

# A New Framework for Identifying and Investigating Seasonal Climate Extremes

MATTHIAS RÖTHLISBERGER,<sup>a</sup> MAURO HERMANN,<sup>a</sup> CHRISTOPH FREI,<sup>b</sup> FLAVIO LEHNER,<sup>c,a,d</sup> ERICH M. FISCHER,<sup>a</sup>  
RETO KNUTTI,<sup>a</sup> AND HEINI WERNLI<sup>a</sup>

<sup>a</sup> *Institute for Atmospheric and Climate Science, ETH Zürich, Zürich, Switzerland*

<sup>b</sup> *Federal Office of Meteorology and Climatology MeteoSwiss, Zürich, Switzerland*

<sup>c</sup> *Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, New York*

<sup>d</sup> *National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 10 December 2020, in final form 17 June 2021)

**ABSTRACT:** Previous studies have recognized the societal relevance of climatic extremes on the seasonal time scale and examined physical processes leading to individual high-impact extreme seasons (e.g., extremely wet or warm seasons). However, these findings have not yet been generalized beyond case studies since at any specific location only very few seasonal events of such rarity occurred in the observational record. In this concept paper, a pragmatic approach to pool seasonal extremes across space is developed and applied to investigate hot summers and cold winters in ERA-Interim and the Community Earth System Model Large Ensemble (CESM-LENS). We identify spatial extreme season objects as contiguous regions of extreme seasonal mean temperatures based on statistical modeling. Regional pooling of extreme season objects in CESM-LENS then yields considerable samples of analogs to even the most extreme ERA-Interim events. This approach offers numerous opportunities for systematically analyzing large samples of extreme seasons, and several such analyses are illustrated. We reveal a striking co-occurrence of El Niño to La Niña transitions and the largest ERA-Interim midlatitude extreme summer events. Moreover, we perform a climate model evaluation with regard to extreme season size and intensity measures and estimate how often an extreme winter like the cold North American 2013/14 winter is expected anywhere in midlatitude regions. Furthermore, we present a large set of simulated analogs to this event, which makes it possible to study commonalities and differences of their underlying physical processes. Finally, substantial but spatially varying climatological differences in the size of extreme summer and extreme winter objects are identified.

**SIGNIFICANCE STATEMENT:** Recent climate extremes on the seasonal time scale, for example, extremely hot summers, had massive socioeconomic impacts. However, by definition, only very few such rare seasonal events occurred at any specific location during the last few decades, which strongly limits the ways in which they can be studied scientifically. Here we propose a pragmatic approach for pooling comparable seasonal extremes occurring in different locations, to increase the number of events available for subsequent analyses. We first identify the most extreme hot summers and cold winters in 40 years of observation-based data and 1200 years of climate model data. Then, we demonstrate how spatial pooling of comparable events makes it possible to study high-impact seasonal extremes in novel and so far unexplored ways.

**KEYWORDS:** Climate variability; Surface temperature; Statistical techniques; Statistics; Climate models; Ensembles; General circulation models; Model evaluation/performance; Reanalysis data; Seasonal variability

## 1. Introduction

Defining weather and climate extremes is a multidimensional problem (Cattiaux and Ribes 2018), which involves at least three dimensions, namely the spatial extent, the intensity, and the duration of an extreme event. Along the last of those dimensions, the relevant meteorology varies greatly. In the midlatitudes, for example, 10-min precipitation extremes typically occur within intense convective systems, while extreme

precipitation accumulated over 1–3 days may be associated with the passage of an intense front (Catto and Pfahl 2013), a generally wavy upper-level flow configuration (Röthlisberger et al. 2016) with upper-level Rossby wave breaking (Martius et al. 2006; Moore et al. 2019), an extratropical cyclone (Raveh-Rubin and Wernli 2015), or a combination thereof. On the weekly time scale, precipitation extremes can arise due to a temporal clustering of extratropical cyclones and landfalling atmospheric rivers (e.g., Moore et al. 2020), and seasonal precipitation extremes may result from anomalously recurring synoptic upper-level flow features, whose temporal organization can be influenced by slower-varying tropical sea surface temperature anomalies (Davies 2015).

The socioeconomic impacts of extremes also vary substantially with the time scale of the extremes. For example, multihour precipitation extremes in the European Alps led to local

 Denotes content that is immediately available upon publication as open access.

 Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-20-0953.s1>.

*Corresponding author:* Matthias Röthlisberger, [matthias.roethlisberger@env.ethz.ch](mailto:matthias.roethlisberger@env.ethz.ch)

DOI: 10.1175/JCLI-D-20-0953.1



This article is licensed under a [Creative Commons Attribution 4.0 license](http://creativecommons.org/licenses/by/4.0/) (<http://creativecommons.org/licenses/by/4.0/>).

flooding and destroyed infrastructure (Rössler et al. 2014). A 10-day wet spell during an already wet winter in Northern California led to an emergency spill-over of the Oroville Dam and forced the temporary evacuation of 188 000 people living downstream of this dam (White et al. 2019; Moore et al. 2020), while the extremely cold and wet European summer of 1816, which followed the Tambora volcanic eruption, caused widespread famine in Europe and allegedly spurred mass emigration to North America (Luterbacher and Pfister 2015). These examples illustrate that the impacts of extremes on different time scales vary not only in magnitude (damage costs, fatalities, etc.) but also in their nature. Clearly, the distinct physical causes and impacts of extremes on different time scales underline the importance of investigating extremes on each time scale.

On the seasonal time scale, particularly high-impact events have been observed during the last decades. Recent years have seen numerous hot summers in many regions of the world (e.g., Schär et al. 2004; Barriopedro et al. 2011; Wehner et al. 2016) but also cold winters (Cattiaux et al. 2010; Wolter et al. 2015), wet and stormy winters (Huntingford et al. 2014) or particularly active hurricane seasons (Klotzbach et al. 2018). The socioeconomic impacts of these seasonal events were severe, often long-lasting; they occurred over almost continental-scale areas and were extremely diverse. During recent hot and dry European summers such as 2003 and 2018, for instance, mortality rates strongly increased in some parts of Europe (Díaz et al. 2006; Fouillet et al. 2006; Robine et al. 2008; García-Herrera et al. 2010; Barriopedro et al. 2011). Moreover, reduced gross primary production has been reported in these two summers (Buras et al. 2020), resulting in additional net release of carbon dioxide to the atmosphere (Ciais et al. 2005), alongside with forest damage, unusual wild fire activity (Krikken et al. 2021), increased energy prices (Fink et al. 2004), and reduced transport capacity of the Rhine River (Jonkeren et al. 2014). Similar lists of diverse impacts of seasonal extreme events could be compiled also for other such events, for example, the wet and stormy winter over the United Kingdom in 2013/14 (Huntingford et al. 2014; Matthews et al. 2014; Davies 2015). The evident socioeconomic relevance of seasonal extremes makes them an attractive and relevant research topic.

Numerous previous studies have used both reanalysis and climate model data to assess the physical drivers of extreme seasons, however, almost exclusively focusing on individual case studies. Examples include the pioneering study of Namias (1978) on “the abnormal North American winter 1976/77” or, more recently Fischer et al. (2007), focusing on soil moisture–atmosphere interactions during the European summer 2003, Cattiaux et al. (2010), discussing the cold 2009/10 European winter, and Davies (2015), who investigated the temporal organization of synoptic flow features during the 2013/14 winter in the Northern Hemisphere. The case study approach provides a detailed process understanding and “storyline” (Trenberth et al. 2015; Shepherd et al. 2018) of individual extreme events. Therefore, all these studies provide valuable insight into the functioning of the selected seasonal extreme, and, in combination they offer a rough idea about their diversity. However, more systematic analyses that aim at a statistical (climatological) characterization of the spatiotemporal

characteristics of and the relevant physical mechanisms behind a certain type of seasonal extremes are not available so far. Consequently, these climatological aspects of extreme seasons so far remain poorly understood, which is unsatisfactory given the socioeconomic impact of extreme seasons.

The purpose of this concept paper is to develop a pragmatic approach for spatially pooling seasonal extreme events, that enables such climatological analyses of seasonal extremes at a regional scale. To this end, we propose an identification scheme for spatial extreme season objects, which we define as contiguous regions of extreme (i.e., locally rare) seasonal conditions with regard to a specific meteorological parameter, for example, 2-m temperature (section 3). Hereby, the rarity of the seasonal extremes within their local (i.e., grid point level) climatology is comparable across space. We apply the scheme to ERA-Interim data (ERA-I; Dee et al. 2011) and data from the Community Earth System Model Large Ensemble project (CESM-LENS; Kay et al. 2015) to identify large sets of extremely hot summers and cold winters, respectively (sections 4–6).

Two terms are central to the remainder of this paper, namely, *climatological analyses* and *seasonal extremes*. Here we use the term *climatological analyses* in contrast to case study analyses. The defining characteristic of a climatological analysis is that it uses statistical techniques to analyze a large sample of events. Consequently, the identified general or climatological characteristics of extreme seasons are not only valid for an individual event but are more general characteristics of a whole sample of extreme season events. With *seasonal extreme* we refer to a seasonal mean value that is only very rarely exceeded locally, that is, a seasonal mean value with a large return period at the exact location where it occurs [local return period (LRP)]. We consider LRPs on the order of a few decades and larger as seasonal extremes.

## 2. Data and data preprocessing

We use JJA and December–February (DJF) seasonal-mean 2-m temperature ( $T_{2m}$ ) data from ERA-I and CESM-LENS. CESM-LENS is the National Center for Atmospheric Research (NCAR) large ensemble, which is based on CESM, version 1 (Hurrell et al. 2013), and contains an ensemble of 30 members that were integrated from 1920 to 2100, using historical forcing up to 2005 and RCP8.5 forcing thereafter. We use  $T_{2m}$  seasonal means at their native spatial resolution of  $1.25^\circ$  longitude  $\times$   $\sim 0.9^\circ$  latitude for the period March 1979–February 2019, yielding 1200 seasonal mean values per grid point and type of season. We consider January and February 2019 as part of the 2018 winter and hereafter refer to this 40-yr period as the 1979–2018 period. ERA-I data are originally produced on T255 spatial resolution with 60 vertical levels (Dee et al. 2011). Here we use JJA and DJF seasonal  $T_{2m}$  means for the same 1979–2018 period that were first computed from 6-hourly  $T_{2m}$  values and then linearly interpolated to the CESM grid. For both winter and summer, we thus obtain 40 values per grid point in ERA-I.

We detrend the seasonal means before identifying extremes, by subtracting a spatially varying and nonlinear forced trend,

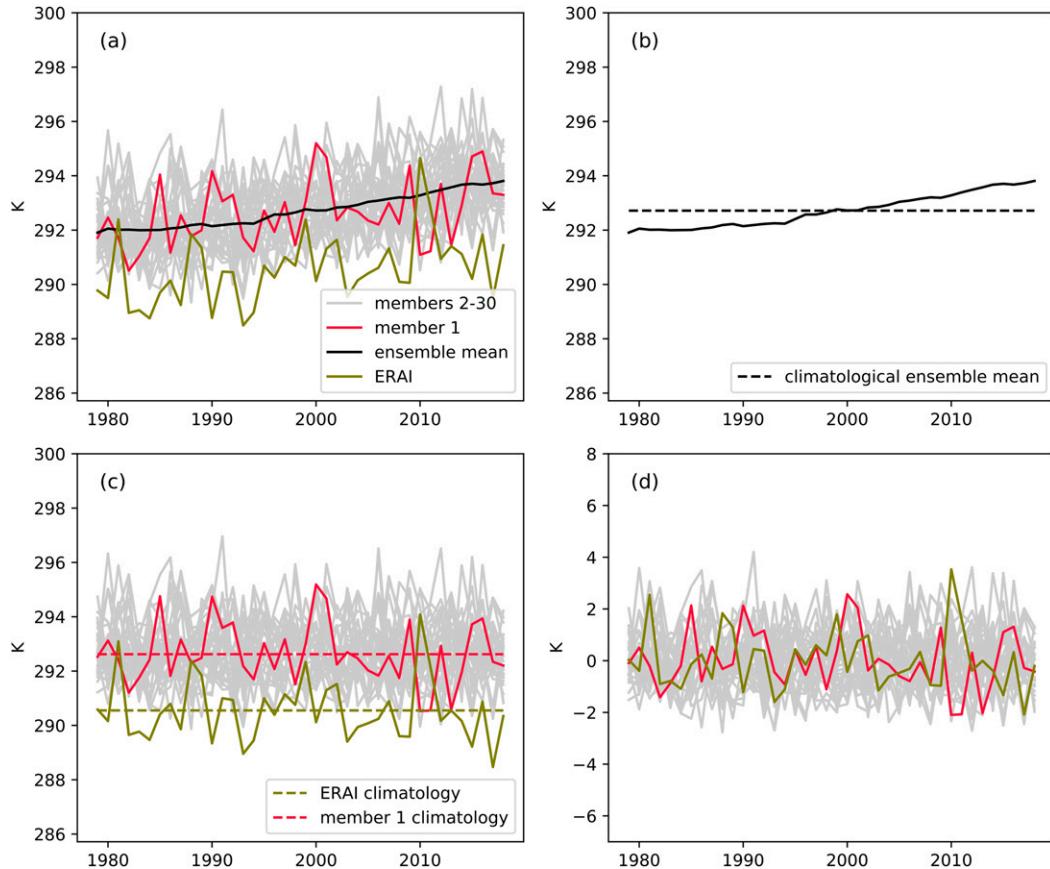


FIG. 1. The detrending procedure used in this study illustrated for the grid point  $56^{\circ}\text{N}$ ,  $38^{\circ}\text{E}$  (Moscow, Russia). (a) Raw JJA means of  $T_{2m}$  for the individual CESM-LENS members (red and gray lines) and ERAI (green). The solid black line in (a) and (b) depicts the ensemble mean smoothed with a 5-yr running average. (b) The climatological ensemble mean is shown as dashed line, and the difference between the solid and the dashed line is the “forced trend,” which is subtracted from the raw data. (c) The detrended JJA means of  $T_{2m}$  and (d) the anomalies  $T_{\text{CESM}}^{\text{JJA}}$  and  $T_{\text{ERAI}}^{\text{JJA}}$  obtained by subtracting the local 1979–2018 JJA  $T_{2m}$  climatological value, computed respectively within each ensemble member and ERAI [shown for CESM-LENS member 1 and ERAI in (c)].

estimated as the seasonal CESM-LENS ensemble mean anomaly relative to the CESM-LENS ensemble mean climatology of 1979–2018, smoothed with a 5-yr running average (Fig. 1, note that for computing the smoothed forced trend we use CESM-LENS data from 1977 to 2020). Removing this forced trend estimate from both datasets retains unforced multiannual to decadal-scale variability in the individual CESM-LENS members as well as in ERAI, which is important as such internal variability conceivably plays an important role in the occurrence of extreme seasons. Furthermore, in the supplemental material we show that over extratropical land areas the 1979–2018 *linear* trends in ERAI typically fall within the range of the linear trends from the individual CESM-LENS ensemble members, while over tropical and subtropical oceans CESM-LENS appears to overestimate the linear trend in JJA and DJF  $T_{2m}$  means (see Fig. S1 in the online supplemental material), and thus likely also the forced trend. However, as the focus of this study will be primarily on extreme seasons over extratropical land areas, this model bias unlikely affects our conclusions.

In a second step, seasonal mean anomalies are computed in both datasets. To ensure maximum comparability between ERAI and CESM-LENS data, CESM-LENS seasonal mean anomalies are computed separately for each ensemble member, relative to the 1979–2018 climatology of the respective member, while ERAI seasonal anomalies are computed by removing the ERAI 1979–2018 climatological mean (Figs. 1c,d). For each grid point  $g$  and dataset  $M$ , the detrended seasonal mean  $T_{2m}$  anomaly in year  $y$  is hereafter denoted by  $T_{g,y}^M$ , but the subscripts  $g$  and  $y$  are omitted whenever possible without loss of clarity. Hot boreal summer extremes and cold boreal winter extremes are hereafter referred to as  $T_{\text{JJA}+}$  and  $T_{\text{DJF}-}$ , respectively. Both datasets used in this study serve as examples, but our conceptual reasoning applies equally to other observation-based datasets and climate simulations.

Furthermore, in section 4 we use a time series of the Niño-3.4 index as downloaded from NOAA’s web page (NOAA 2021), and in section 6, ERAI and CESM-LENS sea surface temperature (SST) data are used and preprocessed exactly as  $T_{2m}$ .

### 3. An identification scheme for extreme season objects

The purpose of this section is to develop a generic identification scheme for spatially coherent extreme season objects that can be applied to short reanalysis datasets as well as potentially much longer climate simulations. We propose an identification scheme consisting of the following three steps:

- 1) Estimate at each grid point the distribution of seasonal mean values by statistical modeling and calculate the LRP for each value.
- 2) Identify locally extreme seasonal mean values based on exceedances of an LRP threshold  $\tau$ .
- 3) Form spatially coherent objects within which  $LRP > \tau$ , and quantify their characteristics, for example, their size and intensity.

Somewhat arbitrarily, we choose  $\tau = 40$  years, as discussed and justified in detail in section 3b. However, none of our conceptual arguments depend on the exact value of  $\tau$ .

#### a. Statistical modeling of seasonal $T_{2m}$ anomalies

Robust estimation of LRPs for rare events requires statistical modeling, particularly when the record length is limited, as is the case with ERAI. Distributional assumptions based on the asymptotic property of block maxima (extreme value theory; see Fisher and Tippett 1928; Coles 2001), which are commonly chosen for modeling shorter time-scale extreme events (e.g., Zwiers and Kharin 1998), are poorly justified for seasonal means. Moreover, the assumption of a normal (Gaussian) distribution, motivated by the central limit theorem and implicitly made with standardized anomalies (e.g., Chase et al. 2006; Barriopedro et al. 2011), may be too simplistic for seasonal mean temperatures at the grid point scale. For example, Figs. 2a and 2b illustrate that  $T^{\text{CESM}}$  values have skewed empirical distributions in many regions, which cannot be adequately described with a symmetric normal distribution.

For the statistical modeling in this study, we therefore combine the normal distribution with a parameterized data transformation to allow for modeling skewed distributions. Specifically, we assume that seasonal mean  $T_{2m}$  anomalies follow a Yeo–Johnson transformed normal distribution  $N^{\text{YJ}}$ . Its cumulative distribution (cdf) is given by

$$G(x; \mu, \sigma^2, \lambda) = F[k(x; \lambda); \mu, \sigma^2], \quad (1)$$

where  $k(x; \lambda)$  is the Yeo–Johnson power transform of a variable  $x$  (Yeo and Johnson 2000) with transformation parameter  $\lambda$  and  $F(y; \mu, \sigma^2)$  is the cdf of the normal distribution with mean  $\mu$  and variance  $\sigma^2$ . By the chain rule, the probability density function (pdf) of  $N^{\text{YJ}}$  is thus

$$g(x; \mu, \sigma^2, \lambda) = f[k(x; \lambda); \mu, \sigma^2] \frac{dk(x; \lambda)}{dx}, \quad (2)$$

where  $f(y; \mu, \sigma^2)$  is the pdf of the normal distribution [see the supplemental material for more details and an explicit formulation of  $g(x; \mu, \sigma^2, \lambda)$ ]. The distribution parameters  $\mu$ ,  $\sigma^2$ , and  $\lambda$  (the transformation exponent) are considered to vary between grid points  $g$  and datasets  $M$ . For their estimation we adopt maximum likelihood estimation,

using procedures implemented in Python's SciPy package (Oliphant 2007).

The incorporation of a power transformation, such as that of Yeo and Johnson (2000), is an essential element of our statistical modeling, because it introduces the flexibility to account for the variable levels of skewness observed over the globe (Figs. 2a,b). It is worth mentioning that alternatives would have been available to model modest departures from normality, for example, the five-parameter skewed generalized  $T$  distribution (Theodossiou 1998), or subfamilies of it (see, e.g., McDonald and Newey 1988; Hansen 1994). In view of the limited sample size in ERAI, and the satisfactory goodness-of-fit analysis (see below), we have opted for the more parsimonious  $N^{\text{YJ}}$ .

Figures 2c and 2d show that in most regions of the world, the empirical distribution of the transformed  $T^{\text{CESM}}$  values  $k(T^{\text{CESM}})$  is not in contradiction with the normal distribution, suggesting there is no evident conflict between the data and our assumption of the  $N^{\text{YJ}}$ . Exceptions are found primarily over the tropical oceans and the high Arctic, which are not the focus of this study. Also, for ERAI we find only very few rejections of the  $N^{\text{YJ}}$  with summer and winter mean  $T_{2m}$  data (Fig. S2), but bear in mind that a goodness-of-fit test for ERAI, with 40 values per grid point only, has a much smaller power to identify conflicts, than for CESM with 1200 values per grid point. Clearly, conflicts between the  $N^{\text{YJ}}$  and seasonal  $T_{2m}$  data may still be present at higher moments (e.g., heavy or short tails), which are difficult to identify with limited sample sizes. Nevertheless, based on Figs. 2c and 2d and Fig. S2, we consider the  $N^{\text{YJ}}$  a reasonable compromise between flexibility and parsimony that lends itself for modeling seasonal mean temperatures over a wide range of climates as well as for the small samples of ERAI.

For any JJA  $T_{g,y}^M$  value in dataset  $M$ , a parametric estimate of its local return period, denoted as  $LRP_p(T_{g,y}^M)$ , is then

$$LRP_p(T_{g,y}^M) = \frac{1}{1 - G(T_{g,y}^M)} = \frac{1}{1 - F[k(T_{g,y}^M)]}. \quad (3)$$

For  $T_{\text{DJF-}}$  extremes, the interest is in the lower (cold) tail of the respective distribution of seasonal mean values and, hence, the LRP is computed as

$$LRP_p(T_{g,y}^M) = \frac{1}{G(T_{g,y}^M)} = \frac{1}{F[k(T_{g,y}^M)]}. \quad (4)$$

#### b. Choosing an LRP threshold $\tau$ to identify seasonal extremes

In particular in ERAI, estimates of large LRPs are uncertain because of the limited sample size. An identification scheme that relies on the criterion  $LRP > \tau$  is, therefore, subject to identification errors, particularly if  $\tau$  is comparable to or larger than the record length. False positives occur when LRPs near  $\tau$  are overestimated, and conversely for false negatives. In the following we investigate the skill of the identification scheme for a range of  $\tau$  values to find a good compromise between our interest in extremes (large  $\tau$ ) and the accuracy of their identification (small  $\tau$ ). This skill is examined for the example of

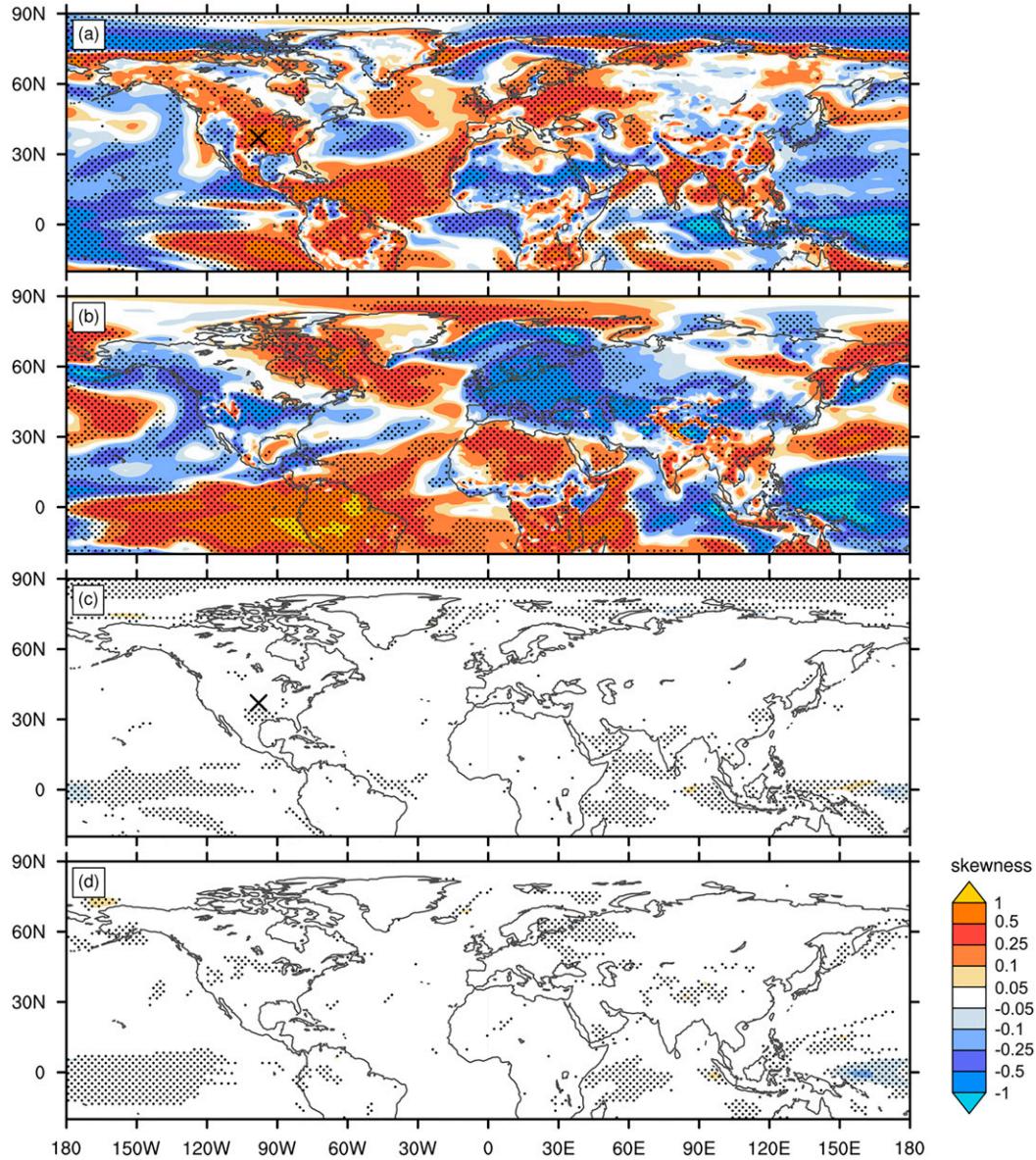


FIG. 2. Skewness of the 1200 (a) JJA and (b) DJF  $T^{\text{CESM}}$  values. Stippling indicates grid points where the empirical distribution of seasonal  $T_{2m}$  anomalies deviates from a normal distribution according to a Shapiro–Wilks test at a significance level of 0.05. (c),(d) As in (a) and (b), but for the power-transformed JJA and DJF  $T^{\text{CESM}}$  values  $k(T^{\text{CESM}})$ . The black cross in (a) and (c) marks a grid point for which the goodness-of-fit of the  $N^{YJ}$  distribution is examined in more detail in Fig. S3.

$T_{\text{JJA}+}$  extremes in each individual CESM-LENS member  $n$  (denoted as  $\text{CESM}n$ ). By construction, the remaining members (referred to as  $\text{CESM}^*n$ ) share the same climate as  $\text{CESM}n$ , and an empirical LRP estimate for any  $T_{g,y}^{\text{CESM}n}$  value,  $\text{LRP}_e(T_{g,y}^{\text{CESM}n})$ , can be calculated from simply counting how often  $T_{g,y}^{\text{CESM}n}$  is exceeded in the remaining 29 CESM-LENS members, that is,

$$\text{LRP}_e(T_{g,y}^{\text{CESM}n}) = \frac{1160}{|\{T_{g,y}^{\text{CESM}^*n} | T_{g,y}^{\text{CESM}^*n} \geq T_{g,y}^{\text{CESM}n}\}|} \quad (5)$$

In this way, an *empirical* LRP estimate for any CESM-LENS seasonal anomaly,  $\text{LRP}_e(T_{g,y}^{\text{CESM}n})$ , can be computed and compared to the respective parametric LRP estimate  $\text{LRP}_p(T_{g,y}^{\text{CESM}n})$ , whereby  $\text{LRP}_p(T_{g,y}^{\text{CESM}n})$  is computed using only the 40 values in member  $n$ . Note that  $\text{LRP}_e(T_{g,y}^{\text{CESM}n})$  uses 1160 values and is thus a much more informed estimate of the true LRP compared to  $\text{LRP}_p(T_{g,y}^{\text{CESM}n})$ . Evaluating  $\text{LRP}_p(T_{g,y}^{\text{CESM}n})$  against  $\text{LRP}_e(T_{g,y}^{\text{CESM}n})$  therefore allows us to assess how well  $T_{\text{JJA}+}$  extremes can be identified in a 40-yr dataset like ERAI based on the  $N^{YJ}$ . This

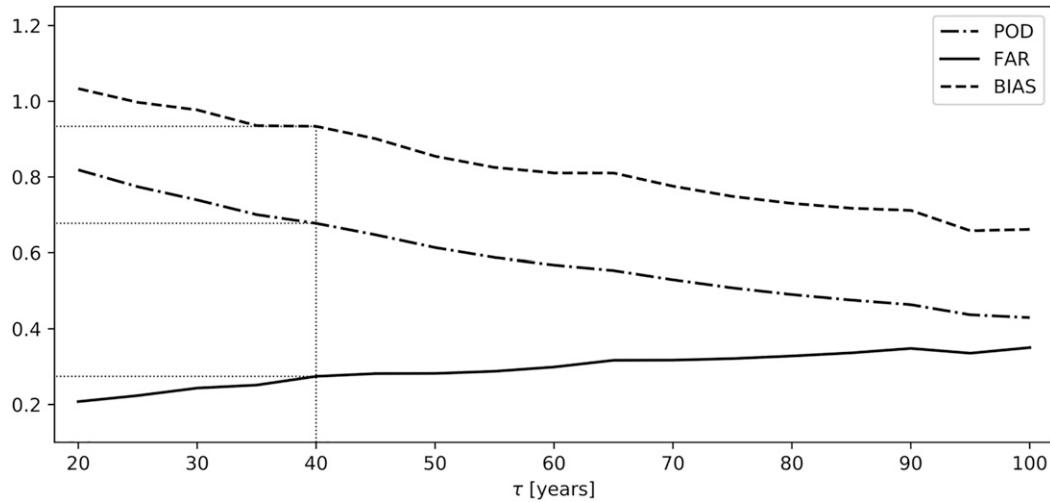


FIG. 3. Skill measures for identifying extreme  $T^{\text{CESM}}$  values based on  $\text{LRP}_p(T^{\text{CESM}})$  for different values of  $\tau$  from 20 to 100 years. The  $\text{BIAS}(\tau)$  is depicted with the dashed line while the  $\text{POD}(\tau)$  and the  $\text{FAR}(\tau)$  are shown by the dash-dotted and solid lines, respectively. Light dotted lines help to read off the values for  $\tau = 40$  years.

evaluation is performed using data from each member, and therefore we omit the ensemble member index  $n$  hereafter.

Figure 3 depicts the probability of detection (POD), the false alarm ratio (FAR), and the bias (BIAS) of this evaluation for various values of  $\tau$ . Each score is computed globally, that is, by aggregating data from all grid points and years. The POD quantifies the fraction of the extreme  $T^{\text{CESM}}$  values identified based on  $\text{LRP}_e(T^{\text{CESM}})$  that is also identified as extreme using  $\text{LRP}_p(T^{\text{CESM}})$ , that is,

$$\text{POD}(\tau) = \frac{|\{T^{\text{CESM}} | \text{LRP}_e(T^{\text{CESM}}) > \tau \wedge \text{LRP}_p(T^{\text{CESM}}) > \tau\}|}{|\{T^{\text{CESM}} | \text{LRP}_e(T^{\text{CESM}}) > \tau\}|} \quad (6)$$

The FAR quantifies the fraction of the extremes identified based on  $\text{LRP}_p(T^{\text{CESM}})$  that are not deemed extreme based on  $\text{LRP}_e(T^{\text{CESM}})$  and is computed as

$$\text{FAR}(\tau) = \frac{|\{T^{\text{CESM}} | \text{LRP}_p(T^{\text{CESM}}) > \tau \wedge \text{LRP}_e(T^{\text{CESM}}) \leq \tau\}|}{|\{T^{\text{CESM}} | \text{LRP}_p(T^{\text{CESM}}) > \tau\}|} \quad (7)$$

while the BIAS relates the total number of extremes identified based on  $\text{LRP}_p(T^{\text{CESM}})$  to the total number of extremes identified based on  $\text{LRP}_e(T^{\text{CESM}})$ , that is,

$$\text{BIAS}(\tau) = \frac{|\{T^{\text{CESM}} | \text{LRP}_p(T^{\text{CESM}}) > \tau\}|}{|\{T^{\text{CESM}} | \text{LRP}_e(T^{\text{CESM}}) > \tau\}|} \quad (8)$$

A value of 1 thus indicates no bias in the number of extremes, while  $\text{BIAS}(\tau) < 1$  and  $\text{BIAS}(\tau) > 1$  imply that too few or too many  $T^{\text{CESM}}$  values are identified as extremes based on the parametric estimate, respectively.

Figure 3 shows that  $\text{BIAS}(\tau)$  decreases from 1.03 for  $\tau = 20$  years to 0.66 for  $\tau = 100$  years, and thus for larger  $\tau$  the number of extremes is substantially underestimated based

on  $\text{LRP}_p(T^{\text{CESM}})$ . This bias arises because an unusually rare value in a small sample (i.e., an “outlier,” e.g., a 100-yr event in a sample of 40 seasonal means) influences the parameter estimates such that the fitted distribution will consider it more likely. Such bias is rather fundamental and cannot be overcome with an alternative estimation method. Underestimation of extremes (and overestimation of average conditions) is a side effect of statistical estimation and predictions that aim at minimizing error variance. It is commonly described as “conditional bias” or “regression to the mean” (e.g., Stigler 1997; Potts 2012) and is a notorious problem in the estimation of extreme events in the presence of uncertainty (e.g., Seo et al. 2014; Frei and Isotta 2019). A further consequence of the conditionally biased parameter estimation is that the occurrence of multiple large  $\text{LRP}_p$  values at the same grid point within a 40-yr period will be underestimated.

The FAR and POD metrics in Fig. 3 have to be interpreted with this bias in mind (for any  $\tau$  the largest possible POD equals the BIAS), and show that for  $\tau = 40$  years, 68% (POD) of the extremes identified with  $\text{LRP}_p(T^{\text{CESM}})$  are also identified as extremes based on  $\text{LRP}_e(T^{\text{CESM}})$ , while 27% (FAR) of the extremes identified with  $\text{LRP}_p(T^{\text{CESM}})$  are not extremes according to  $\text{LRP}_e(T^{\text{CESM}})$ . The FAR for  $\tau = 40$  years thus implies that if an extreme is identified based on  $\text{LRP}_p(T^{\text{CESM}})$  it is also identified as extreme based on  $\text{LRP}_e(T^{\text{CESM}})$  in 73% of all such cases. Furthermore, Figs. S4 and S5 reveal that for a given season the spatial patterns of  $\text{LRP}_e(T^{\text{CESM}})$  and  $\text{LRP}_p(T^{\text{CESM}})$  are very similar, but the bias of  $\text{LRP}_p(T^{\text{CESM}})$  for large LRPs leads to frequent “near misses,” which are nevertheless strictly counted as misses in the POD. Our choice of  $\tau = 40$  years is somewhat arbitrary but given the conceptual nature of the following analyses we find the skill measures for this threshold acceptable. We nevertheless test the sensitivity of our results to variations in  $\tau$  from 20 to 100 years in the following sections.

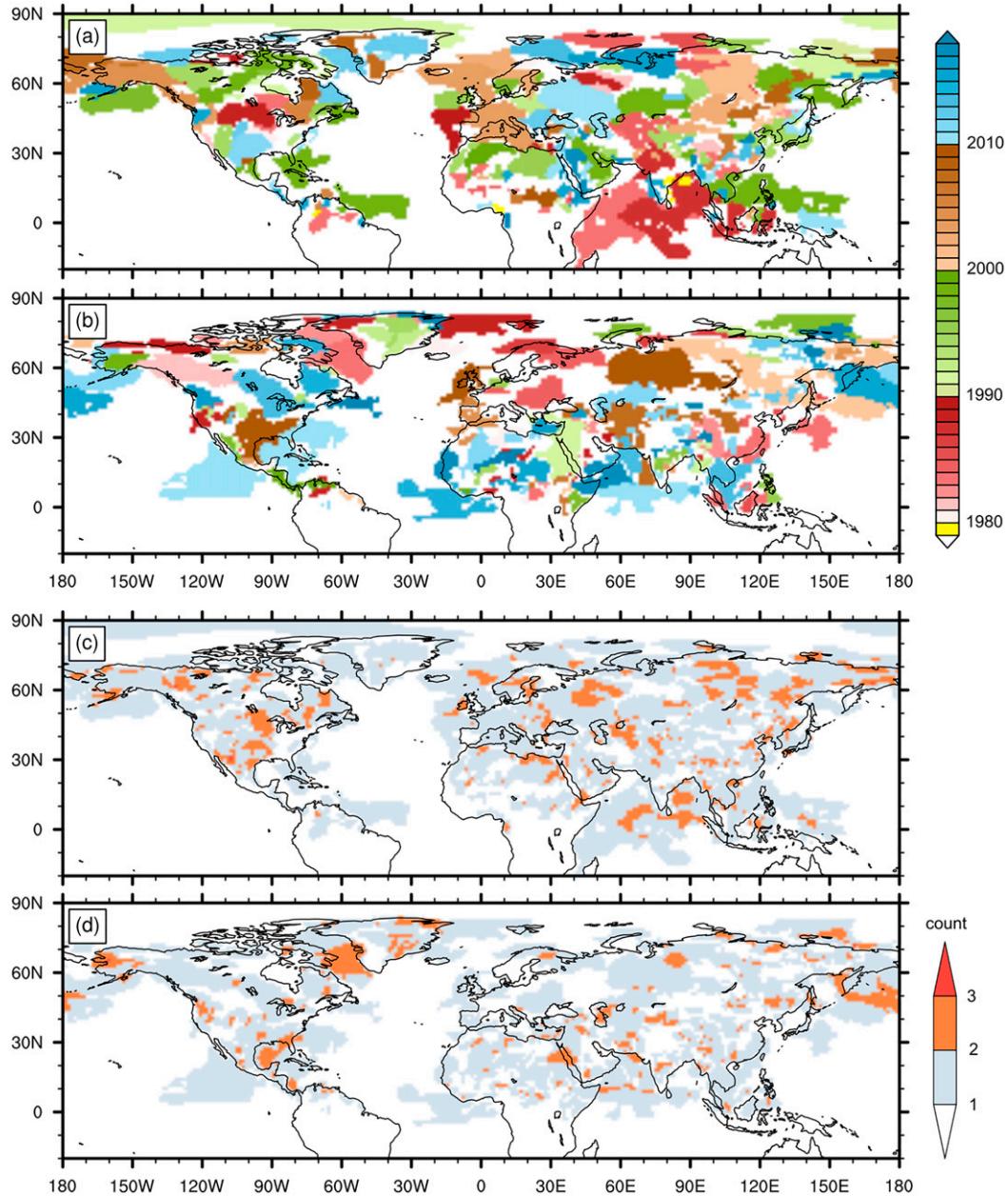


FIG. 4. (a),(b) Years of Northern Hemisphere ERAI  $T_{JJA+}$  and  $T_{DJF-}$  extreme season objects with  $A_L > 10^5 \text{ km}^2$ , respectively. For grid points that feature more than one object, the year of the latest object is shown. (c),(d) The number of ERAI  $T_{JJA+}$  and  $T_{DJF-}$  objects with  $A_L > 10^5 \text{ km}^2$  per grid point, respectively.

### c. Forming coherent spatial extreme season objects

In the third step we form coherent spatial extreme season objects for each season  $y$  in dataset  $M$ , by considering for each grid point where  $\text{LRP} > \tau$  all eight neighboring grid points as valid connections. A neighboring grid point is connected if its  $\text{LRP}$  also exceeds  $\tau$ . For the so constructed extreme season objects, we then compute the center of mass ( $\text{COM}$ ) as well as the area ( $A$ ), an intensity measure ( $I$ ), computed as the area weighted mean  $T^M$  (i.e., seasonal  $T_{2m}$  anomaly) over the object

and, analogously, the land center of mass ( $\text{COM}_L$ ), land area ( $A_L$ ), and land intensity ( $I_L$ ), by only considering land grid points of a particular extreme season object.

## 4. Extremely hot summers and cold winters in ERAI

Applying this identification scheme to ERAI  $T_{2m}$  anomalies in JJA and DJF yields 175  $T_{JJA+}$  objects and 147  $T_{DJF-}$  objects with a land area exceeding  $10^5 \text{ km}^2$  in the Northern

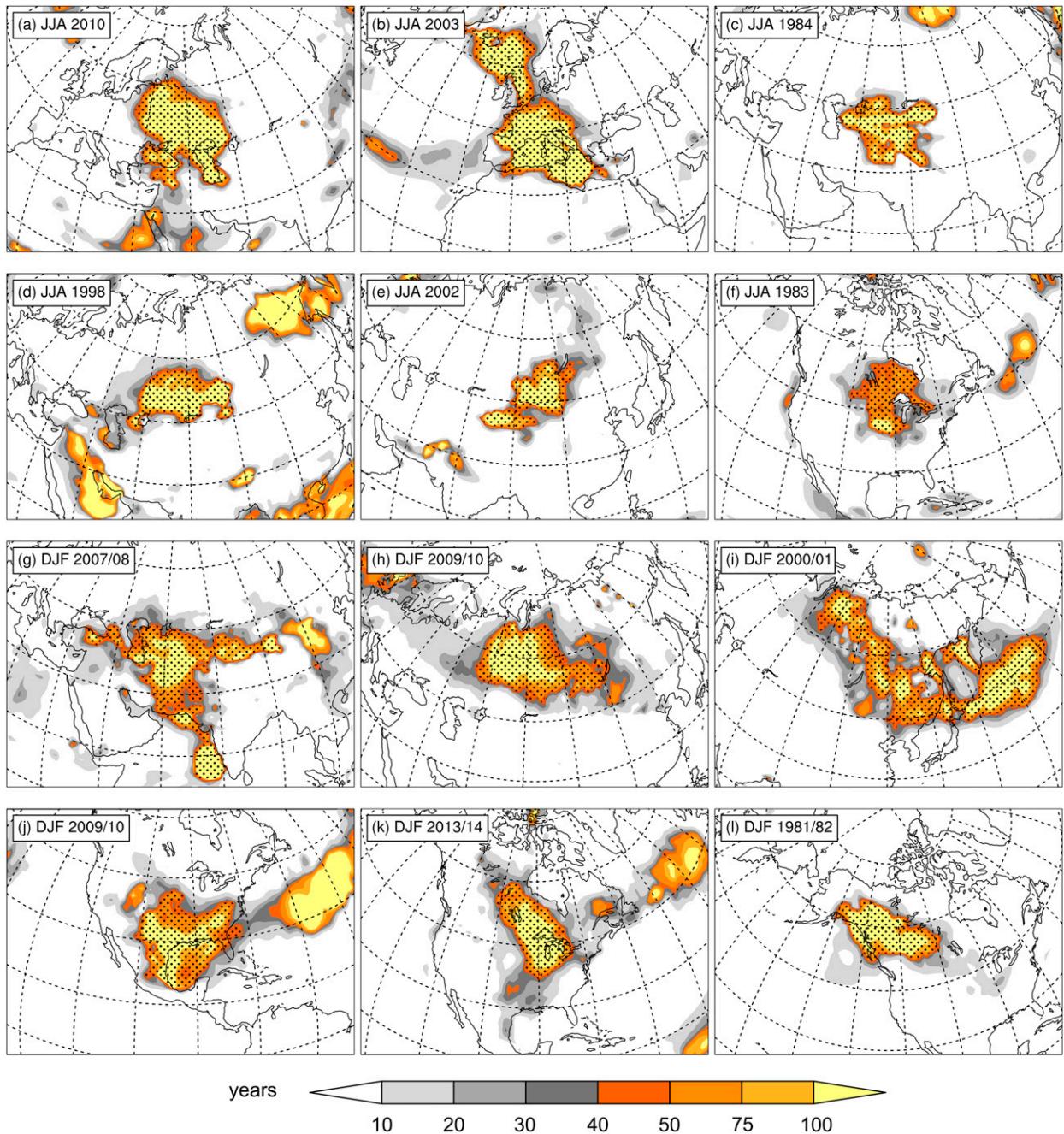


FIG. 5. The top six ERAI (a)–(f)  $T_{JJA+}$  and (g)–(l)  $T_{DJF-}$  objects according to  $A_L$  with a  $COM_L$  between  $30^\circ$  and  $60^\circ\text{N}$ . Stippled areas mark the respective objects. All panels show  $LRP_p(T_{gy}^{ERA1})$  of the respective season. Contour levels have been chosen such that they correspond to local return periods values that also appear in Tables 2 and 3. For reference, gridlines are drawn every  $15^\circ$  longitude and latitude.

Hemisphere (NH; Figs. 4a,b). Smaller objects as well as objects located entirely over the oceans are excluded from further analyses in this study. Most but by far not all NH land grid points in ERAI contribute to one  $T_{JJA+}$  object, while some even contribute to up to three  $T_{JJA+}$  objects (Fig. 4c; e.g., at  $30^\circ\text{N}$ ,  $110^\circ\text{W}$ ). Similar results with regard to the spatial distribution of events are obtained for  $T_{DJF-}$  seasons (Fig. 4d).

Focusing now on midlatitude objects with a  $COM_L$  between  $30^\circ$  and  $60^\circ\text{N}$ , Fig. 5 depicts the top six ERAI  $T_{JJA+}$  and  $T_{DJF-}$  events according to  $A_L$ , and Table 1 summarizes their characteristics. For a few events in Fig. 5 there is clearly some ambiguity as to whether patches of large LRPs should be considered as a single or as multiple separate events. In such cases, varying  $\tau$  in our identification scheme leads to splitting or

TABLE 1. The top six  $T_{\text{JJA}+}$  and  $T_{\text{DJF}-}$  extreme seasons in ERAI according to  $A_L$  in the midlatitudes (i.e., with a  $\text{COM}_L$  between  $30^\circ$  and  $60^\circ\text{N}$ ) and their land area and land intensity characteristics.

	Year	Region/country	$A_L$ ( $10^6 \text{ km}^2$ )	$I_L$ (K)
$T_{\text{JJA}+}$	2010	Western Russia	3.85	2.92
	2003	Europe	2.91	2.36
	1984	Central Asia	2.53	1.89
	1998	Kazakhstan	2.43	2.49
	2002	Mongolia/China	2.02	1.98
	1983	United States/Canada	1.83	2.21
$T_{\text{DJF}-}$	2007/08	Central Asia	4.41	-3.69
	2009/10	Russia	4.02	-7.33
	2000/01	Siberia/China	4.01	-5.45
	2009/10	Southern United States/Mexico	2.96	-2.89
	2013/14	Central United States	2.61	-5.01
	1981/82	Canada/Alaska	2.23	-6.10

merging of objects. Table 2 therefore examines how sensitive the identification of the largest midlatitude objects is to variations of  $\tau$  from 20 to 100 years. For such variations in  $\tau$ , all but two of the events in Fig. 5 remain within the top 10 NH midlatitude events with regard to  $A_L$ . Larger changes in the ranks occur for extreme season objects that either break up when  $\tau$  is increased or merge with nearby objects when  $\tau$  is decreased (e.g.,  $T_{\text{DJF}-}$  in 2000/01 in Siberia when increasing  $\tau$  from 40 to 75 years).

Several of the extreme seasons in Fig. 5 are well-known high-impact events, for example, the hot summers in western Russia in 2010 (Fig. 5a; Barriopedro et al. 2011; Dole et al. 2011) and in Europe in 2003 (Fig. 5b; Schär et al. 2004; Fink et al. 2004; Black et al. 2004; Fischer et al. 2007), or the highly abnormal Northern Hemisphere winter 2009/10 (Fig. 5h; Cattiaux et al. 2010; Sprenger et al. 2017) and the 2013/14 cold winter in the United States (Fig. 5k; Davies 2015; Wolter et al. 2015). However, also some less-well-studied events appear, such as

the summer 1984 in central Asia (Fig. 5c) or the winter 1981/82 in Canada and Alaska (Fig. 5l).

Despite their distinct geographical location, it is interesting to compare the physical mechanisms that have been reported for some of these events. For several  $T_{\text{JJA}+}$  events for example, persistent upper-level anticyclonic flow, often in the form of atmospheric blocking, has been identified as a key characteristic (Black et al. 2004; Fink et al. 2004; Trenberth and Fasullo 2012; Schneidereit et al. 2012). Blocking leads to subsidence and clear-sky conditions, which both foster hot surface conditions through adiabatic compression and increase incoming shortwave radiation at the surface (Bieli et al. 2015; Zschenderlein et al. 2019). Furthermore, in several of these events, dry soils preceding and persisting during the events have further exacerbated the surface heat (Fischer et al. 2007; Lau and Kim 2012; Wehrli et al. 2019). The physical mechanism linking persistent upper-level anticyclonic flow to persistent surface heat is hence clear and basic meteorological reasoning suggests that it operates everywhere in

TABLE 2. The top six  $T_{\text{JJA}+}$  and  $T_{\text{DJF}-}$  extreme seasons in ERAI according to  $A_L$  in the midlatitudes (i.e., with a  $\text{COM}_L$  between  $30^\circ$  and  $60^\circ\text{N}$ ). The five rightmost columns list their ranks (by  $A_L$ ) when using various values of  $\tau$  for their identification.

	Year	Region/country	Rank $_{\tau=100}$	Rank $_{\tau=75}$	Rank $_{\tau=40}$	Rank $_{\tau=30}$	Rank $_{\tau=20}$
$T_{\text{JJA}+}$	2010	Western Russia	1	1	1	1	1
	2003	Europe	2	2	2	2	3
	1984	Central Asia	5	3	3	3	7
	1998	Kazakhstan	3	4	4	4	4
	2002	Mongolia/China	8	7	5	5	5
	1983	United States/Canada	30	32	6	6	6
$T_{\text{DJF}-}$	2007/08	Central Asia	5	5	1	3	2
	2009/10	Russia	1	1	2	1	2
	2000/01	Siberia/China	12	10	3	2	3
	2009/10	Southern United States/Mexico	4	4	4	6	6
	2013/14	Central United States	3	2	5	7	7
	1981/82	Canada/Alaska	2	3	6	8	9

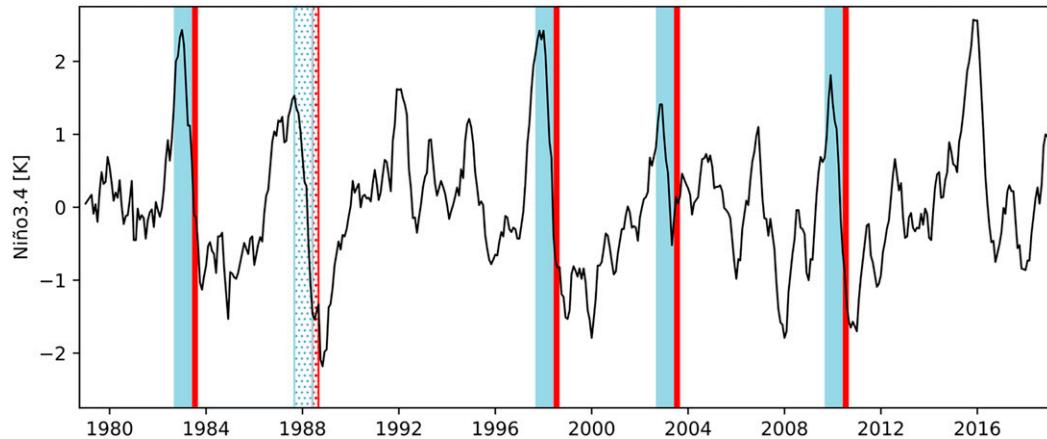


FIG. 6. Time series of Niño-3.4 anomalies relative to their 1981–2010 mean and the periods of the 1983, 1998, 2003, and 2010  $T_{JJA+}$  events (Figs. 5a,b,d,f; red bars in this figure). The blue bars show the 9 months preceding the respective event. The 1988 event is shown with stippled bars.

the midlatitudes. The key question is thus why the anticyclonic circulation was so prevalent and persistent during these seasons in the respective regions.

For the  $T_{JJA+}$  event in western Russia, Fragkoulidis et al. (2018) emphasized the role of recurrent transient Rossby wave packets and illustrated that during this summer a sequence of such wave packets approached western Russia in just the right phasing to repeatedly amplify the block there. Moreover, Sedláček et al. (2011) argued that the joint action of teleconnections from the Arabian and Barents Seas instigated a stationary wave pattern that organized the transients favorably to the development of blocking over western Russia. Somewhat similarly, Trenberth and Fasullo (2012) and Schneider et al. (2012) underlined the relevance of the rapid El Niño to La Niña transition in the first half of 2010 for setting up a stationary wave pattern with an anticyclonic center over western Russia. Interestingly, also for the  $T_{JJA+}$  events in 1983, 1998, and 2003 (Figs. 5b,d,f) as well as the hot North American summer of 1988 (which ranks seventh with regard to  $A_L$  in the midlatitudes), the peak period of an El Niño event occurred within nine months prior to the respective event, while neutral or La Niña conditions prevailed during the respective summer (Fig. 6). Thus, despite occurring on three different continents, these five major  $T_{JJA+}$  events evidently share characteristics in their physical storyline. For the 2010  $T_{JJA+}$  event, Trenberth and Fasullo (2012) argued that the stationary wave pattern forced from the emerging La Niña conditions was enhanced over western Russia, by wave forcing from anomalously high sea surface temperatures over the Caribbean and the Bay of Bengal, which they considered to be remnants of the previous El Niño episode. To what extent similar El Niño remnants played a role in the 1983, 1988, 1998, or 2003  $T_{JJA+}$  events is as of yet unclear, but the example certainly illustrates the value of spatially pooling events for unraveling common physical characteristics of major seasonal extreme events. Yet, analyses of shared or “climatological” extreme season characteristics based on reanalyses data alone are strongly limited

by the small number of comparable events. Therefore, in the subsequent sections we explore to what extent CESM-LENS can be used to expand the sample size.

### 5. Climatological evaluation of CESM-LENS with regard to extreme season characteristics

Model evaluation is of fundamental importance when attempting to use climate model data for expanding the sample of observed seasonal extreme events. CESM has already been evaluated in much detail in previous studies with regard to numerous aspects of synoptic to seasonal-scale dynamics that are relevant to extreme seasons. The model appeared to perform rather well in simulating, for example, the relationship between blocking and heat waves over western Russia (Schaller et al. 2018), weather regime frequencies over Europe (Huguenin et al. 2020), and distributions of daily maximum and minimum temperatures, as well as trends in heat-wave frequencies (Perkins-Kirkpatrick et al. 2017) and the substructure of hot summers in the Northern Hemisphere (Röthlisberger et al. 2020). Our extreme season identification scheme now enables further ways of model evaluation, regarding specifically seasonal extreme characteristics.

To maximize the comparability of CESM-LENS and ERAI extreme season objects we apply the scheme separately to each CESM-LENS member, and first contrast the 40-yr return levels for JJA and DJF  $T_{2m}$  anomalies in ERAI and CESM-LENS (Figs. 7 and 8). Herein these 40-yr return levels are denoted by  $RL_g^M$ , and correspond to the respective  $T_g^M$  value for which  $LRP_p(T_g^M) = 40$  years. Note that  $RL^{CESM}$  in Figs. 7b and 8b is the average of the 30 return level values from the individual CESM-LENS members. For JJA  $T_{2m}$  anomalies the hemispheric spatial pattern of  $RL^{CESM}$  and  $RL^{ERAI}$  share many common features, such as  $RL < 0.5$  K over some tropical oceans and return levels of several K over mid- and high-latitude landmasses. The absolute values of  $RL^{CESM}$  and  $RL^{ERAI}$ , however, can differ substantially. For example, over northern Siberia CESM clearly overestimates the 40-yr return

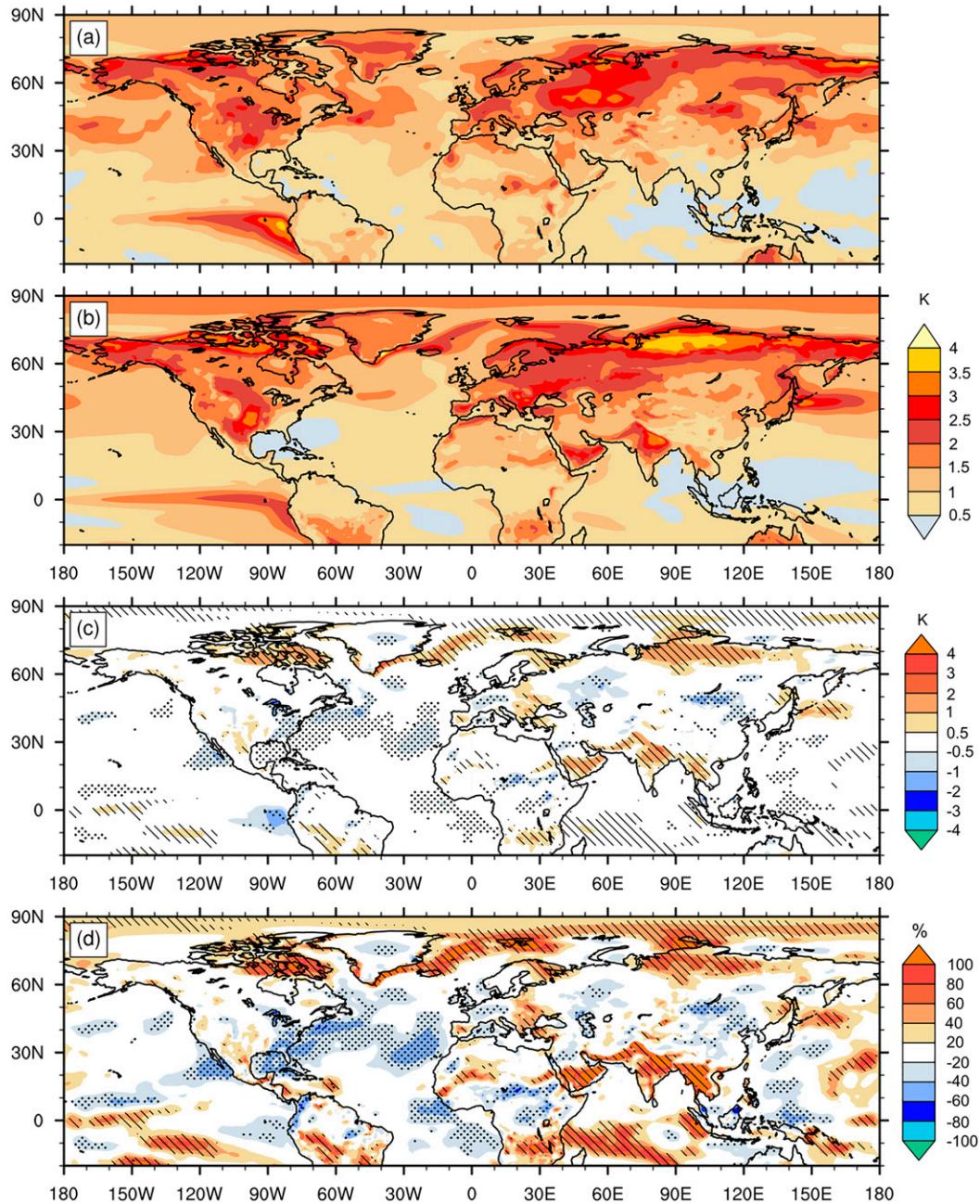


FIG. 7. The 40-yr return levels for (a) positive JJA  $T_{2m}$  anomalies in ERAI ( $RL^{ERAI}$ ) and (b) the ensemble-average 40-yr return level for positive JJA  $T_{2m}$  anomalies in CSM-LENS ( $RL^{CESM}$ ). (c),(d) Difference (b) minus (a) in K and relative to  $RL^{ERAI}$  in %, respectively. Hatching in (c) and (d) denotes grid points where all 30 return levels from the individual CSM-LENS members exceed  $RL^{ERAI}$ , and stippling indicates where they are all below  $RL^{ERAI}$ .

level (by up to 1.5 K), and  $RL^{ERAI}$  falls outside the range of the 30 individual ensemble member return levels (Fig. 7d). Over the midlatitude landmasses  $RL^{ERAI}$  and  $RL^{CESM}$  typically differ by less than 0.5 K (mostly less than 40% of  $RL^{ERAI}$ ; Figs. 7c,d).

Also, for DJF  $T_{2m}$  anomalies, the hemispheric patterns of  $RL^{ERAI}$  and  $RL^{CESM}$  are in reasonable agreement and feature return levels below  $-8$  K in some regions (Figs. 8a,b). The

absolute values of the return levels over the extratropical landmasses are typically larger for DJF than for JJA and thus also differ more strongly between ERAI and CSM in DJF compared to JJA (Figs. 7c and 8c). For both seasons,  $RL^{ERAI}$  over the midlatitude land areas is within the range of the individual member  $RL^{CESM}$  values almost everywhere (Figs. 7c,d and 8c,d), which, however, is not true for many subtropical and tropical ocean regions.

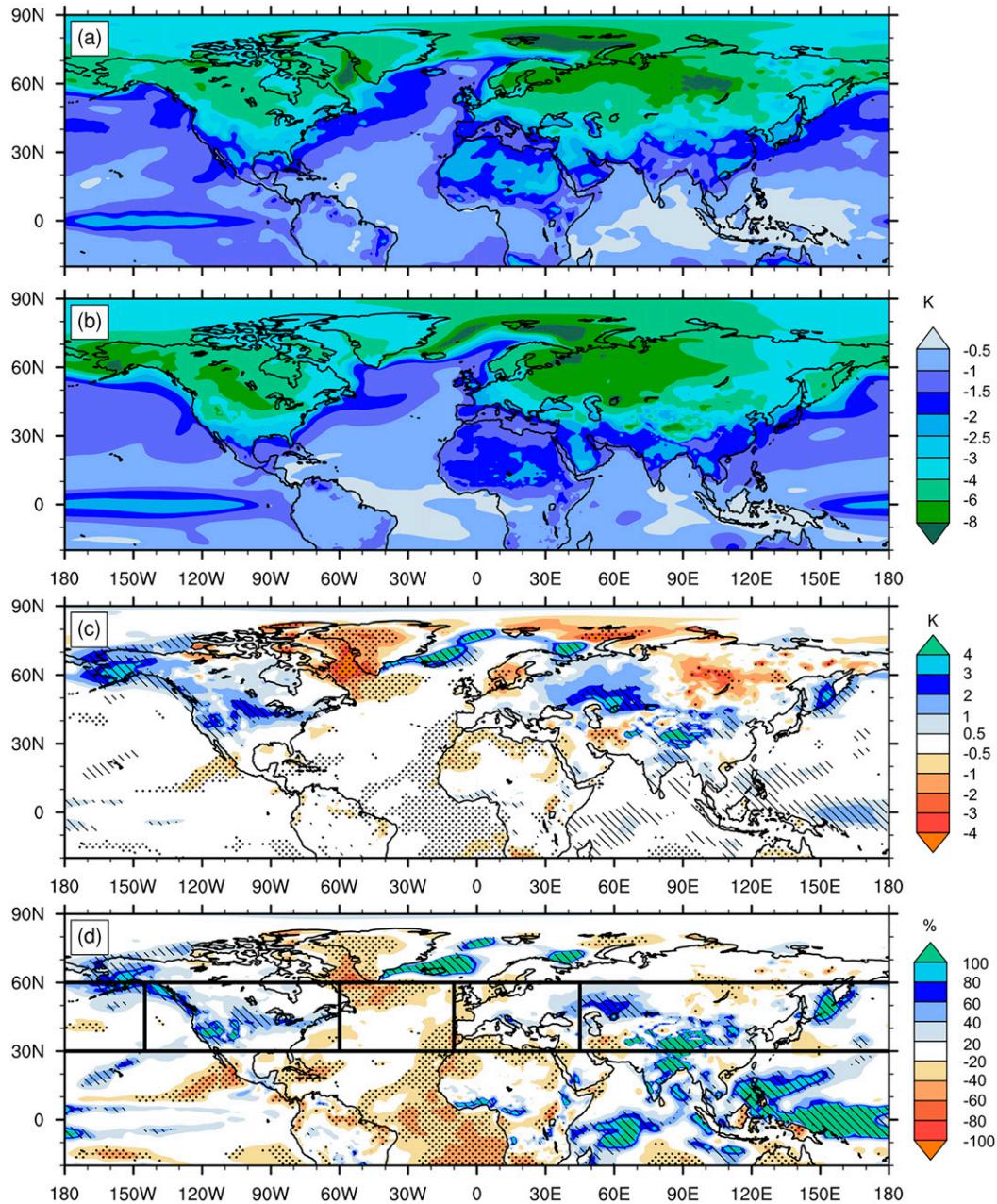


FIG. 8. The 40-yr return levels for (a) negative DJF  $T_{2m}$  anomalies in ERAI ( $RL^{ERAI}$ ) and (b) the ensemble average 40-yr return level for negative DJF  $T_{2m}$  anomalies in CSM-LENS ( $RL^{CESM}$ ). (c),(d) As in Figs. 7c and 7d, but for absolute values of (a) and (b) in this figure. Hatching in (c) and (d) denotes grid points where the absolute values of all 30 return levels from the individual CSM-LENS members exceed the absolute value of  $RL^{ERAI}$ , and stippling indicates where they are all below  $RL^{ERAI}$ . The black lines in (d) illustrate the extent of the three regions Europe, North America, and midlatitudes.

Substantial differences between  $RL^{ERAI}$  and  $RL^{CESM}$  in some regions are expected even if CSM were a perfect model, first because  $RL^{CESM}$  is the average over 30 return levels and second because there are large uncertainties in the estimation of the return levels for ERAI and each individual ensemble member. Nevertheless, widespread areas of under-/overestimated return levels in CSM (Figs. 7 and 8) point to model biases in the

variability of seasonal  $T_{2m}$  over numerous ocean regions, India, and northern Siberia. The best agreement in the spatial patterns as well as absolute values of  $RL^{ERAI}$  and  $RL^{CESM}$  is found over the midlatitude landmasses ( $30^{\circ}$ – $60^{\circ}$ N).

We therefore next focus on extreme season objects with a  $COM_L$  in the midlatitudes and examine their size and intensity in both datasets. Figure 9 shows scatterplots of  $A_L$  versus  $I_L$  for

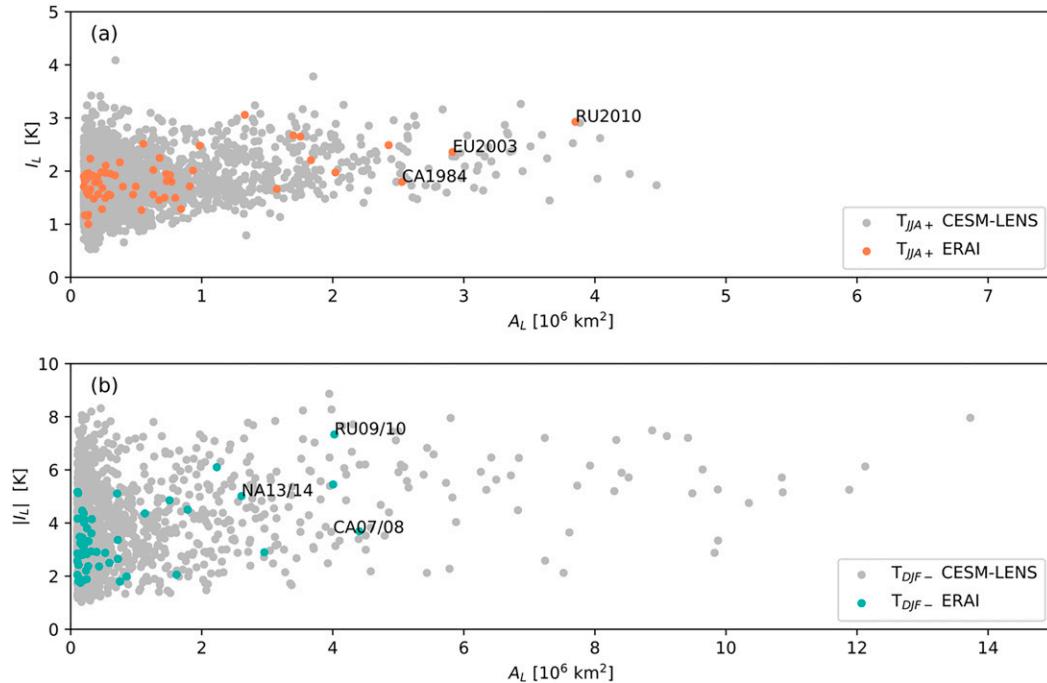


FIG. 9. (a) Scatterplot of  $A_L$  vs  $I_L$  for  $T_{JJA+}$  objects with a  $COM_L$  between  $30^\circ$  and  $60^\circ\text{N}$  in ERAI (orange) and CESM-LENS (gray). (b) As in (a), but for  $T_{DJF-}$  objects (ERAI objects are indicated with turquoise dots) with absolute values of  $I_L$  on the y axis. Selected ERAI events from Fig. 5 are highlighted with labels indicating the respective region and year. Note the different axes in the two panels.

midlatitude  $T_{JJA+}$  and  $T_{DJF-}$  objects from both datasets. In this  $A_L$ - $I_L$  space, the ERAI objects fall within the point clouds of the respective CESM-LENS objects for both types of seasons. There is a large variability of  $A_L$  and  $I_L$  in both types of seasons and for both characteristics, but numerous CESM-LENS objects with comparable characteristics to the largest/most intense ERAI objects are found. A more quantitative evaluation of the CESM-LENS  $T_{JJA+}$  and  $T_{DJF-}$  objects is provided in Fig. 10, which shows quantile–quantile (Q–Q) plots of  $A_L$  and  $I_L$  for both types of seasons. For the characteristics of  $T_{JJA+}$  objects there is little difference even for the top quantiles, and all ERAI  $A_L$  and most  $I_L$  quantiles remain within the range expected from sampling uncertainty (Figs. 10a,b). For  $T_{DJF-}$  objects, Figs. 10c and 10d reveals that CESM slightly overestimates  $I_L$ , but almost all  $A_L$  quantiles remain within the sampling uncertainty range.

This section reveals substantial but geographically confined model biases with regard to the 40-yr return levels of both JJA and DJF seasonal mean  $T_{2m}$ . However, the model evaluation also unravels a consistent geographical pattern of the 40-yr return levels across datasets, and, moreover, identifies remarkably good agreement between CESM-LENS and ERAI for some midlatitude extreme season characteristics.

## 6. Starting points for climatological extreme season analyses

In this section we illustrate possible starting points for climatological extreme season analyses that serve to illustrate the potential of our approach. Common to all three of these

analyses is that they are using CESM-LENS data to examine climatological characteristics of seasonal extremes that cannot be studied with reanalysis data alone, due to too few observed events. Consequently, the climatological results presented below are valid first and foremost for the CESM climate, and transferring these results to reality requires trusting the model outside of the observable. Section 5 has revealed reasonable agreement between simulated and observed  $A_L$  and  $I_L$  characteristics for midlatitude  $T_{JJA+}$  and  $T_{DJF-}$  objects. Moreover, as discussed above, previous studies have revealed good model performance for other aspects of synoptic to seasonal-scale dynamical processes that are relevant to extreme seasons. They, together with the results shown in Figs. 7–10, provide evidence that drawing inferences about climatological extreme season characteristics from CESM-LENS data is meaningful.

### a. Regional-scale climatological analyses of physical extreme season characteristics

The identification of spatial extreme season objects with specific characteristics makes it possible to assemble large sets of analogs (with regard to size, intensity, location, etc.) to specific observed events. They enable assessing to what extent key physical characteristics of the observed event are robust characteristics of a larger set of comparable (CESM-LENS) events.

For example, Fig. 11 illustrates a set of 15 CESM-LENS analogs to the 2013/14  $T_{DJF-}$  extreme season in North America (NA2013/14 event). These analogs are the 15 North American  $T_{DJF-}$  objects in CESM-LENS whose  $COM_L$  is closest to

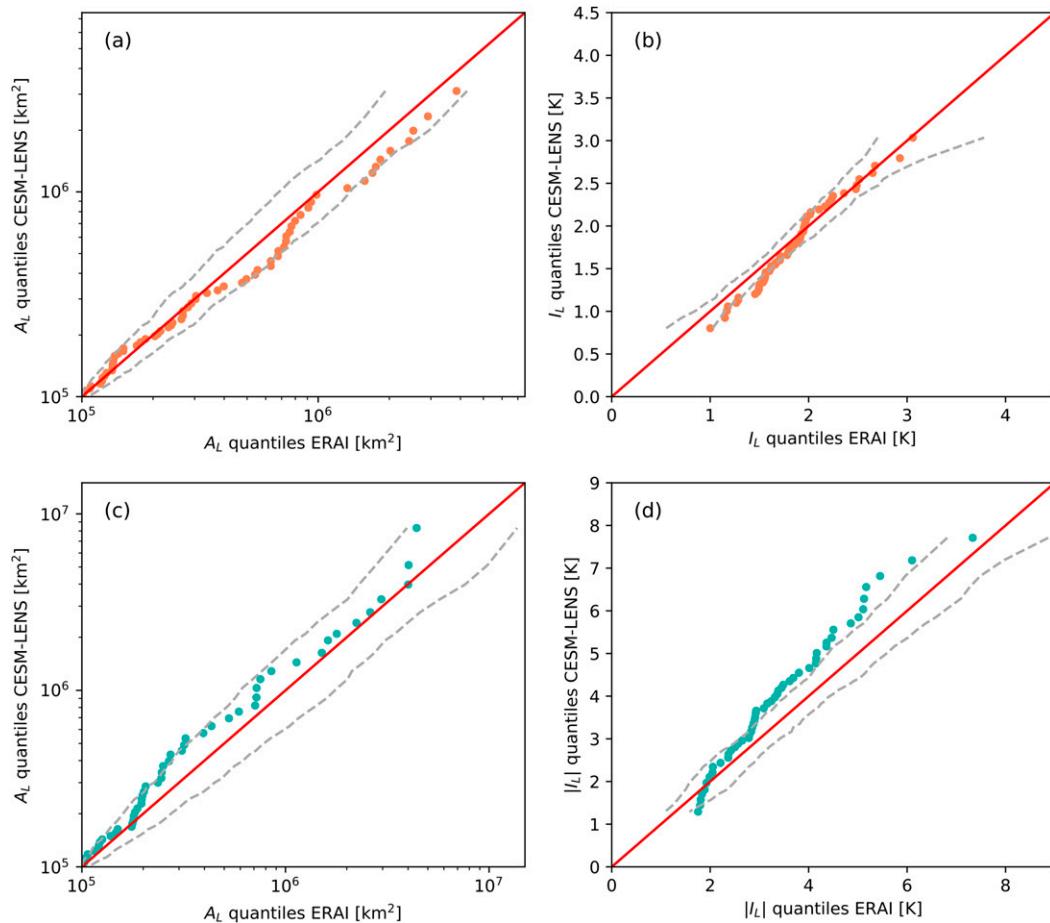


FIG. 10. (a),(b) Q–Q plots for  $A_L$  and  $|L_L|$  of ERAI vs CESM-LENS  $T_{JJA+}$  objects with a  $COM_L$  between  $30^\circ$  and  $60^\circ\text{N}$ . (c) As in (a), but for  $T_{DJF-}$  objects, and (d) Q–Q plot for  $|L_L|$  of midlatitude  $T_{DJF-}$  objects. Dashed lines indicate the 90% confidence bands within which the dots are expected under the null hypothesis that the respective ERAI and CESM-LENS extreme season objects are samples from the same distribution of extreme season objects. The confidence bands have been constructed by randomly drawing samples of CESM-LENS objects with a sample size corresponding to the number of ERAI objects. This procedure has been repeated 1000 times, which yields the respective null distribution for each quantile. Note that there are logarithmic axes in (a) and (c) and linear axes in (b) and (d).

that of the NA2013/14 object and which have an  $A_L$  between  $2 \times 10^6$  and  $4 \times 10^6 \text{ km}^2$ . Hartmann (2015) related the cold over North America in DJF 2013/14 to a particular SST anomaly pattern in the extratropical North Pacific, which projects strongly onto the North Pacific mode (NPM) of Deser and Blackmon (1995). It is therefore interesting to examine North Pacific SST anomalies also for the identified CESM-LENS analogs. Some of them indeed feature SST anomaly patterns similar to the ones observed in DJF 2013/14, for example, the  $T_{DJF-}$  event in Fig. 11o (Figs. 12a,b), that is, a positive anomaly in the northern North Pacific with a peak in the Gulf of Alaska, negative anomalies off the coast of Japan, and strongly positive anomalies in the Arctic Ocean (cf. Figs. 12a,b). For the event shown in Fig. 11p, however, the SST anomaly pattern in the North Pacific and the Arctic Ocean is almost reversed compared to DJF 2013/14, suggesting that the pattern identified by Hartmann (2015) is not a necessary

condition for extreme North American winters comparable to the NA2013/14 event. Nevertheless, when compositing the SST anomalies of all 15 analogs in Fig. 11, the resulting pattern does resemble the NPM [compare Figs. 12d and 1c in Hartmann (2015)], which suggests that NPM-like SST anomalies are common to extremely cold North American winters in CESM-LENS that are comparable to the NA2013/14 event.

#### b. Regional-scale analyses of statistical extreme season characteristics

Besides process-oriented analyses, the large set of CESM-LENS extreme season objects also enables statistical analyses with regard to individual observed events, for example, an estimation of regional return periods (RRPs). Previous studies have estimated return periods of seasonal extremes (e.g., Wolter et al. 2015 for the NA2013/14 event), but such analyses typically focused on the specific location of the respective

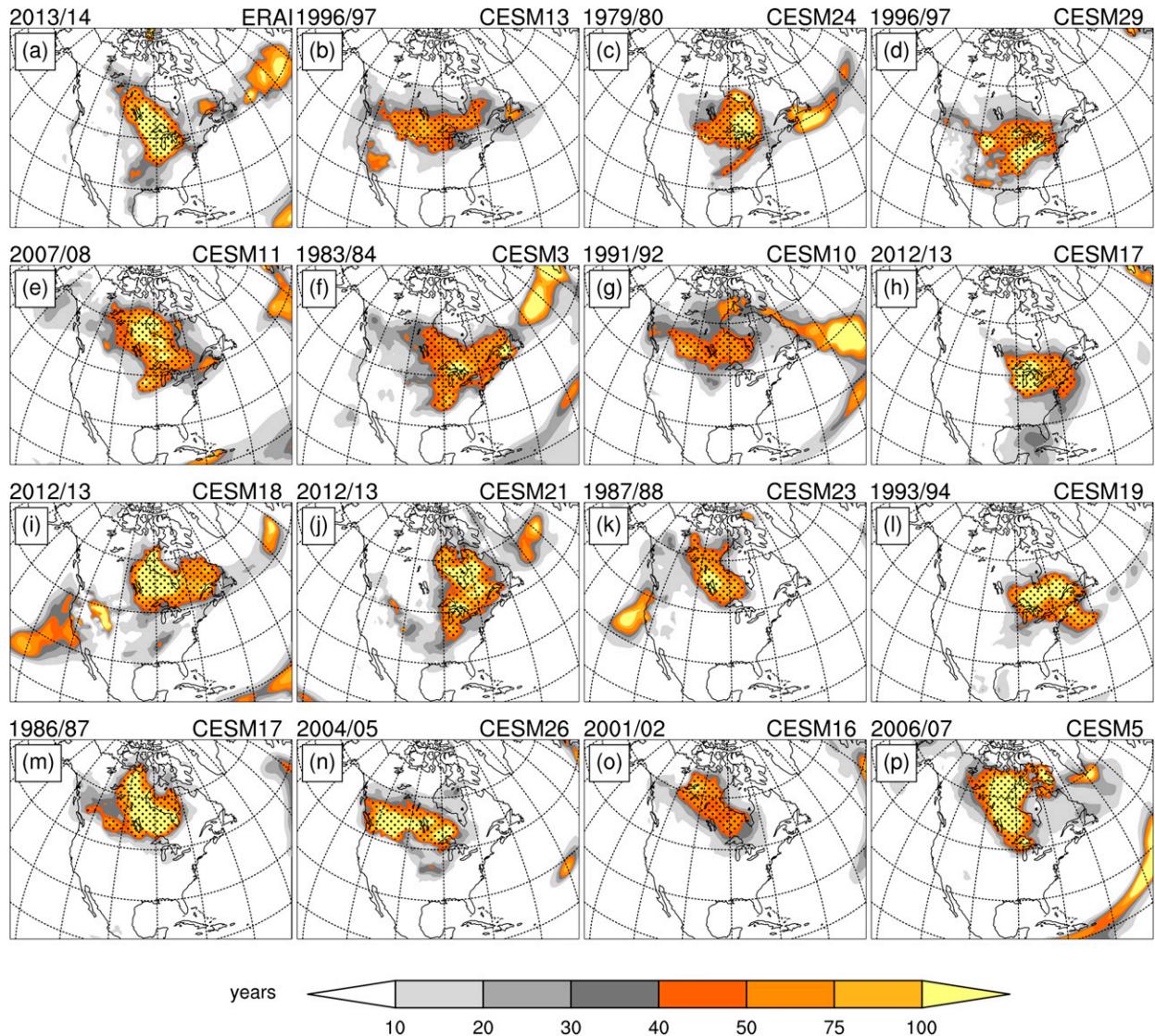


FIG. 11. North American CEM-LENS analogs to the NA2013/14  $T_{DJF-}$  extreme season. All panels depict  $LRP_p(T_v^M)$  for the respective dataset and year. (a) The observed NA2013/14  $T_{DJF-}$  extreme season; (b)–(p) the 15 CEM-LENS analogs to the NA2013/14 event (see text for details of the selection criteria). The respective year and CEM-LENS member are indicated above each panel. For reference, gridlines are drawn every  $15^\circ$  longitude and latitude.

event. However, in particular for decision makers acting at a continental or even global level, knowledge about the return period of such events anywhere within a larger region (e.g., North America, the NH midlatitudes) might be of interest too—if not even more relevant. Assuming that the statistics of CEM-LENS extreme season characteristics are accurate, such RRP can be estimated by counting how often a CEM-LENS extreme season object exceeds one or several characteristics of a specific observed event in ERAI event within a predefined geographical region.

Figure 13 shows illustrative examples of the computation of RRP for the NA2013/14 event for three regions: North America, Europe, and the entire NH midlatitudes (these regions are indicated in Fig. 8d). Within the 1200 years of CEM-LENS

simulations there are 41  $T_{DJF-}$  objects with a  $COM_L$  in North America for which  $A_L$  exceeds that of the NA2013/14 event, which yields an RRP of about 29 years for a  $T_{DJF-}$  object as large as the NA2013/14 event in North America. When considering both  $A_L$  and  $I_L$ , we find 26 North American  $T_{DJF-}$  objects in CEM-LENS that exceed the NA2013/14 event with regard to both characteristics, yielding an RRP of approximately 46 years. In Europe, events with an  $A_L$  at least as large as that of the NA2013/14 event occurred 35 times in CEM-LENS (RRP of 34 years); however, the intensity of the NA2013/14 event was highly unusual by European standards, and a European event with larger  $A_L$  and  $I_L$  than the NA2013/14 event only occurs 13 times in CEM-LENS, that is, with an RRP of 92 years. Finally,

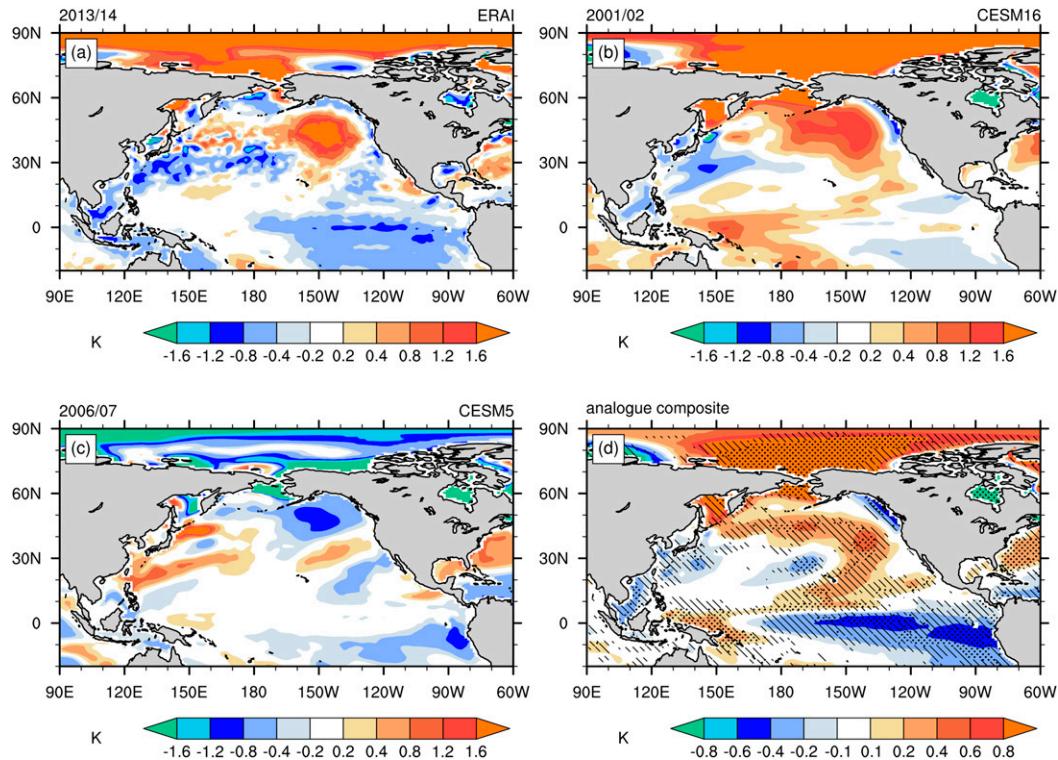


FIG. 12. SST anomalies for (a) the ERAI 2013/14 winter, (b),(c) the CESM-LENS winters of the  $T_{DJF-}$  events shown in Figs. 11o and 11p, and (d) the composite SST anomaly of all 15 analogs shown in Fig. 11. Note the different color scale in (d) compared to (a)–(c). Hatching (stippling) in (d) indicates grid points for which at least in 10 (in 12) of the analogs the SST anomalies have the same sign.

considering CESM-LENS  $T_{DJF-}$  objects anywhere in the midlatitudes we find numerous events that top the NA2013/14 event with regard to only  $A_L$  (135 events) and both characteristics (74 events), yielding RRP of 9 and 16 years, respectively. By construction, RRP are lower than local return periods. Consequently, focusing on local return periods only does not adequately describe how rare a particular observed event is in the regional-scale context relevant to many supra-national stakeholders.

Furthermore, Table 3 lists the midlatitude RRP with regard to  $A_L$  for the  $T_{JJA+}$  and  $T_{DJF-}$  objects presented in Fig. 5 for various values of  $\tau$ . For some events (in particular  $T_{DJF-}$  events) the RRP are remarkably insensitive to variations of  $\tau$ , while for others, RRP estimates change by almost an order of magnitude when varying  $\tau$  between 20 and 100 years. Clearly, the RRP thus depend on the exact definition of the events. Moreover, considering data from different climate models could further reconcile the numbers presented in Table 3.

### c. Contrasting extreme season characteristics across regions and seasons

Finally, the large sets of  $T_{JJA+}$  and  $T_{DJF-}$  events in CESM-LENS allow us to contrast characteristics of extreme seasons across regions and seasons. Such an analysis is presented in Figs. 14a,b, which shows for each grid point the median area  $A$  of all CESM-LENS extreme season objects covering the

respective grid point. Note that Fig. 14 considers the area  $A$  of all objects and not their land area ( $A_L$ ), in order to ensure that the results in Fig. 14 are not affected by the land–sea geometry. There are considerable regional variations in the median  $A$ , with largest objects over the tropical oceans and smallest objects over the Himalayan region, parts of Africa, and Mexico. Over extratropical land areas, the largest  $T_{JJA+}$  objects are found in northern Canada, Europe, and Russia, while  $T_{DJF-}$  objects are particularly large over central Canada and Russia. Interestingly, over the extratropical ocean regions, CESM-LENS  $T_{DJF-}$  and  $T_{JJA+}$  objects are of roughly the same size (Fig. 14c). However, over extratropical land areas  $T_{DJF-}$  objects are larger than  $T_{JJA+}$  objects by a factor of 2–4 over most regions. The main exceptions are areas of elevated topography in North America and Asia (Fig. 14c).

The physical causes for these regional and seasonal differences in the size of extreme season objects are hitherto unknown, but, given that the size of these events is undoubtedly a pivotal factor determining their overall impact, they should be further explored and understood.

## 7. Discussion

Our extreme season identification scheme is conceptually similar to existing heat-wave identification schemes (Stefanon et al. 2012; Russo et al. 2014; Vogel et al. 2020).

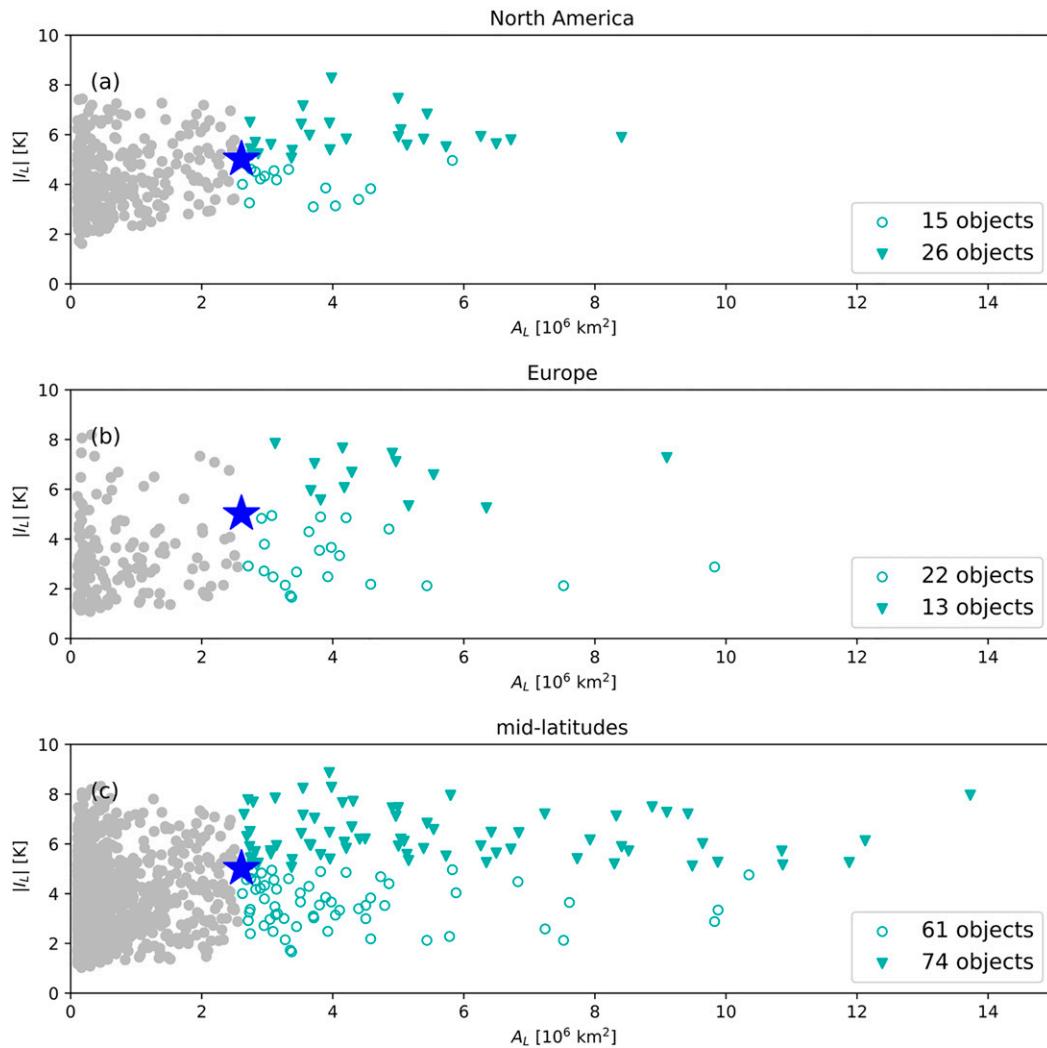


FIG. 13.  $A_L$ – $|I_L|$  space for  $T_{DJF-}$  objects in CESM-LENS for three different regions: (a) North America, (b) Europe, and (c) the midlatitudes ( $30^{\circ}$ – $60^{\circ}$ N). The extent of these regions is indicated in Fig. 8d. The blue star depicts the NA2013/14 event (see text for details). Turquoise open circles show CESM-LENS  $T_{DJF-}$  objects for which the  $A_L$  exceeds the  $A_L$  of the NA2013/14 event, but  $|I_L|$  is below that of the NA2013/14 event. Filled triangles depict CESM-LENS  $T_{DJF-}$  objects for which both  $A_L$  and  $|I_L|$  exceed the respective value of the NA2013/14 event.

However, these schemes use daily exceedances of percentile thresholds to identify hot days from which spatial heat-wave objects are constructed. Evidently, such a percentile-based approach is not applicable when identifying seasonal extreme events with LRPs of several decades in ERAI, and thus we base our identification scheme on statistical modeling of seasonal-mean temperatures.

The scheme involves at least three nontrivial decisions, namely, choosing an appropriate statistical model, selecting an LRP threshold to identify the locally extreme seasonal means, and deciding for characteristics of the extreme season objects that are subsequently analyzed. Each of these decisions should be guided by the specific research question in mind, but still various options may often be sensible. We have shown that the  $N^{YJ}$  distribution fits the JJA and DJF  $T_{2m}$  anomalies, but in

particular for strongly skewed variables (e.g., seasonal mean precipitation) more flexible parametric probability distributions such as the gamma or skewed  $T$  distribution could be chosen (Hansen 1994; Theodossiou 1998; Wilks 2011). Furthermore, sophisticated spatial models for extremes (e.g., Cooley and Sain 2010) could further improve the estimation of LRPs. Also, the selection of  $\tau$  as well as the extreme season characteristics that are analyzed should depend on the type of impacts that one is interested in, and reasonable choices for  $\tau$  are also constrained by the record length of the available data. In this concept paper, we propose and illustrate pragmatic and easily implementable solutions for how to take these decisions when analyzing seasonal  $T_{2m}$  anomalies, but we emphasize that these are neither the only possible nor perfect solutions.

TABLE 3. The top six  $T_{\text{JJA}+}$  and  $T_{\text{DJF}-}$  extreme seasons in ERAI according to  $A_L$  in the midlatitudes (i.e., with a  $\text{COM}_L$  between  $30^\circ$  and  $60^\circ\text{N}$ ). The five rightmost columns list the RRP of an object with the same land area anywhere in the midlatitudes ( $30^\circ$ – $60^\circ\text{N}$ ) in CESM-LENS for various  $\tau$ .

	Year	Region/country	RRP $_{\tau=100}$	RRP $_{\tau=75}$	RRP $_{\tau=40}$	RRP $_{\tau=30}$	RRP $_{\tau=20}$
$T_{\text{JJA}+}$	2010	Western Russia	240	400	240	100	1200
	2003	Europe	56	44	39	29	17
	1984	Central Asia	16	32	24	16	9
	1998	Kazakhstan	27	29	21	15	17
	2002	Mongolia/China	8	8	13	12	10
	1983	United States/Canada	2	1	10	11	9
$T_{\text{DJF}-}$	2007/08	Central Asia	11	11	20	19	27
	2009/10	Russia	16	16	18	24	21
	2000/01	Siberia/China	3	4	18	22	21
	2009/10	Southern United States/Mexico	12	12	11	8	10
	2013/14	Central United States	13	13	9	8	9
	1981/82	Canada/Alaska	15	13	7	5	4

To enable climatological extreme season analyses we propose to spatially pool individual seasonal extreme events. Similar spatial pooling strategies for rare weather and climate events have been employed frequently and successfully in the past, for instance, for identifying robust trends in their number, duration, and intensity (Fischer et al. 2013; Coumou et al. 2013; Fischer and Knutti 2014; Pfeleiderer et al. 2019). In these studies, the motivation for regional pooling of events was that the climate change signal is often masked at a local scale by internal variability and emerges robustly only at a regional scale (Fischer and Knutti 2014). Here, the purpose of the pooling is slightly different, as we aim to enable process-oriented climatological analyses, which require the analyzed extreme seasons to be comparable with regard to their underlying physical processes. In this study we have not thoroughly assessed these processes, and it thus remains to be examined how similar or dissimilar they are for the identified seasonal extreme events. Nevertheless, the identification of events with comparable spatial, temporal, and intensity characteristics performed here is a necessary starting point for such analyses.

Furthermore, the analyses presented here are limited by a number of factors, which are rather fundamental to the analysis of rare events. First, identifying extreme seasons with LRPs of several decades in reanalysis datasets is inherently uncertain with any statistical model and, second, suffers from a general underestimation of large LRPs in small samples. Third, the climatological extreme season analyses proposed in section 6 rely on the statistics and physics of extreme seasons in climate model data, and transferring results for simulated extreme seasons to reality requires trusting the model. At the same time, model evaluation with regard to seasonal extremes is difficult, because the observed statistics of extreme season characteristics become more robust only at large spatial scales (i.e., the entire midlatitudes). All three of these caveats can be partly addressed when working with longer reanalyses such as the new ERA5 dataset (Hersbach et al. 2020). Finally, the identification of coherent spatial extreme season objects is ambiguous in situations where multiple proximate patches of

large LRPs occur during a particular season. In such situations, varying  $\tau$  often leads to the breakup or merger of objects. This problem is inherent to any object identification in weather and climate data and has previously been referred to as ‘‘camel effect’’ (Wernli et al. 2008). We argue that it is best treated by exploring the sensitivity of any results to variations in  $\tau$ .

## 8. Summary and conclusions

In this study, we highlight the current lack of climatological studies on the dynamics and physics of seasonal extremes and propose a pragmatic approach for spatial pooling of seasonal extreme events. This approach enables climatological analyses of seasonal extremes at a regional scale. The centerpiece of this approach is a generic scheme that identifies spatially coherent extreme season objects in reanalysis and climate model data. Here we develop the identification scheme for JJA and DJF  $T_{2m}$  extreme seasons, but its principles are transferable to any variable of interest.

The scheme consists of three steps: 1) fitting a statistical model to seasonal mean values at each grid point, 2) selecting a local return period (LRP) threshold above which seasonal mean values are deemed extreme, and 3) forming spatially coherent extreme season objects, whose characteristics (area, intensity, etc.) can then be computed. We apply the scheme to Northern Hemisphere JJA and DJF  $T_{2m}$  anomalies from ERAI and the 30-member CESM-LENS dataset, whereby the scheme is applied to each member separately. A Yeo–Johnson power-transformed normal distribution is fitted to the data at each grid point, which allows for contradiction-free modeling of  $T_{2m}$  seasonal means over almost all land areas. Spatially coherent extreme season objects are identified based on an LRP threshold of 40 years. In the Northern Hemisphere and the four decades of ERAI data we identify 175  $T_{\text{JJA}+}$  and 147  $T_{\text{DJF}-}$  extreme season objects, while the 1200-yr CESM-LENS dataset yields 5285  $T_{\text{JJA}+}$  and 3556  $T_{\text{DJF}-}$  objects.

The application of this identification scheme to a short reanalysis dataset is not without challenges, because it involves

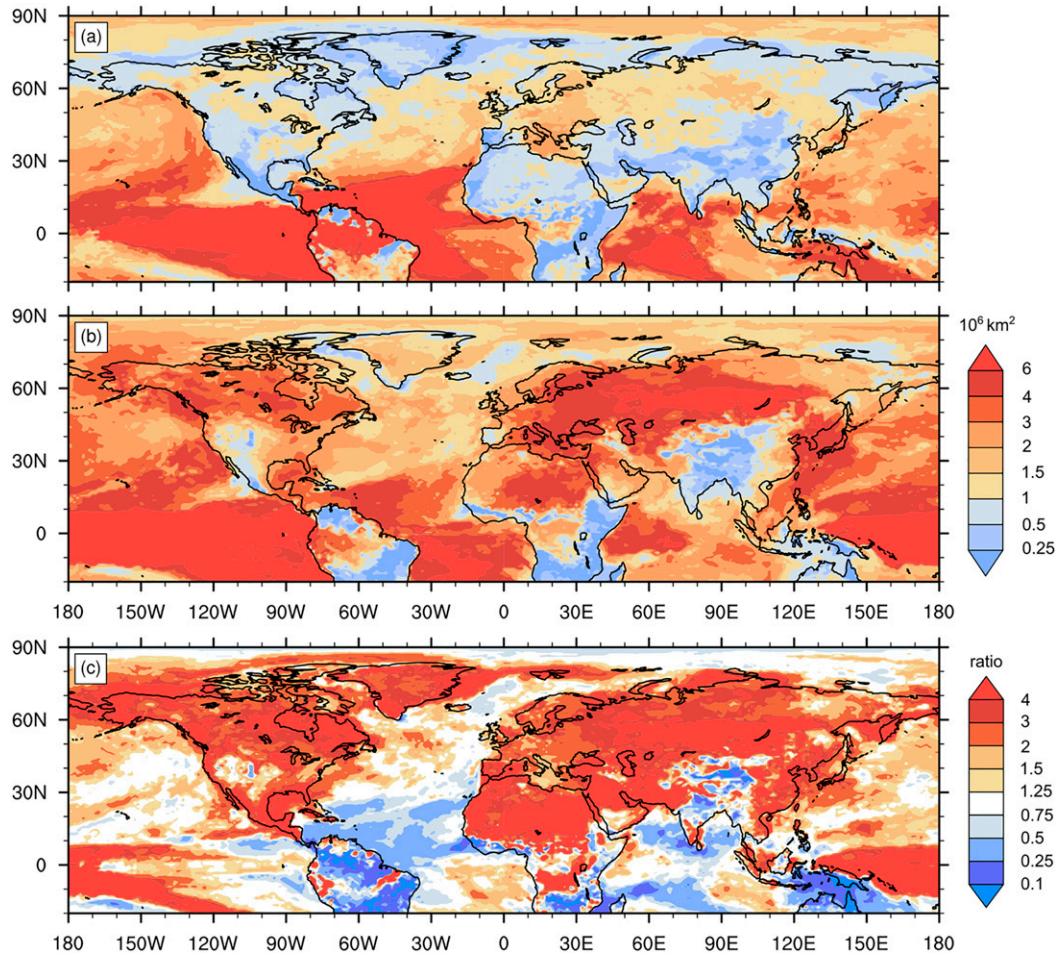


FIG. 14. (a) Median area  $A$  of all  $T_{JJA+}$  objects covering the respective grid point; (b) as in (a), but for  $T_{DJF-}$  objects; (c) the ratio of the quantities shown in (b) and (a), i.e., the field in (b) divided by the field in (a). Note that this figure displays quantities derived from the area  $A$ , while Figs. 8–9, 11, and 13 show quantities related to the land area  $A_L$  of extreme season objects.

estimating large LRP in small samples, which is inherently uncertain with any statistical model. Using CESM-LENS data, we show that in a dataset of 40 years JJA  $T_{2m}$  extremes can be identified with a probability of detection (POD) of 0.68 and a false alarm rate of 0.27 with our statistical model and an LRP threshold of 40 years. Hereby, the LRP of unusually rare values in small samples is on average underestimated because of outlier effects on the estimated distribution parameters (Stigler 1997; Potts 2012). This general underestimation of large LRPs in small samples leads to numerous “near misses” and partly explains the limited POD.

Despite these challenges, we identify the largest ERAI  $T_{JJA+}$  and  $T_{DJF-}$  extreme season objects in the midlatitudes and find numerous well-known high-impact events among them, for example, the summers 2010 and 2003 in western Russia and Europe, respectively, or the cold winters in the United States in 2009/10 and 2013/14. Furthermore, five out of the seven largest ERAI midlatitude  $T_{JJA+}$  events coincided with periods of El Niño to La Niña transitions. This illustrates the value of spatially pooling events, as, despite their occurrence

on three different continents, these major extreme summer events share hitherto unknown physical commonalities.

The proposed approach offers vast opportunities for climatological extreme season analyses, in particular when considering both simulated and observed extreme season objects. However, using model data in such ways also requires thorough model evaluation. We therefore first compare ERAI and CESM-LENS extreme season characteristics and find substantial but geographically confined model biases in 40-yr return levels for JJA and DJF  $T_{2m}$  anomalies. At the same time, the size of midlatitude  $T_{JJA+}$  and  $T_{DJF-}$  objects are remarkably well captured by the model. Our feature-based approach allows us to compile large samples of simulated analogs (e.g., with regard to location, size, and intensity) to even the most extreme ERAI events, which can serve as a starting point for climatological analyses of simulated extreme seasons with characteristics comparable to those of observed high-impact events. Furthermore, the large set of CESM-LENS  $T_{JJA+}$  and  $T_{DJF-}$  extreme season objects can be exploited to estimate regional return periods (RRPs) of seasonal extremes with

particular characteristics. For the North American 2013/14  $T_{DJF-}$  event, for example, CESM-LENS data suggests that an event with at least the same area and intensity occurs roughly every 46 years in North America, while in Europe such an event is expected only every 92 years. By construction, such RRP are distinctly lower than the strictly local return periods that have commonly been reported by previous studies, yet in particular for decision makers acting at continental to global scales (e.g., global reinsurers, the European Union, or United Nations organizations), such RRP might be more relevant than local return periods and should thus be quantified and communicated. Finally, the CESM-LENS  $T_{JJA+}$  and  $T_{DJF-}$  extreme season objects also makes it possible to identify and study regional and seasonal variations in extreme season characteristics.

We conclude that already existing and freely available reanalyses and large-ensemble climate model simulations (Deser et al. 2020; Maher et al. 2021) yield immense potential for studying climatological aspects of seasonal extremes. This study now provides tools for better evaluating climate models with regard to extreme season characteristics and makes it possible to address these research questions. The methods developed here are deliberately generic so that they can easily be adapted to other variables and datasets and expanded with more sophisticated statistical tools. It is therefore hoped that this study will trigger further and diverse research on climatological aspects of extreme seasons.

**Acknowledgments.** MR, MH, and HW acknowledge funding of the INTEXseas project from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement 787652). FL acknowledges the Swiss National Science Foundation (Grant 174128) and the Regional and Global Model Analysis (RGMA) component of the Earth and Environmental System Modeling Program of the U.S. DOE's Office of BER (via NSF IA 1947282). The authors declare no conflicts of interests. We thank the three anonymous reviewers for their critical and constructive comments.

**Data availability statement.** ERA-Interim data can be downloaded freely from the ECMWF webpage (<https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/>) and the CESM-LENS data is freely available via the Earth System Grid (<https://www.earthsystemgrid.org>). Derived data displayed in the figures is available from the corresponding author upon request.

## REFERENCES

- Barriopedro, D., E. Fischer, J. Luterbacher, R. Trigo, and G.-H. Ricardo, 2011: The hot summer of 2010: Redrawing the temperature record map of Europe. *Science*, **332**, 220–224, <https://doi.org/10.1126/science.1201224>.
- Bieli, M., S. Pfahl, and H. Wernli, 2015: A Lagrangian investigation of hot and cold temperature extremes in Europe. *Quart. J. Roy. Meteor. Soc.*, **141**, 98–108, <https://doi.org/10.1002/qj.2339>.
- Black, E., M. Blackburn, R. G. Harrison, B. J. Hoskins, and J. Methven, 2004: Factors contributing to the summer 2003 European heatwave. *Weather*, **59**, 217–223, <https://doi.org/10.1256/wea.74.04>.
- Buras, A., A. Rammig, and C. S. Zang, 2020: Quantifying impacts of the 2018 drought on European ecosystems in comparison to 2003. *Biogeosciences*, **17**, 1655–1672, <https://doi.org/10.5194/bg-17-1655-2020>.
- Cattiaux, J., and A. Ribes, 2018: Defining single extreme weather events in a climate perspective. *Bull. Amer. Meteor. Soc.*, **99**, 1557–1568, <https://doi.org/10.1175/BAMS-D-17-0281.1>.
- , R. Vautard, C. Cassou, P. Yiou, V. Masson-Delmotte, and F. Codron, 2010: Winter 2010 in Europe: A cold extreme in a warming climate. *Geophys. Res. Lett.*, **37**, L20704, <https://doi.org/10.1029/2010GL044613>.
- Catto, J. L., and S. Pfahl, 2013: The importance of fronts for extreme precipitation. *J. Geophys. Res.*, **118**, 10 791–10 801, <https://doi.org/10.1002/jgrd.50852>.
- Chase, T. N., K. Wolter, R. A. Pielke, and I. Rasool, 2006: Was the 2003 European summer heat wave unusual in a global context? *Geophys. Res. Lett.*, **33**, L23709, <https://doi.org/10.1029/2006GL027470>.
- Ciais, P., and Coauthors, 2005: Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature*, **437**, 529–533, <https://doi.org/10.1038/nature03972>.
- Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer, 219 pp.
- Cooley, D., and S. R. Sain, 2010: Spatial hierarchical modeling of precipitation extremes from a regional climate model. *J. Agric. Biol. Environ. Stat.*, **15**, 381–402, <https://doi.org/10.1007/s13253-010-0023-9>.
- Coumou, D., A. Robinson, and S. Rahmstorf, 2013: Global increase in record-breaking monthly-mean temperatures. *Climatic Change*, **118**, 771–782, <https://doi.org/10.1007/s10584-012-0668-1>.
- Davies, H. C., 2015: Weather chains during the 2013/2014 winter and their significance for seasonal prediction. *Nat. Geosci.*, **8**, 833–837, <https://doi.org/10.1038/ngeo2561>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Deser, C., and M. L. Blackmon, 1995: On the relationship between tropical and North Pacific sea surface temperature variations. *J. Climate*, **8**, 1677–1680, [https://doi.org/10.1175/1520-0442\(1995\)008<1677:OTRBTA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<1677:OTRBTA>2.0.CO;2).
- , and Coauthors, 2020: Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Climate Change*, **10**, 277–286, <https://doi.org/10.1038/s41558-020-0731-2>.
- Díaz, J., R. García-Herrera, R. M. Trigo, C. Linares, M. A. Valente, J. M. De Miguel, and E. Hernández, 2006: The impact of the summer 2003 heat wave in Iberia: How should we measure it? *Int. J. Biometeor.*, **50**, 159–166, <https://doi.org/10.1007/s00484-005-0005-8>.
- Dole, R., and Coauthors, 2011: Was there a basis for anticipating the 2010 Russian heat wave? *Geophys. Res. Lett.*, **38**, L06702, <https://doi.org/10.1029/2010GL046582>.
- Fink, A. H., T. Brücher, A. Krüger, G. C. Leckebusch, J. G. Pinto, and U. Ulbrich, 2004: The 2003 European summer heatwaves and drought—Synoptic diagnosis and impacts. *Weather*, **59**, 209–216, <https://doi.org/10.1256/wea.73.04>.
- Fischer, E. M., and R. Knutti, 2014: Detection of spatially aggregated changes in temperature and precipitation extremes. *Geophys. Res. Lett.*, **41**, 547–554, <https://doi.org/10.1002/2013GL058499>.

- , S. I. Seneviratne, P. L. Vidale, D. Lüthi, and C. Schär, 2007: Soil moisture–atmosphere interactions during the 2003 European summer heat wave. *J. Climate*, **20**, 5081–5099, <https://doi.org/10.1175/JCLI4288.1>.
- , U. Beyerle, and R. Knutti, 2013: Robust spatially aggregated projections of climate extremes. *Nat. Climate Change*, **3**, 1033–1038, <https://doi.org/10.1038/nclimate2051>.
- Fisher, R. A., and L. H. C. Tippett, 1928: Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Math. Proc. Cambridge Philos. Soc.*, **24**, 180–190, <https://doi.org/10.1017/S0305004100015681>.
- Fouillet, A., and Coauthors, 2006: Excess mortality related to the August 2003 heat wave in France. *Int. Arch. Occup. Environ. Health*, **80**, 16–24, <https://doi.org/10.1007/s00420-006-0089-4>.
- Fragkoulidis, G., V. Wirth, P. Bossmann, and A. H. Fink, 2018: Linking Northern Hemisphere temperature extremes to Rossby wave packets. *Quart. J. Roy. Meteor. Soc.*, **144**, 553–566, <https://doi.org/10.1002/qj.3228>.
- Frei, C., and F. A. Isotta, 2019: Ensemble spatial precipitation analysis from rain gauge data: Methodology and application in the European Alps. *J. Geophys. Res.*, **124**, 5757–5778, <https://doi.org/10.1029/2018JD030004>.
- García-Herrera, R., J. Díaz, R. M. Trigo, J. Luterbacher, and E. M. Fischer, 2010: A review of the European summer heat wave of 2003. *Crit. Rev. Environ. Sci. Technol.*, **40**, 267–306, <https://doi.org/10.1080/10643380802238137>.
- Hansen, B. E., 1994: Autoregressive conditional density estimation. *Int. Econ. Rev.*, **35**, 705, <https://doi.org/10.2307/2527081>.
- Hartmann, D. L., 2015: Pacific sea surface temperature and the winter of 2014. *Geophys. Res. Lett.*, **42**, 1894–1902, <https://doi.org/10.1002/2015GL063083>.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Huguenin, M. F., E. M. Fischer, S. Kotlarski, S. C. Scherrer, C. Schwierz, and R. Knutti, 2020: Lack of change in the projected frequency and persistence of atmospheric circulation types over Central Europe. *Geophys. Res. Lett.*, **47**, e2019GL086132, <https://doi.org/10.1029/2019GL086132>.
- Huntingford, C., and Coauthors, 2014: Potential influences on the United Kingdom's floods of winter 2013/14. *Nat. Climate Change*, **4**, 769–777, <https://doi.org/10.1038/nclimate2314>.
- Hurrell, J. W., and Coauthors, 2013: The Community Earth System Model: A framework for collaborative research. *Bull. Amer. Meteor. Soc.*, **94**, 1339–1360, <https://doi.org/10.1175/BAMS-D-12-00121.1>.
- Jonkeren, O., P. Rietveld, J. van Ommeren, and A. te Linde, 2014: Climate change and economic consequences for inland waterway transport in Europe. *Reg. Environ. Change*, **14**, 953–965, <https://doi.org/10.1007/s10113-013-0441-7>.
- Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) Large Ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, <https://doi.org/10.1175/BAMS-D-13-00255.1>.
- Klotzbach, P. J., C. J. Schreck III, J. M. Collins, M. M. Bell, E. S. Blake, and D. Roache, 2018: The extremely active 2017 North Atlantic hurricane season. *Mon. Wea. Rev.*, **146**, 3425–3443, <https://doi.org/10.1175/MWR-D-18-0078.1>.
- Krikken, F., F. Lehner, K. Haustein, I. Drobyshev, and G. J. van Oldenborgh, 2021: Attribution of the role of climate change in the forest fires in Sweden 2018. *Nat. Hazards Earth Syst. Sci.*, **21**, 2169–2179, <https://doi.org/10.5194/nhess-21-2169-2021>.
- Lau, W. K. M., and K. M. Kim, 2012: The 2010 Pakistan flood and Russian heat wave: Teleconnection of hydrometeorological extremes. *J. Hydrometeorol.*, **13**, 392–403, <https://doi.org/10.1175/JHM-D-11-016.1>.
- Luterbacher, J., and C. Pfister, 2015: The year without a summer. *Nat. Geosci.*, **8**, 246–248, <https://doi.org/10.1038/ngeo2404>.
- Maher, N., S. Milinski, and R. Ludwig, 2021: Large ensemble climate model simulations: Introduction, overview, and future prospects for utilising multiple types of large ensemble. *Earth Syst. Dyn.*, **12**, 401–418, <https://doi.org/10.5194/esd-12-401-2021>.
- Martius, O., E. Zenklusen, C. Schwierz, and H. C. Davies, 2006: Episodes of alpine heavy precipitation with an overlying elongated stratospheric intrusion: A climatology. *Int. J. Climatol.*, **26**, 1149–1164, <https://doi.org/10.1002/joc.1295>.
- Matthews, T., C. Murphy, R. L. Wilby, and S. Harrigan, 2014: Stormiest winter on record for Ireland and UK. *Nat. Climate Change*, **4**, 738–740, <https://doi.org/10.1038/nclimate2336>.
- McDonald, J. B., and W. K. Newey, 1988: Partially adaptive estimation of regression models via the generalized T distribution. *Econom. Theor.*, **4**, 428–457, <https://doi.org/10.1017/S0266466600013384>.
- Moore, B. J., D. Keyser, and L. F. Bosart, 2019: Linkages between extreme precipitation events in the central and eastern United States and Rossby wave breaking. *Mon. Wea. Rev.*, **147**, 3327–3349, <https://doi.org/10.1175/MWR-D-19-0047.1>.
- , A. B. White, D. J. Gottas, and P. J. Neiman, 2020: Extreme precipitation events in Northern California during winter 2016–17: Multiscale analysis and climatological perspective. *Mon. Wea. Rev.*, **148**, 1049–1074, <https://doi.org/10.1175/MWR-D-19-0242.1>.
- Namias, J., 1978: Multiple causes of the North American abnormal winter 1976–77. *Mon. Wea. Rev.*, **106**, 279–295, [https://doi.org/10.1175/1520-0493\(1978\)106<0279:MCOTNA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1978)106<0279:MCOTNA>2.0.CO;2).
- NOAA, 2021: Niño 3.4 SST index. Accessed 3 May 2021, [https://psl.noaa.gov/gcos\\_wgsp/Timeseries/Nino34/](https://psl.noaa.gov/gcos_wgsp/Timeseries/Nino34/).
- Oliphant, T. E., 2007: Python for scientific computing. *Comput. Sci. Eng.*, **9**, 10–20, <https://doi.org/10.1109/MCSE.2007.58>.
- Perkins-Kirkpatrick, S. E., E. M. Fischer, O. Angéllil, and P. B. Gibson, 2017: The influence of internal climate variability on heatwave frequency trends. *Environ. Res. Lett.*, **12**, 044005, <https://doi.org/10.1088/1748-9326/aa63fe>.
- Pfleiderer, P., C. F. Schlessner, K. Kornhuber, and D. Coumou, 2019: Summer weather becomes more persistent in a 2°C world. *Nat. Climate Change*, **9**, 666–671, <https://doi.org/10.1038/s41558-019-0555-0>.
- Potts, J. M., 2012: Basic concepts. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, I. T. Jolliffe and D. B. Stephenson, Eds., John Wiley & Sons, Inc., 11–29.
- Raveh-Rubin, S., and H. Wernli, 2015: Large-scale wind and precipitation extremes in the Mediterranean: A climatological analysis for 1979–2012. *Quart. J. Roy. Meteor. Soc.*, **141**, 2404–2417, <https://doi.org/10.1002/qj.2531>.
- Robine, J. M., S. L. K. Cheung, S. Le Roy, H. Van Oyen, C. Griffiths, J. P. Michel, and F. R. Herrmann, 2008: Death toll exceeded 70,000 in Europe during the summer of 2003. *C. R. Biol.*, **331**, 171–178, <https://doi.org/10.1016/j.crv.2007.12.001>.
- Rössler, O., P. Froidevaux, U. Börst, R. Rickli, O. Martius, and R. Weingartner, 2014: Retrospective analysis of a non-forecasted rain-on-snow flood in the Alps—A matter of model limitations or unpredictable nature? *Hydrol. Earth Syst. Sci.*, **18**, 2265–2285, <https://doi.org/10.5194/hess-18-2265-2014>.
- Röthlisberger, M., S. Pfahl, and O. Martius, 2016: Regional-scale jet waviness modulates the occurrence of midlatitude weather

- extremes. *Geophys. Res. Lett.*, **43**, 10 989–10 997, <https://doi.org/10.1002/2016GL070944>.
- , M. Sprenger, E. Flaounas, U. Beyerle, and H. Wernli, 2020: The substructure of extremely hot summers in the Northern Hemisphere. *Wea. Climate Dyn.*, **1**, 45–62, <https://doi.org/10.5194/wcd-1-45-2020>.
- Russo, S., and Coauthors, 2014: Magnitude of extreme heat waves in present climate and their projection in a warming world. *J. Geophys. Res. Atmos.*, **119**, 12 500–12 512, <https://doi.org/10.1002/2014JD022098>.
- Schaller, N., J. Sillmann, J. Anstey, E. M. Fischer, C. M. Grams, and S. Russo, 2018: Influence of blocking on northern European and western Russian heatwaves in large climate model ensembles. *Environ. Res. Lett.*, **13**, 054015, <https://doi.org/10.1088/1748-9326/aaba55>.
- Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller, 2004: The role of increasing temperature variability in European summer heatwaves. *Nature*, **427**, 332–336, <https://doi.org/10.1038/nature02300>.
- Schneiderer, A., S. Schubert, P. Vargin, F. Lunkeit, X. Zhu, D. H. W. Peters, and K. Fraedrich, 2012: Large-scale flow and the long-lasting blocking high over Russia: Summer 2010. *Mon. Wea. Rev.*, **140**, 2967–2981, <https://doi.org/10.1175/MWR-D-11-00249.1>.
- Sedláček, J., O. Martius, and R. Knutti, 2011: Influence of subtropical and polar sea-surface temperature anomalies on temperatures in Eurasia. *Geophys. Res. Lett.*, **38**, L12803, <https://doi.org/10.1029/2011GL047764>.
- Seo, D. J., R. Siddique, Y. Zhang, and D. Kim, 2014: Improving real-time estimation of heavy-to-extreme precipitation using rain gauge data via conditional bias-penalized optimal estimation. *J. Hydrol.*, **519**, 1824–1835, <https://doi.org/10.1016/j.jhydrol.2014.09.055>.
- Shepherd, T. G., and Coauthors, 2018: Storylines: An alternative approach to representing uncertainty in physical aspects of climate change. *Climatic Change*, **151**, 555–571, <https://doi.org/10.1007/s10584-018-2317-9>.
- Sprenger, M., and Coauthors, 2017: Global climatologies of Eulerian and Lagrangian flow features based on ERA-Interim. *Bull. Amer. Meteor. Soc.*, **98**, 1739–1748, <https://doi.org/10.1175/BAMS-D-15-00299.1>.
- Stefanon, M., F. Dandrea, and P. Drobinski, 2012: Heatwave classification over Europe and the Mediterranean region. *Environ. Res. Lett.*, **7**, 014023, <https://doi.org/10.1088/1748-9326/7/1/014023>.
- Stigler, S. M., 1997: Regression towards the mean, historically considered. *Stat. Methods Med. Res.*, **6**, 103–114, <https://doi.org/10.1177/096228029700600202>.
- Theodossiou, P., 1998: Financial data and the skewed generalized T distribution. *Manage. Sci.*, **44**, 1650–1661, <https://doi.org/10.1287/mnsc.44.12.1650>.
- Trenberth, K. E., and J. T. Fasullo, 2012: Climate extremes and climate change: The Russian heat wave and other climate extremes of 2010. *J. Geophys. Res.*, **117**, D17103, <https://doi.org/10.1029/2012JD018020>.
- , —, and T. G. Shepherd, 2015: Attribution of climate extreme events. *Nat. Climate Change*, **5**, 725–730, <https://doi.org/10.1038/nclimate2657>.
- Vogel, M. M., J. Zscheischler, E. M. Fischer, and S. I. Seneviratne, 2020: Development of future heatwaves for different hazard thresholds. *J. Geophys. Res. Atmos.*, **125**, e2019JD032070, <https://doi.org/10.1029/2019JD032070>.
- Wehner, M., D. Stone, H. Krishnan, K. AchutaRao, and F. Castillo, 2016: The deadly combination of heat and humidity in India and Pakistan in summer 2015 [in “Explaining Extreme Events of 2015 from a Climate Perspective”]. *Bull. Amer. Meteor. Soc.*, **97** (12), S81–S86, <https://doi.org/10.1175/BAMS-D-16-0145.1>.
- Wehrli, K., B. P. Guillod, M. Hauser, M. Leclair, and S. I. Seneviratne, 2019: Identifying key driving processes of major recent heat waves. *J. Geophys. Res. Atmos.*, **124**, 11 746–11 765, <https://doi.org/10.1029/2019JD030635>.
- Wernli, H., M. Paulat, M. Hagen, and C. Frei, 2008: SAL—A novel quality measure for the verification of quantitative precipitation forecasts. *Mon. Wea. Rev.*, **136**, 4470–4487, <https://doi.org/10.1175/2008MWR2415.1>.
- White, A. B., B. J. Moore, D. J. Gattas, and P. J. Neiman, 2019: Winter storm conditions leading to excessive runoff above California’s Oroville Dam during January and February 2017. *Bull. Amer. Meteor. Soc.*, **100**, 55–70, <https://doi.org/10.1175/BAMS-D-18-0091.1>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 663 pp.
- Wolter, K., M. Hoerling, J. K. Eischeid, G. J. Van Oldenborgh, X. W. Quan, J. E. Walsh, T. N. Chase, and R. M. Dole, 2015: How unusual was the cold winter of 2013/14 in the Upper Midwest? [in “Explaining Extreme Events of 2014 from a Climate Perspective”]. *Bull. Amer. Meteor. Soc.*, **96** (12), S10–S14, <https://doi.org/10.1175/BAMS-D-15-00126.1>.
- Yeo, I.-K., and R. A. Johnson, 2000: A new family of power transformations to improve normality or symmetry. *Biometrika*, **87**, 954–959, <https://doi.org/10.1093/biomet/87.4.954>.
- Zschenderlein, P., A. H. Fink, S. Pfahl, and H. Wernli, 2019: Processes determining heat waves across different European climates. *Quart. J. Roy. Meteor. Soc.*, **145**, 2973–2983, <https://doi.org/10.1002/qj.3599>.
- Zwiers, F. W., and V. V. Kharin, 1998: Changes in the extremes of the climate simulated by CCC GCM2 under CO<sub>2</sub> doubling. *J. Climate*, **11**, 2200–2222, [https://doi.org/10.1175/1520-0442\(1998\)011<2200:CITEOT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<2200:CITEOT>2.0.CO;2).