEAS5 does not perform better than the simple statistical model. Over the whole European domain, SEAS5 shows potentially useful skill for most of the considered time horizons, however it does not perform significantly better than ModTrend, except around June-July. SEAS5 exhibits significant correlation (above 0.4) and significantly higher than ModTrend, for two regions—Eastern Europe and the Mediterranean—for almost all forecast horizons considered, suggesting that in these two regions it could provide skilful heatwave forecasts up to three months in advance. The heatwave predictability beyond the warming trend in these two regions could be related to the strong soil–atmosphere coupling and the impact of soil pre-conditioning, as shown by Prodhomme et al. (2016a).

3.3 Severe heatwave analysis

As shown above, SEAS5 shows some predictive capability for summer heatwave indices over some parts of Europe, especially Eastern Europe and the Mediterranean region, which offers important prospects for public health and socio-economic activities. However, societal adaptation would be easier if forecasts could also provide information about the timing, duration, location, magnitude and spatial extent of a single severe heatwave. Heatwaves with large spatial extent and long duration have much bigger impact on society than short-lived local heatwaves. It remains unclear if seasonal forecasts can predict single severe heatwaves such as the 2003, 2010 and 2015 events (Prodhomme et al. 2016b; Feudale and Shukla 2011a, b; Fischer et al. 2007; Dole et al. 2011; Katsafados et al. 2014; Weisheimer et al. 2011; Mecking et al. 2019). In order to assess the performance of SEAS5 in this regard, we use the method described in Sect. 2.3.1, that considers as a single heatwave all the sub-heatwave points separated by less than 2° and 1 day. We
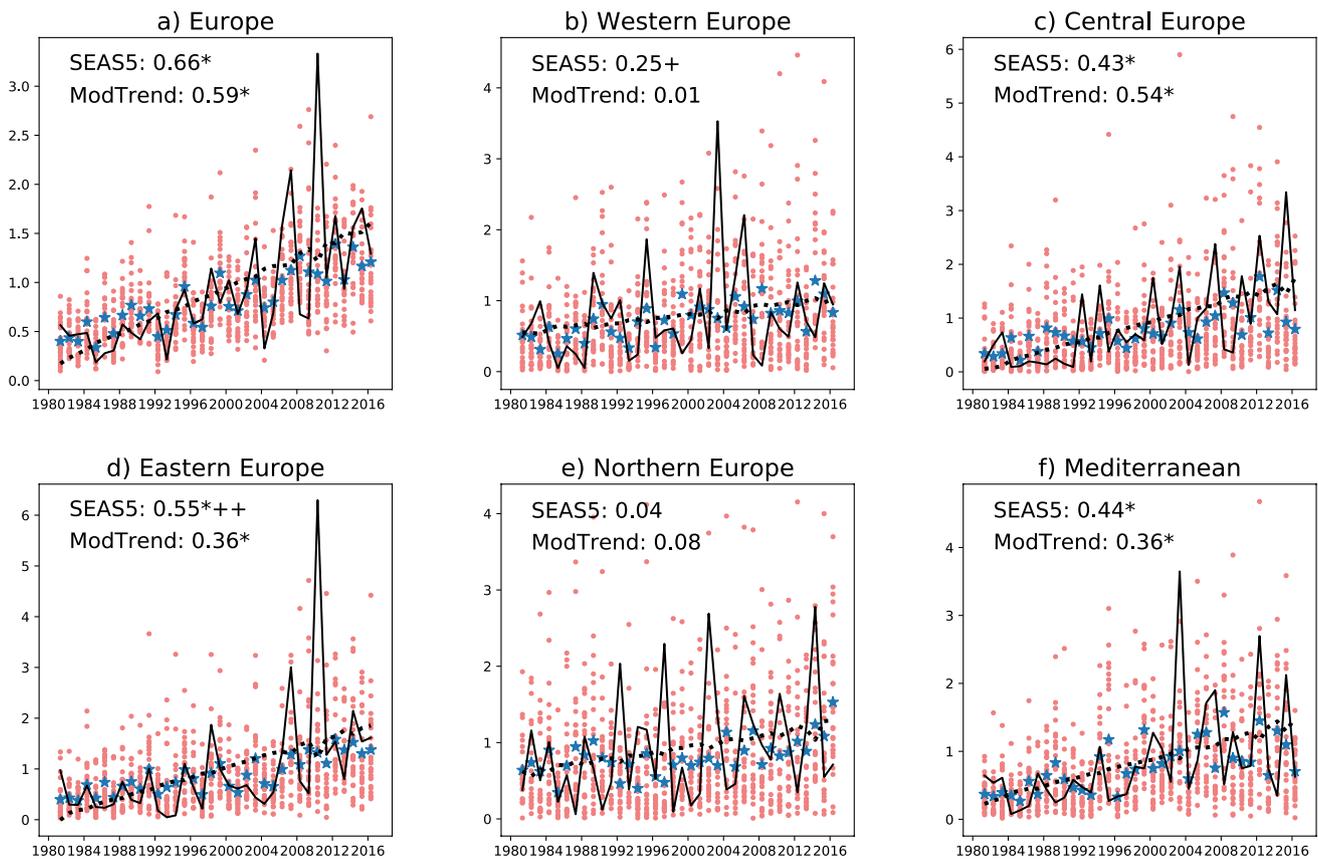
Fig. 5 Same as Fig. 4 but for THWM

consider that SEAS5 predicts a severe heatwave if at least 30% of its members (i.e 8 members) predict it (see Sect. 2 for more details).

Figure 7a shows the performance in 15MJJA prediction of severe heatwaves from May 1st, using the severity threshold $HST = 5$ and $HST = 10$. We compare the dynamical seasonal prediction system with ModTrend, which would predict the severe heatwaves in the last 5 (or 10) years of the re-forecast period. Results show that the dynamical model performs better than using either random predictions (ETS > 0) or simple statistical predictions. Consistent with Katsafados et al. (2014), who suggested that the 2010 heatwave was not predictable at a seasonal scale, it is not predicted by SEAS5 as a severe heatwave, when $HST = 5$. However, for $HST = 10$, the 2010 heatwave is successfully predicted. Therefore SEAS5 does predict that a heatwave occurs, but with a too small magnitude. It is interesting to note that the 2010 heatwave is not detected when we consider simple area-average metrics and ensemble means (Figs. 4, 5). This confirms that for predicting extreme events and extracting the maximum of information from seasonal forecast systems, special methodologies designed for extreme events are needed and simple spatio-temporal averages are not appropriate.

In order to better understand the sensitivity of this heatwave detection method to the choice of the severity threshold and the probability threshold, Fig. 8 shows the ETS score of SEAS5 and ModTrend as a function of HST. We also compare two different choices for the probability threshold, either using a fixed probability threshold, or using an optimal threshold maximizing the ETS in cross-validation (sea more details in Appendix 1). For both types of probability thresholds, SEAS5 performs generally better than ModTrend and a simple random forecast (ETS > 0) for a severity threshold $5 < HST < 20$. The skill, measured with the ETS, is strongly dependant on the parameter HST, and is maximum for $10 < HST < 15$. Choosing a $HST < 10$ is too demanding for the model, since it would require a precise prediction of the total heatwave magnitude. SEAS5 is able to predict a severe heatwave with some uncertainty on the magnitude of the heatwave, as illustrated previously for the 2010 heatwave. Using the optimal probability threshold gives higher skill than using a fixed threshold at 30%, for $HST < 5$ and $HST > 15$. Therefore, in order to optimize the forecast skill of heatwaves, the severity threshold should be between 10 and 15 and the probability threshold fixed at 30%. Fig. S4 shows the same figure computed only over the hindcast period (1981–2016) and shows similar results,
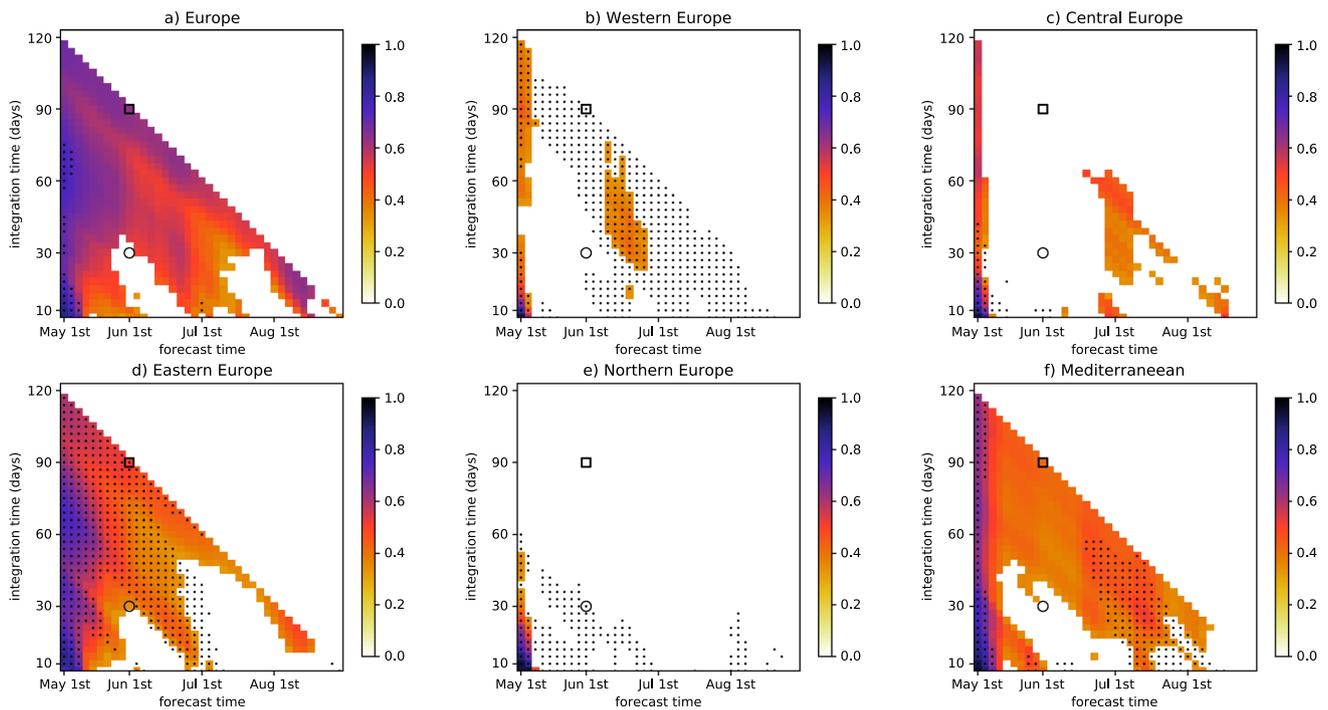
Fig. 6 Correlation values for THWM computed over different forecast windows between SEAS5 and ERA5 for different European regions (different panels). Only correlation values superior to the ones achieved by ModTrend and significant at the 95% confidence level are shown. Each point shows the correlation achieved for a different sub-season: from the date represented on the x-axis and over the integration period given on the y-axis. For example, the top left point corresponds to the correlation for MJJA, indicated in Fig. 5. The bottom-left point corresponds to the weather prediction from May 1st to May 10th. The black circle corresponds to the prediction of the month of June initialized on May 1st, while the black square corresponds to the prediction of JJA. Stippling indicates correlations significantly higher at the 90% confidence interval than the ones achieved by ModTrend

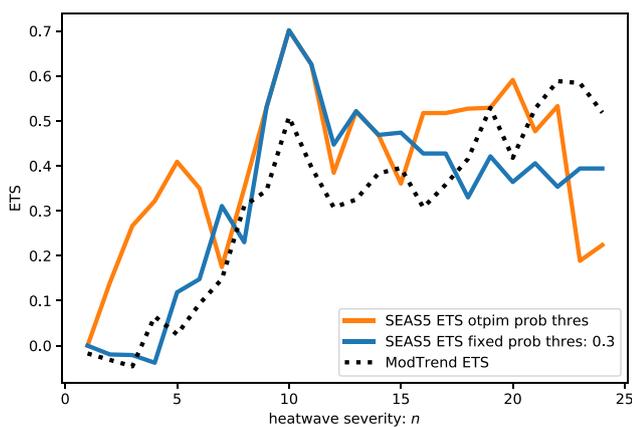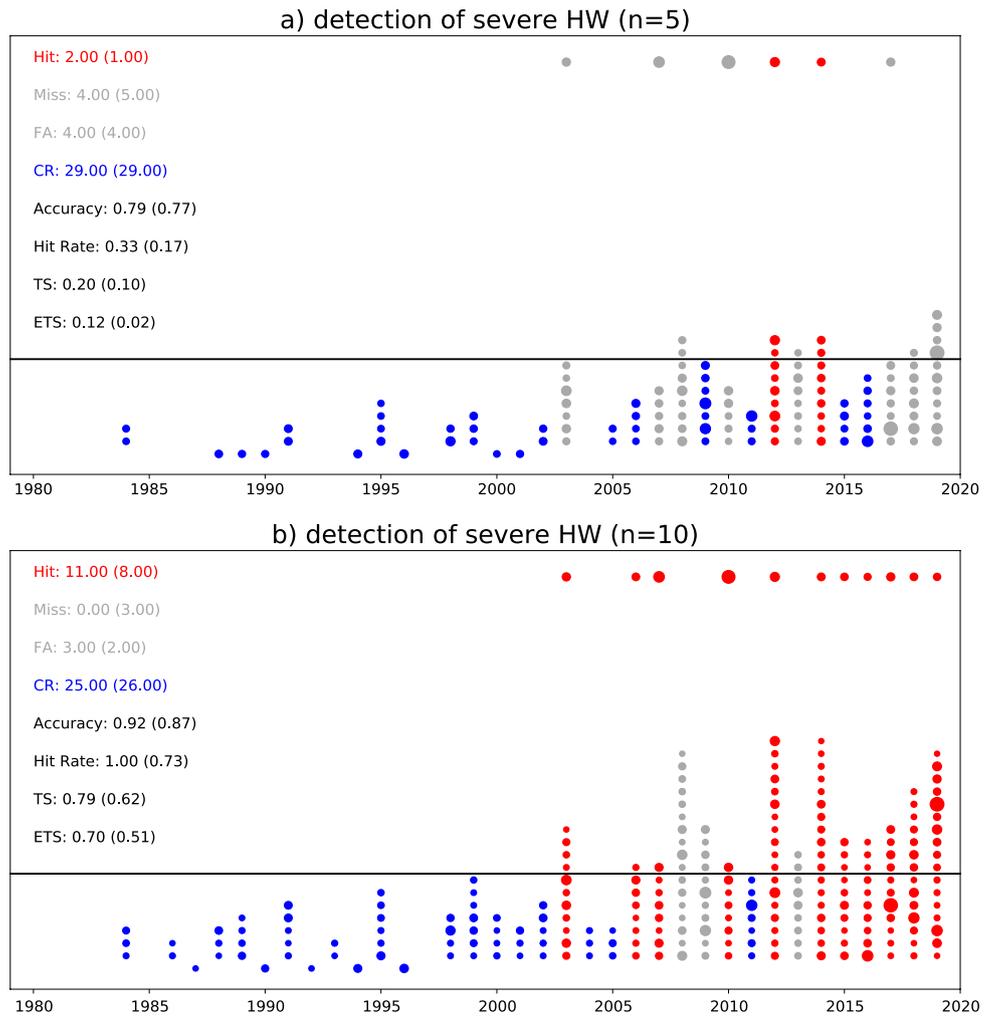which suggest that this result is not dependant to the chosen period.

To get a broader insight into individual heatwave characteristics (magnitude, spatial pattern, timing) captured by the forecast system, Fig. 9 compares the patterns of the severe observed heatwaves with their predicted counterparts, using the Heatwave detection method defined in Sect. 2. We choose to illustrate the results of this method for a severity threshold $N = 10$, for which a high ETS is achieved, in order to see what could be the information that SEAS5 could provide with an optimal choice of severe heatwave parameters. The first column shows the observed heatwave pattern, sum over time of all sub-heatwaves belonging to the heatwave (as defined in Sect. 2). The second column shows the ensemble-mean heatwave pattern, including only the members in which a heatwave is detected. This figure shows that, for some severe heatwaves, the location of the heatwave predicted by the model is very close to the observed one, such as 2007, 2010 and 2012, despite an underestimation of the magnitude. Individual members can forecast heatwaves with magnitudes that are comparable to observed heatwaves but with a different exact location. In the ensemble-averaging process, as expected, we obtain a heatwave pattern with

weaker magnitude and larger spatial extent than in the observations. For the rest of the years, the spatial pattern predicted by the model does not match the observed one. Therefore, the reliability of SEAS5 to predict severe heatwave location from May seems relatively low. The third column of Fig. 9 shows the time evolution of the observed heatwave (black line) as well as the maximal magnitude of the heatwaves predicted in the different members of SEAS5 and their timing (dots). The blue line shows the ensemble mean of heatwave magnitudes, for all the members predicting a heatwave. This figure illustrates that there is very little agreement among members about the timing of the heatwave peak, confirming that there is no deterministic predictability of the time evolution of the heatwave.

4 Discussion and outlook

In this study, we propose a comprehensive methodology to evaluate the heatwave propensity prediction skill using dynamical seasonal prediction systems and illustrate this methodology with the ECMWF operational seasonal forecast system SEAS5. We show that seasonal forecasts

Fig. 7 a Prediction of severe heatwaves with a threshold defined as the $HST = 5$ most severe heatwaves over the whole period, excluding the predicted year (e.g cross-validation). Top circles indicate observed severe heatwaves. Diameters of circles correspond to the observed magnitude. The bottom circles indicate the severe heatwaves that were predicted. Stacked circles indicate that different members predicted severe heatwaves. The size of the circles indicates the magnitude of the simulated heatwaves. The horizontal black line corresponds to the percentage of members needed to formally detect an event, in this case we use a threshold of 30% (i.e., 8 members). The color corresponds to the contingency table, grey: False Alarm or missed event, red: hit, blue: correct rejection. The numbers show the score obtained by SEAS5 and by the simple statistical model in parentheses (more information in Sect. 1). **b** Same as **a** but for the severity threshold $N = 10$

a) detection of severe HW (n=5)

Hit: 2.00 (1.00)
Miss: 4.00 (5.00)
FA: 4.00 (4.00)
CR: 29.00 (29.00)
Accuracy: 0.79 (0.77)
Hit Rate: 0.33 (0.17)
TS: 0.20 (0.10)
ETS: 0.12 (0.02)

b) detection of severe HW (n=10)

Hit: 11.00 (8.00)
Miss: 0.00 (3.00)
FA: 3.00 (2.00)
CR: 25.00 (26.00)
Accuracy: 0.92 (0.87)
Hit Rate: 1.00 (0.73)
TS: 0.79 (0.62)
ETS: 0.70 (0.51)

Fig. 8 ETS for severe heatwave prediction for different values of the parameter HST. The black dashed line shows the ETS achieved with ModTrend. Blue and orange lines show the ETS achieved with SEAS 5 using two different probability thresholds: a fixed threshold at 30% (blue) and an optimal threshold estimated by maximising the ETS in cross-validation (orange)

skilfully predict several local heatwave indices, especially over eastern and southern Europe. Consistent with previous studies, the prediction skill of extreme indices is bounded by the skill of mean temperature. Generally, the dynamical forecast system performs better than a simple statistical model based on the global warming trend. Dynamical forecasts also provide potentially useful information at the regional scale up two months ahead for THWM integrated over one, two or three months for the whole European domain, the Mediterranean region and Eastern Europe. Seasonal forecasts are able to provide reliable information about the propensity of occurrence of individual severe heatwaves, performing better than random predictions or a simple statistical model based on a linear trend. The prediction system does not seem able to give reliable information about the timing of the heatwave peak nor about the heatwave spatial pattern though.

Two main limitations of the present study are that it focused exclusively on one dynamical prediction system (SEAS5) and that it used a fairly simple statistical prediction model as a benchmark to the SEAS5 performance.
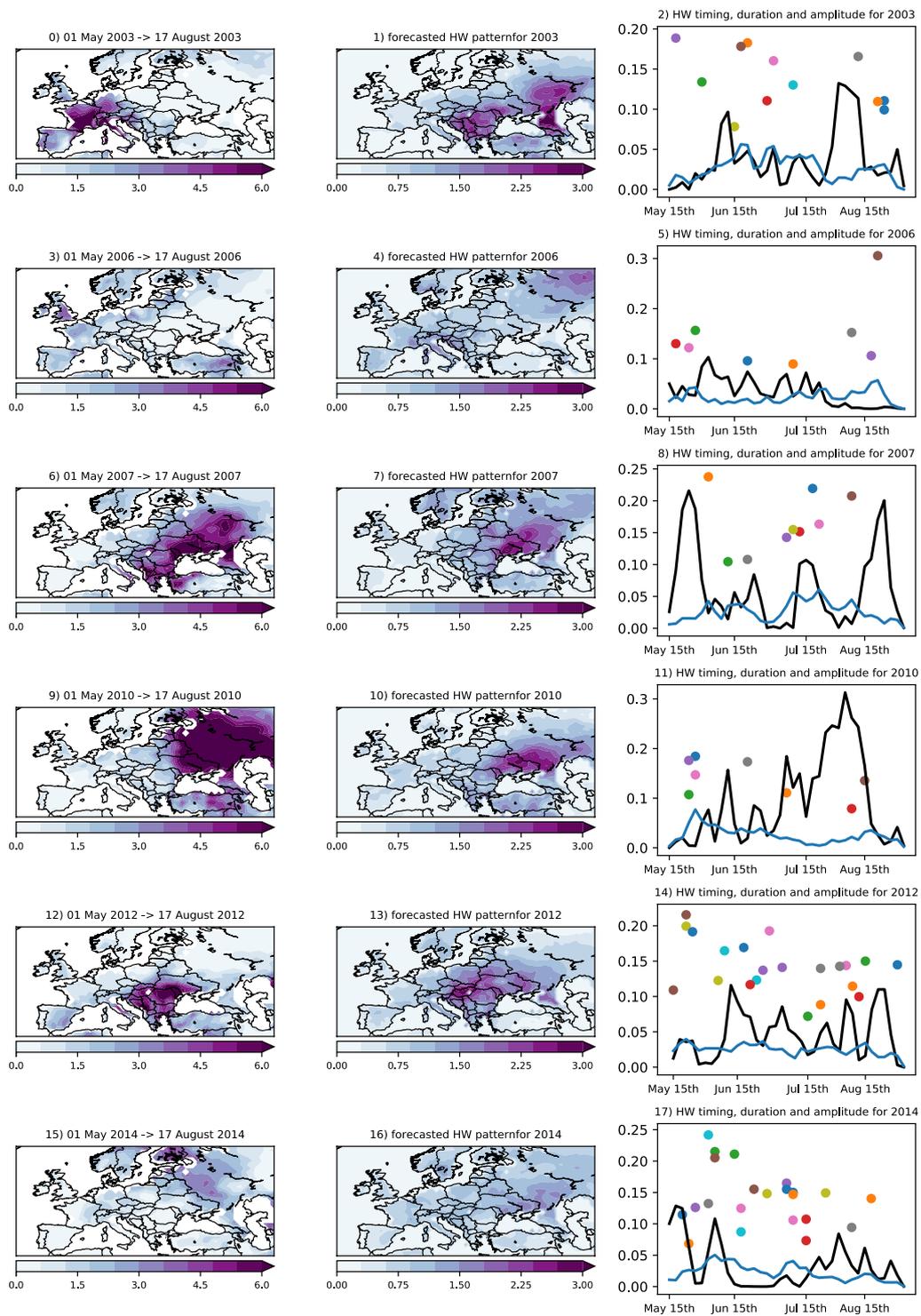
Fig. 9 (First column) Observed heatwave patterns (see Sect. 2) of the severe heatwaves defined with *HST* = 10. (Second column) Simulated heatwave patterns averaged for all heatwaves detected in SEAS5. (Third column) Black line: time evolution of the observed heatwave magnitude (weighted spatial sum of all sub-heatwaves as a function of time), blue line: ensemble mean of heatwave magnitude as a function of time for all members in which a heatwave is detected, colored dots show the day of maximum heatwave magnitude for each member of SEAS5 in which an heatwave is detected
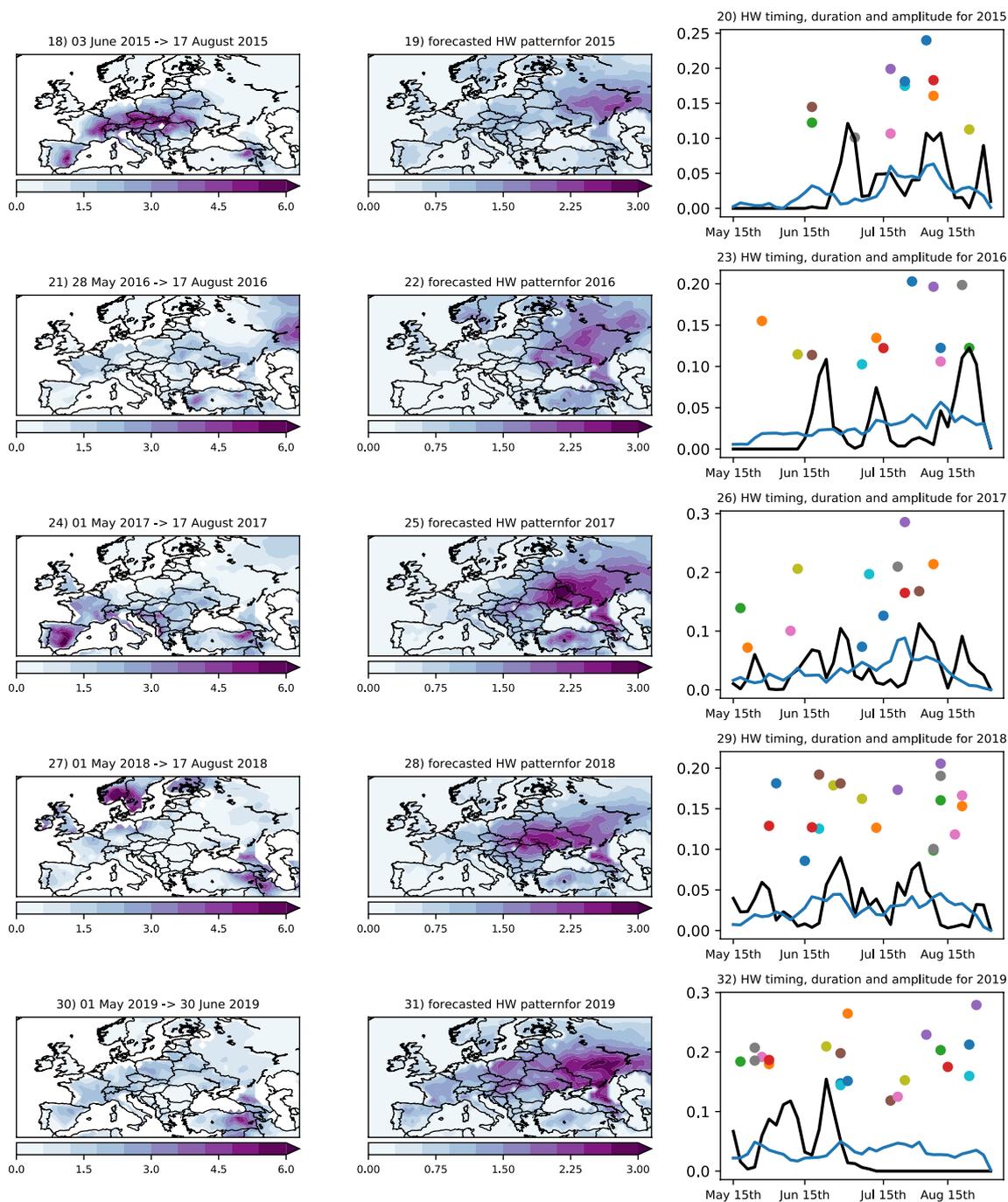
Fig. 9 (continued)

From our results, the predictability of Eastern Europe and Mediterranean heatwaves is not only related to the warming trend. This could allow us to investigate the source of predictability and mechanisms leading to the formations of these heatwaves (atmospheric circulation, large-scale teleconnections and land-atmosphere coupling). By comparing the ability to re-forecast past events of different seasonal prediction systems as a function of their different characteristics, we

could investigate potential mechanisms underlying severe heatwaves. Analysing the predictability of severe heatwave, using the systematic method presented in this study, in sensitivity experiments with different initial conditions, such Prodhomme et al. (2016b) and Ardilouze et al. (2017) for soil-moisture, could be used to quantify the impact of soil pre-conditioning on the occurrence and characteristics of severe heatwave. The application of the heatwave detection algorithm in sensitivity

experiments similar to (Wehrli et al. 2019), that prescribed SSTs, large scale circulation and/or soil conditions in a coupled models, would allow to understand more precisely the contribution of the different components and their biases on heatwave development and draw prospect for heatwave seasonal prediction improvements.

For European summer mean temperature, statistical prediction systems including large scale teleconnections, already exist (Della-Marta et al. 2007); our results suggest that similar work, targeting specifically heatwaves indices, would be beneficial. Knowing the importance of soil atmosphere-coupling in heatwave development, such models could also include soil preconditioning.

Our study focused on Europe. Although heatwaves are a global issue, very few studies have investigated heatwave seasonal predictability in Africa (Batté et al. 2018), America (Luo and Zhang 2012) and Asia (Gao et al. 2018). A similar methodology to the one presented here could be applied in these less studied regions, and also extended to marine heatwaves.

Our work demonstrates that heatwaves are partly predictable one or two months ahead, this could have important impacts on society for water management (Hartick et al. 2019; Turco et al. 2017), agriculture (Santos et al. 2020; Ceglar et al. 2018), health (Tompkins et al. 2018) and fire risk (Marcos et al. 2015). However, the heatwave predictability strongly varies depending on the considered index, regions, seasons and forecast horizons (Fig. 6). Similarly, Straaten et al. (2020) found that, for sub-seasonal temperature forecasts, skill horizons strongly differ depending on the region and that temporal and spatial aggregation does not systematically result in higher predictability. Therefore, in order to transform this predictability into valuable information for stakeholders and policy makers, careful evaluations of dedicated metrics, specifically targeted to the sector of interests should be performed. This kind of assessment implies strong trans-disciplinary collaboration, including social scientists, and integration of stakeholders and policy-makers. In addition, further work should be done to maximize heatwave predictability using statistical post-processing and multi-model combination (Mishra et al. 2019).

Appendix 1: Scores description

The skill scores used throughout this article are described here. The Pearson correlation coefficient is given by:

$$Cor(f) = \frac{\sum_{i=1}^{N_{year}} (m_i(f) - \bar{m}_i(f))(r_i(f) - \bar{r}(f))}{\sqrt{\sum_{i=1}^{N_{year}} (m_i(f) - \bar{m}(f))^2 \sum_{i=1}^{N} (r_i(f) - \bar{r}(f))^2}}$$

where f is the forecast time, $m_i(f)$ the forecast for the year i, $\bar{m}_i(f)$ the forecast averaged over the hindcast period excluding the year i (anomaly computed in cross-validation), $r_i(f)$ and $\bar{r}_i(f)$ the corresponding reference data at forecast time

f, and N_{year} the number of years in the re-forecast period. We provide uncertainty estimates and confidence interval with all the correlations values. We use the Student's t-distribution with N degrees of freedom to estimate the significance level of correlation. The significance of the difference between two correlations is estimated using the methodology of Siegert et al. (2016), which takes into account the dependence from sharing the same observations in both correlation coefficients and the number of independent data in each time series, which is necessary given the serial correlation typical of the time series considered. The Brier score is defined as:

$$BS(f) = \frac{1}{N} \sum_{i=1}^{N} (y_i(f) - o_i(f))^2$$

where $y_i(f)$ is the forecast probability of occurrence of a given event estimated from the ensemble distribution. In our case, an event could be the temperature reaching the upper/lower climatological quintile or to remain in the climatological inter-quantile range. Its probability of occurrence would be computed from the fraction of ensemble members predicting this event. $o_i(f)$ is the corresponding "observation" and has a value of 1 if the event happens, 0 otherwise. The Brier score is a distance in probability space. The Brier Skill Score is defined as:

$$BSS(f) = 1.0 - \frac{BS(f)}{BS(r)}$$

where $BS(r)$ is the Brier score of the reference probability forecast, typically a climatological forecast, for which e.g. the probability of occurrence of the upper and lower quintile are 20% and the interquintile range 60%.

For the reference model accounting for the warming trend used in the calculation of the BSS for Fig. S3, the probability for each year is defined by the following equation:

$$P_{1/5}(year) = \frac{2N_{year}}{N_{year} + 1}(1 - year)$$

$$P_{4/5}(year) = \frac{2N_{year}}{N_{year} + 1} year$$

where $P(1/5)$ and $P(4/5)$ are the probability to be below the first quintile and above last quintile, respectively. The probability to be in the interquintile range is constant: 3/5.

For verification of the detection of large heatwaves, as defined in Sect. 2.3.1, we use scores based on the contingency table:

Model	Observation	
	Heatwave	No heatwave
Heatwave	Hit	False alarm (FA)
No heatwave	Miss	Correct rejection (CR)

– Accuracy: proportion of forecasts which are correct between 0 and 1.

$$accuracy = \frac{hit + CR}{nyear}$$

– Hitrate: proportion of extreme events which are captured (between 0 and 1).

$$hitrate = \frac{hit}{hit + miss}$$

– Threat Score (TS): It measures the fraction of observed and predicted events that were both observed and predicted. It can be thought of as the accuracy when correct negatives have been removed from consideration. This score is sensitive to hits, penalizes both misses and false alarms. It depends on climatological frequency of events (poorer scores for rarer events) since some hits can occur purely due to random chance.

$$TS = \frac{hit}{hit + miss + FA}$$

– Equitable Threat Score: It measures the fraction of observed and predicted events that were both observed and predicted, adjusted for hits associated with random chance (for example, it is easier to correctly forecast rain occurrence in a wet climate than in a dry climate). It is sensitive to hits and it penalises both misses and false alarms in the same way.

$$ETS = \frac{hit - hitrandom}{hit + miss + FA - hitrandom}$$

where

$$hitrandom = \frac{(hit + miss)(hit + FA)}{nyear}$$

Appendix 2: ETS optimal probability threshold

In order to test the sensitivity of the severe heatwave detection to the probability threshold, we use two different methods for Fig. 8, a fixed probability threshold at 30% (in other words at least 8 members detecting a severe heatwave) and an optimal probability threshold calculated in cross-validation, as described below:

– Step 1: with a fixed value of the heatwave severity parameter n, we compute the ETS, for the whole hindcast period, except the year we want to forecast (cross-validation) for different value of the probability threshold: 10%, 20%, 30%, 40%, 50%, 60%.

– Step 2: Keep the probability threshold for which the ETS has the highest value.
– Step 3: Using the probability threshold defined in Step 2, we estimate if the model detect an heatwave for the considered year.
– Step 4: Perform Steps 1, 2 and for all the years and computed the ETS for the whole period.

References

Ardilouze C, Batté L, Bunzel F, Decremer D, Déqué M, Doblas-Reyes FJ, Douville H, Fereday D, Guemas V, MacLachlan C, Müller W, Prodhomme C (2017) Multi-model assessment of the impact of soil moisture initialization on mid-latitude summer predictability. Clim Dyn 49(11–12):3959–3974. https://doi.org/10.1007/s00382-017-3555-7

Ardilouze C, Batté L, Déqué M, van Meijgaard E, van den Hurk B (2019) Investigating the impact of soil moisture on European summer climate in ensemble numerical experiments. Clim Dyn 52(7–8):4011–4026

Ballester J, Robine JM, Herrmann FR, Rodó X (2011) Long-term projections and acclimatization scenarios of temperature-related mortality in Europe. Nat Commun 2(May):358. https://doi.org/10.1038/ncomms1360arXiv:1507.02142v2

Barbier J, Guichard F, Bouniol D, Couvreux F, Roehrig R (2018) Detection of intraseasonal large-scale heat waves: characteristics and historical trends during the Sahelian spring. J Clim 31(1):61–80. https://doi.org/10.1175/JCLI-D-17-0244.1

Barriopedro D, Fischer EM, Luterbacher J, Trigo RM, García-Herrera R (2011) The hot summer of 2010: redrawing the

temperature record map of Europe. Science (New York, NY) 332. https://doi.org/10.1126/science.1201224. http://www.ncbi.nlm.nih.gov/pubmed/21415316

Batté L, Ardilouze C, Déqué M (2018) Forecasting West African heat waves at subseasonal and seasonal time scales. Mon Weather Rev 146(3):889–907. https://doi.org/10.1175/MWR-D-17-0211.1

Bhend J, Mahlstein I, Liniger MA (2017) Predictive skill of climate indices compared to mean quantities in seasonal forecasts. Q J R Meteorol Soc 143(702):184–194. https://doi.org/10.1002/qj.2908

Brier GW (1950) Verification of forecasts expressed in terms of probability. Mon Weather Rev 78(1):1–3. https://doi.org/10.1175/1520-0493(1950)078<3c0001:VOFEIT>3e2.0.CO;2

Brunner L, Schaller N, Anstey J, Sillmann J, Steiner AK (2018) Dependence of present and future European temperature extremes on the location of atmospheric blocking. Geophys Res Lett 45(12):6311–6320. https://doi.org/10.1029/2018GL077837

Bunzel F, Müller WA, Dobrynin M, Fröhlich K, Hagemann S, Pohlmann H, Stacke T, Baehr J (2018) Improved seasonal prediction of European summer temperatures with new five-layer soil-hydrology scheme. Geophys Res Lett 45(1):346–353. https://doi.org/10.1002/2017GL076204

Cassou C, Terray L, Phillips A (2005) Tropical Atlantic influence on European heat waves. J Clim. https://doi.org/10.1175/JCLI3506.1

Ceglar A, Toreti A, Lecerf R, Van der Velde M, Dentener F (2016) Impact of meteorological drivers on regional inter-annual crop yield variability in France. Agric For Meteorol 216:58–67. https://doi.org/10.1016/j.agrformet.2015.10.004

Ceglar A, Toreti A, Prodhomme C, Zampieri M, Turco M, Doblas-Reyes FJ (2018) Land-surface initialisation improves seasonal climate prediction skill for maize yield forecast. Sci Rep. https://doi.org/10.1038/s41598-018-19586-6

Cornes RC, van der Schrier G, van den Besselaar EJM, Jones PD (2018) An ensemble version of the E-OBS temperature and precipitation data sets. J Geophys Res Atmos 123(17):9391–9409. https://doi.org/10.1029/2017JD028200

Della-Marta PM, Luterbacher J, von Weissenfluh H, Xoplaki E, Brunet M, Wanner H (2007) Summer heat waves over western Europe 1880–2003, their relationship to large-scale forcings and predictability. Clim Dyn 29(2–3):251–275. https://doi.org/10.1007/s00382-007-0233-1

Doblas-Reyes FJ, Hagedorn R, Palmer TN, Morcrette JJ (2006) Impact of increasing greenhouse gas concentrations in seasonal ensemble forecasts. Geophys Res Lett 33(7):L07708. https://doi.org/10.1029/2005GL025061

Dole R, Hoerling M, Perlwitz J, Eischeid J, Pegion P, Zhang T, Quan XW, Xu T, Murray D (2011) Was there a basis for anticipating the 2010 Russian heat wave? Geophys Res Lett 38(6):1–5. https://doi.org/10.1029/2010GL046582

Duchez A, Frajka-Williams E, Josey SA, Evans DG, Grist JP, Marsh R, McCarthy GD, Sinha B, Berry DI, Hirschi JJ (2016) Drivers of exceptionally cold North Atlantic Ocean temperatures and their link to the 2015 European heat wave. Environ Res Lett. https://doi.org/10.1088/1748-9326/11/7/074004

Feudale L, Shukla J (2011a) Influence of sea surface temperature on the European heat wave of 2003 summer. Part I: an observational study. Clim Dyn 36(9–10):1691–1703. https://doi.org/10.1007/s00382-010-0788-0

Feudale L, Shukla J (2011b) Influence of sea surface temperature on the European heat wave of 2003 summer. Part II: a modeling study. Clim Dyn 36(9–10):1705–1715. https://doi.org/10.1007/s00382-010-0789-z

Fischer EM, Schär C (2010) Consistent geographical patterns of changes in high-impact European heatwaves. Nat Geosci 3(6):1–6. https://doi.org/10.1038/ngeo866

Fischer EM, Seneviratne SI, Lüthi D, Schär C (2007) Contribution of land–atmosphere coupling to recent European summer heat waves. Geophys Res Lett 34(6):L06707. https://doi.org/10.1029/2006GL029068

Frich P, Alexander LV, Della-Marta P, Gleason B, Haylock M, Tank Klein AM, Peterson T (2002) Observed coherent changes in climatic extremes during the second half of the twentieth century. Clim Res 19(3):193–212. https://doi.org/10.3354/cr019193

Gao M, Wang B, Yang J, Dong W (2018) Are peak summer sultry heat wave days over the Yangtze–Huaihe River basin predictable? J Clim 31(6):2185–2196. https://doi.org/10.1175/JCLI-D-17-0342.1

García-Herrera R, Díaz J, Trigo RM, Luterbacher J, Fischer EM (2010) A review of the European summer heat wave of 2003. Crit Rev Environ Sci Technol 40(4):267–306. https://doi.org/10.1080/10643380802238137

Hartick C, Furusho C, Goergen K, Kollet S (2019) Interannual, probabilistic prediction of water resources over Europe following the heatwave and drought 2018. https://doi.org/10.31223/osf.io/h43xz

Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A, Soci C, Abdalla S, Abellan X, Balsamo G, Bechtold P, Biavati G, Bidlot J, Bonavita M, De Chiara G, Dahlgren P, Dee D, Diamantakis M, Dragani R, Flemming J, Forbes R, Fuentes M, Geer A, Haimberger L, Healy S, Hogan RJ, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, de Rosnay P, Rozum I, Vamborg F, Villaume S, Thépaut JN (2020) The ERA5 global reanalysis. Q J R Meteorol Soc 146(730):1999–2049. https://doi.org/10.1002/qj.3803

Hughes PD (2008) Response of a montenegro glacier to extreme summer heatwaves in 2003 and 2007. Geografiska Annaler Ser A Phys Geogr 90(4):259–267. https://doi.org/10.1111/j.1468-0459.2008.00344.x

Johnson SJ, Stockdale TN, Ferranti L, Balmaseda MA, Molteni F, Magnusson L, Tietsche S, Decremer D, Weisheimer A, Balsamo G, Keeley SP, Mogensen K, Zuo H, Monge-Sanz BM (2019) SEAS5: the new ECMWF seasonal forecast system. Geosci Model Dev 12(3):1087–1117. https://doi.org/10.5194/gmd-12-1087-2019

Jolliffe IT, Stephenson DB (2012) Forecast verification: a practitioner's guide in atmospheric science. Wiley, New York

Karoly DJ (2009) The recent bushfires and extreme heat wave in southeast Australia. Bull Aust Meteorol Oceanogr Soc 22(1):10–13

Katsafados P, Papadopoulos A, Varlas G, Papadopoulou E, Mavromatidis E (2014) Seasonal predictability of the 2010 Russian heat wave. Nat Hazards Earth Syst Sci 14(6):1531–1542. https://doi.org/10.5194/nhess-14-1531-2014

Kenyon J, Hegerl GC (2008) Influence of modes of climate variability on global temperature extremes. J Clim 21(15):3872–3889. https://doi.org/10.1175/2008JCLI2125.1

Lorenz R, Davin EL, Lawrence DM, Stö Ckli R, Seneviratne SI (2013) How important is vegetation phenology for European climate and heat waves? J Clim 26(24):10077–10100. https://doi.org/10.1175/JCLI-D-13-00040.1

Luo L, Zhang Y (2012) Did we see the 2011 summer heat wave coming? Geophys Res Lett. https://doi.org/10.1029/2012GL051383

Mahlstein I, Spirig C, Liniger MA, Appenzeller C (2015) Estimating daily climatologies for climate indices derived from climate model data and observations. J Geophys Res Atmos 120(7):2808–2818. https://doi.org/10.1002/2014JD022327

Marcos R, Turco M, Bedía J, Llasat MC, Provenzale A (2015) Seasonal predictability of summer fires in a Mediterranean environment. Int J Wildland Fire 24(8):1076. https://doi.org/10.1071/WF15079

Mecking JV, Drijfhout SS, Hirschi JJ, Blaker AT (2019) Ocean and atmosphere influence on the 2015 European heatwave. Environ Res Lett 14(11):114035. https://doi.org/10.1088/1748-9326/ab4d33

Miralles DG, Teuling AJ, Heerwaarden CCV (2014) Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. Nat Geosci 7(5):345–349. https://doi.org/10.1038/ngeo2141

Mishra N, Prodhomme C, Guemas V (2019) Multi-model skill assessment of seasonal temperature and precipitation forecasts over Europe. Clim Dyn 52(7–8):4207–4225. https://doi.org/10.1007/s00382-018-4404-z

Otto FEL, Massey N, van Oldenborgh GJ, Jones RG, Allen MR (2012) Reconciling two approaches to attribution of the 2010 Russian heat wave. Geophys Res Lett. https://doi.org/10.1029/2011GL050422

Pepler AS, Díaz LB, Prodhomme C, Doblas-Reyes FJ, Kumar A (2015) The ability of a multi-model seasonal forecasting ensemble to forecast the frequency of warm, cold and wet extremes. Weather Clim Extremes 9:68–77. https://doi.org/10.1016/j.wace.2015.06.005

Perkins SE (2015) A review on the scientific understanding of heatwaves—their measurement, driving mechanisms, and changes at the global scale. Atmos Res 164–165:242–267. https://doi.org/10.1016/J.ATMOSRES.2015.05.014

Prodhomme C, Batté L, Massonnet F, Davini P, Bellprat O, Guemas V, Doblas-Reyes FJ (2016a) Benefits of increasing the model resolution for the seasonal forecast quality in EC-earth. J Clim 29(24):9141–9162. https://doi.org/10.1175/JCLI-D-16-0117.1

Prodhomme C, Doblas-Reyes F, Bellprat O, Dutra E (2016b) Impact of land-surface initialization on sub-seasonal to seasonal forecasts over Europe. Clim Dyn 47(3–4):919–935. https://doi.org/10.1007/s00382-015-2879-4

Quesada B, Vautard R, Yiou P, Hirschi M, Seneviratne SI (2012) Asymmetric European summer heat predictability from wet and dry southern winters and springs. Nat Clim Change 2(10):736–741. https://doi.org/10.1038/nclimate1536

Robine JM, Cheung SLK, Le Roy S, Van Oyen H, Griffiths C, Michel JP, Herrmann FR (2008) Death toll exceeded 70,000 in Europe during the summer of 2003. Comptes Rendus Biologies 331(2):171–178. https://doi.org/10.1016/J.CRVI.2007.12.001

Russo S, Dosio A, Graversen RG, Sillmann J, Carrao H, Dunbar MB, Singleton A, Montagna P, Barbola P, Vogt JV (2014) Magnitude of extreme heat waves in present climate and their projection in a warming world. J Geophys Res Atmos 119(22):12500–12512. https://doi.org/10.1002/2014JD022098

Santos JA, Ceglar A, Toreti A, Prodhomme C (2020) Agricultural and forest meteorology performance of seasonal forecasts of Douro and Port wine production. Agric For Meteorol 291(June):108095. https://doi.org/10.1016/j.agrformet.2020.108095

Schoetter R, Cattiaux J, Douville H (2015) Changes of western European heat wave characteristics projected by the CMIP5 ensemble. Clim Dyn 45(5–6):1601–1616. https://doi.org/10.1007/s00382-014-2434-8

Siegert S, Bellprat O, Ménégoz M, Stephenson DB, Doblas-Reyes FJ (2017) Detecting improvements in forecast correlation skill: statistical testing and power analysis. Mon Weather Rev 145(2):437–450. https://doi.org/10.1175/MWR-D-16-0037.1

Sousa PM, Trigo RM, Barriopedro D, Soares PM, Santos JA (2018) European temperature responses to blocking and ridge regional patterns. Clim Dyn 50(1–2):457–477. https://doi.org/10.1007/s00382-017-3620-2

Straaten C, Whan K, Coumou D, Hurk B, Schmeits M (2020) The influence of aggregation and statistical post processing on the sub-seasonal predictability of European temperatures. Q J R Meteorol Soc 146(731):2654–2670. https://doi.org/10.1002/qj.3810

Thomson MC, Doblas-reyes FJ, Mason SJ, Hagedorn R, Connor SJ, Phindela T, Morse AP, Palmer TN (2006) Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. Nature 439:576–579. https://doi.org/10.1038/nature04503

Tompkins AM, Lowe R, Nissan H, Martiny N, Roucou P, Thomson MC, Nakazawa T (2019) Predicting climate impacts on health at sub-seasonal to seasonal timescales. In Robertson AW, Vitart F (eds) The gap between weather and climate forecasting: sub-seasonal to seasonal prediction, Elsevier, pp 455–477. https://doi.org/10.1016/B978-0-12-811714-9.00022-X

Turco M, Ceglar A, Prodhomme C, Soret A, Toreti A, Doblas-Reyes Francisco J (2017) Summer drought predictability over Europe: empirical versus dynamical forecasts. Environ Res Lett. https://doi.org/10.1088/1748-9326/aa7859

Vogel MM, Zscheischler J, Fischer EM, Seneviratne SI (2020) Development of future heatwaves for different hazard thresholds. J Geophys Res Atmos 125(9):e2019JD032070. https://doi.org/10.1029/2019JD032070

Wehrli K, Guillod BP, Hauser M, Leclair M, Seneviratne S (2019) Identifying key driving processes of major recent heat waves. J Geophys Res Atmos 124(22):11746–11765. https://doi.org/10.1029/2019JD030635

Weisheimer A, Doblas-Reyes FJ, Jung T, Palmer TN (2011) On the predictability of the extreme summer 2003 over Europe. Geophys Res Lett. https://doi.org/10.1029/2010GL046455

Welbergen JA, Klose SM, Markus N, Eby P (2008) Climate change and the effects of temperature extremes on Australian flying-foxes. Proc R Soc B Biol Sci 275(1633):419–425. https://doi.org/10.1098/rspb.2007.1385

Whan K, Zscheischler J, Orth R, Shongwe M, Rahimi M, Asare EO, Seneviratne SI (2015) Impact of soil moisture on extreme maximum temperatures in Europe. Weather Clim Extrem. https://doi.org/10.1016/j.wace.2015.05.001

Wolf G, Czaja A, Brayshaw DJ, Klingaman NP (2020) Connection between sea surface anomalies and atmospheric quasi-stationary waves. J Clim 33(1):201–212. https://doi.org/10.1175/JCLI-D-18-0751.1

World Meteorological Organization (2019) July matched, and maybe broke, the record for the hottest month since analysis began. WMO, Geneva

Wulff CO, Domeisen DI (2019) Higher subseasonal predictability of extreme hot European summer temperatures as compared to average summers. Geophys Res Lett 46(20):11520–11529. https://doi.org/10.1029/2019GL084314

Wulff CO, Greatbatch RJ, Domeisen DI, Gollan G, Hansen F (2017) Tropical forcing of the summer east Atlantic pattern. Geophys Res Lett 44(21):11166–11173. https://doi.org/10.1002/2017GL075493

Zhu J, Huang B, Cash B, Kinter JL, Manganello J, Barimalala R, Altshuler E, Vitart F, Molteni F, Towers P (2015) ENSO prediction in project minerva: sensitivity to atmospheric horizontal resolution and ensemble size. J Clim 28(5):2080–2095. https://doi.org/10.1175/JCLI-D-14-00302.1

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