Searching for the most extreme temperature events in recent history

Julien Cattiaux, Aurélien Ribes, Vikki Thompson

Searching for the most extreme temperature events in recent history

Julien Cattiaux[a] Aurélien Ribes,[a] Vikki Thompson,[b]

[a] *Centre National de Recherches Météorologiques (CNRM), Université de Toulouse, CNRS, Météo-France, Toulouse, France*

[b] *Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands*

Corresponding author: Julien Cattiaux, julien.cattiaux@meteo.fr

ABSTRACT: Because they are rare, extreme weather events have critical impacts on societies and ecosystems and attract public and scientific attention. The most unusual events are regularly documented as part of routine climate monitoring by meteorological services. A growing number of attribution studies also aim at quantifying how their probability has evolved under human induced climate change. However, it is often recognized that (i) the selection of studied events is geographically uneven, and (ii) the definition of a given event, in particular its spatio-temporal scale, is subjective, which may impact attribution statements. Here we present an original method that objectively selects, defines, and compares extreme events that have occurred worldwide in the recent years. Building on previous work, the event definition consists of automatically selecting the spatio-temporal scale that maximizes the event rarity, accounting for the non-stationary context of climate change. We then explore all years, seasons, and regions and search for the most extreme events. We demonstrate how our searching procedure can be both useful for climate monitoring over a given territory, and resolve the geographical selection bias of attribution studies. Ultimately, we provide a selection of the most exceptional hot and cold events in the recent past, among which are iconic heatwaves such as those seen in 2021 in Canada or 2003 in Europe.

SIGNIFICANCE STATEMENT: The purpose of the article is to objectively select and rank the most exceptional heatwaves and cold spells that have occurred worldwide in the recent years. As these events often have the greatest socio-economic impacts, better knowledge and characterization of historical events in a changing climate can inform adaptation strategies. We exhaustively scan temperature data over all years and regions to identify extreme events using the event probability as a universal metric. Applied over a specific location, such as on a national level, our method provides useful information for the climate monitoring of weather events. Applied globally, it can help attribution studies to pick events without the selection bias related to authors origin or media coverage.

CAPSULE: The most extreme historical heatwaves and cold spells across the globe are objectively selected from their ocurrence probability by an exhaustive scan of temperature events.

1. Introduction

Extreme weather events are rare, and so they inevitably attract attention and cause socio-environmental impacts when they occur. Explaining events that have occurred and putting them in a climate perspective constitutes a key challenge for both national weather services and the "event attribution" community.

The former have the responsibility to document historical weather events and maintain long-term statistics over their territory. This includes describing events from a synoptic perspective, quantifying their rarity (e.g. return period), and comparing with inventories of past events of the same type. The selection of documented events is generally made on the basis of fixed thresholds, so that events can be characterized by a level and duration of exceedance. As the climate warms, the thresholds for defining heatwaves (cold spells) are exceeded more (less) easily: for instance, 37 (16) of the 46 heatwaves (46 cold spells) officially reported in France by Météo-France over 1947–2022 have occurred after 1985, in the second half of the record period[1]. This raises new questions for climate monitoring. Are recent episodes of threshold exceedance really extreme events in today's climate? How rare were past events relative to their respective climate? Besides, due to possible adaptation, the impacts associated with a given temperature event are not necessarily the

[1]ClimatHD by Météo-France [http://www.meteofrance.fr/climat-passe-et-futur/climathd], online, accessed 2023-10-15.

same today as in the past. Taking non-stationarity into account in regular climate monitoring has therefore become a necessity for weather services.

Placing extreme weather events in the perspective of climate change is precisely the aim of "event attribution" studies. This field seeks to assess how climate change has altered event probability or intensity, through quantities such as the probability ratio or the fraction of attributable risk. It has emerged since the pioneering analysis of the European heatwave of summer 2003 by Stott et al. (2004). It has become a regular activity as illustrated by the series of supplements to the annual State of the Climate report since Peterson et al. (2012)[2] or the creation of the World Weather Attribution (WWA) group[3], and there is a growing interest from weather services in making it operational. Many events have been scrutinized over the past two decades, but it is often expressed that studies suffer from a "selection bias", as in the IPCC AR6 Chapter 11: "studies in the developing world are [...] lacking (Otto et al. 2020)" and "events that have been studied are not representative of all extreme events that occurred [...]" (Seneviratne et al. 2021). There are a variety of reasons for this bias (origin of the authors, quality of the data, media coverage, etc.), but so far little has been proposed to limit it and objectify event selection.

A common issue to climate monitoring and event attribution analyses is that their statements are dependent on the event definition, in particular the spatio-temporal extent considered. Weather services need to identify the duration and extent where the event was extreme to analyze it appropriately and provide relevant information to local decision-makers. In attribution studies, using large scales generally yields a higher detection of the signal (climate change) due to the smoothing out of the noise (internal variability), and therefore larger probability ratios (Angélil et al. 2014, 2018; Cattiaux and Ribes 2018; Kirchmeier-Young et al. 2019). Event definition is often subjective in the literature, which is referred to as another type of "selection bias" (Seneviratne et al. 2021). In a previous paper (Cattiaux and Ribes 2018, hereafter CR18), we proposed an objective definition procedure consisting in automatically selecting the spatio-temporal scale that maximizes the event rarity. This was motivated by the fact that the meteorological extremeness of an event is generally what makes it of interest and generates the greatest socio-economic impacts. Importantly, we showed that our procedure does not artificially bias attribution statements: maximizing the rarity does not systematically maximize (or minimize) the attributable signal.

[2]BAMS explaining extreme events from a climate perspective [https://www.ametsoc.org/ams/index.cfm/publications/bulletin-of-the-american-meteorological-society-bams/explaining-extreme-events-from-a-climate-perspective], online, accessed 2023-10-15.

[3]WWA webpage [https://www.worldweatherattribution.org], online, accessed 2023-10-15.

Here we show that this method can also be used to compare the rarity of different events, for example, two heatwaves of different spatio-temporal scales occurring in different years and at different locations. It can therefore serve both meteorological services to create inventories of historical events, and the event attribution community to address the issue of selection bias. Ultimately, we provide objective selections of historical heatwaves and cold waves at a national level (France) and at the global level, which constitute new information on observed extremes and can serve as databases for the validation of climate models or statistical methods.

2. The event probability as the ranking metric

Our study complements products of national weather services and rankings of extreme temperature events that have already been recently proposed in the literature. Russo et al. (2015) have proposed a ranking of European heatwaves on the basis of an index combining the fraction of area over which daily temperatures exceed reference (fixed) thresholds and the magnitude of the exceedances. Röthlisberger et al. (2021) and Boettcher et al. (2023) have developed a methodology to identify extreme seasonal temperatures at the grid point scale, also from fixed thresholds, and then form spatially coherent objects. Thompson et al. (2022) have compared the spectacular Pacific Northwest heatwave of June 2021 to the most extreme 1-day hot events ever recorded globally, defined at each region of the world as the largest daily temperature anomaly normalized with respect to local mean and variance. In a follow-up paper, Thompson et al. (2023) have also provided estimates of the return period of current records of daily maximum temperatures worldwide.

Here, as in CR18, we wish to introduce flexibility in the definition of the spatio-temporal scale of the events. We thus need a metric enabling the comparison of the temperature events not only over all dates and locations, but also over several temporal durations and spatial domain sizes. The probability of occurrence of the event in its factual climate — which will be denoted as p_1 throughout this paper — perfectly meets this requirement. Mathematically, it writes as the probability that an equally or more intense event occurs at the same place and time:

$$p_1 = \Pr\left\{X^{(t_1)} \geq x_{t_1}\right\} , \tag{1}$$

where $X^{(t_1)}$ is the random variable describing all possible realizations of the climate at time t_1, and x_{t_1} is the value effectively observed at that time. As discussed in CR18, p_1 has the intrinsic

property of being uniformly distributed within the interval [0,1]: by definition, a randomly selected (weather) event can indeed be anywhere in the (climate) distribution of all possible events. Provided that p_1 is well estimated, this property thus enables to confront all dates, locations and scales on the basis of a fair and universal metric.

We perform an exhaustive "scan" of all possible warm and cold events and identify the highest rarity (or lowest p_1) events. As p_1 is the probability of the event occurring the year it actually did, our procedure provides a non-stationary view of the monitoring of extremes — in a warming climate, a heatwave of a given intensity will become increasingly likely (less extreme) —, which can meet the needs of weather services. In attribution studies, p_1 is a key quantity, denoted "factual probability", which reinforces its relevance for our purpose. It is often compared to a "counterfactual probability" p_0 measuring the rarity of the event x_{t_1} in a different climate X, e.g. without human influence or at a other (distant) time t_0:

$$p_0 = \Pr\left\{X^{(t_0)} \geq x_{t_1}\right\} \ . \tag{2}$$

3. Scanning a given region: France

We first illustrate our scanning procedure at a fixed location: metropolitan France. We use the national daily thermal index provided by Météo-France over the 1947–2022 period (Ribes et al. 2022). We search for the most extreme warm and cold events in this time series (i.e. the lowest p_1), exploring all years, all days of a year, and several temporal durations of n days. To keep the computation time reasonable, we restrict n to 1–8, 10, 12, 15, 20, 25, 30, 40, 50, 60, 75, 90, 120, 180, 270, and 365 days (23 values). The methodology for estimating p_1 and p_0 is depicted in Figure 1 using the example of the 5-day heatwave of June 15–19, 2022. This episode is one of the three heatwaves reported by Météo-France in 2022 and the earliest one in records[4] based on the reference threshold[5] of 23.4 °C (Figure 1a).

As in CR18, the event rarity is assessed through an empirical estimation method consisting of (i) detrending the sample of observations x_t to make all values representative of the climate at $t_1 = 2022$ (factual) or $t_0 = 1950$ (counterfactual) and (ii) fitting stationary parametric distributions

[4]Weather extremes of summer 2022 by Météo-France [https://meteofrance.com/actualites-et-dossiers/actualites/changement-climatique-lete-2022-et-ses-extremes-meteorologiques], online, accessed 2023-10-15.

[5]Definition of heatwaves by Météo-France [https://meteofrance.com/actualites-et-dossiers/comprendre-la-meteo/temperature/comment-les-climatologues-evaluent-vagues-chaleur-canicules], online, accessed 2023-10-15.

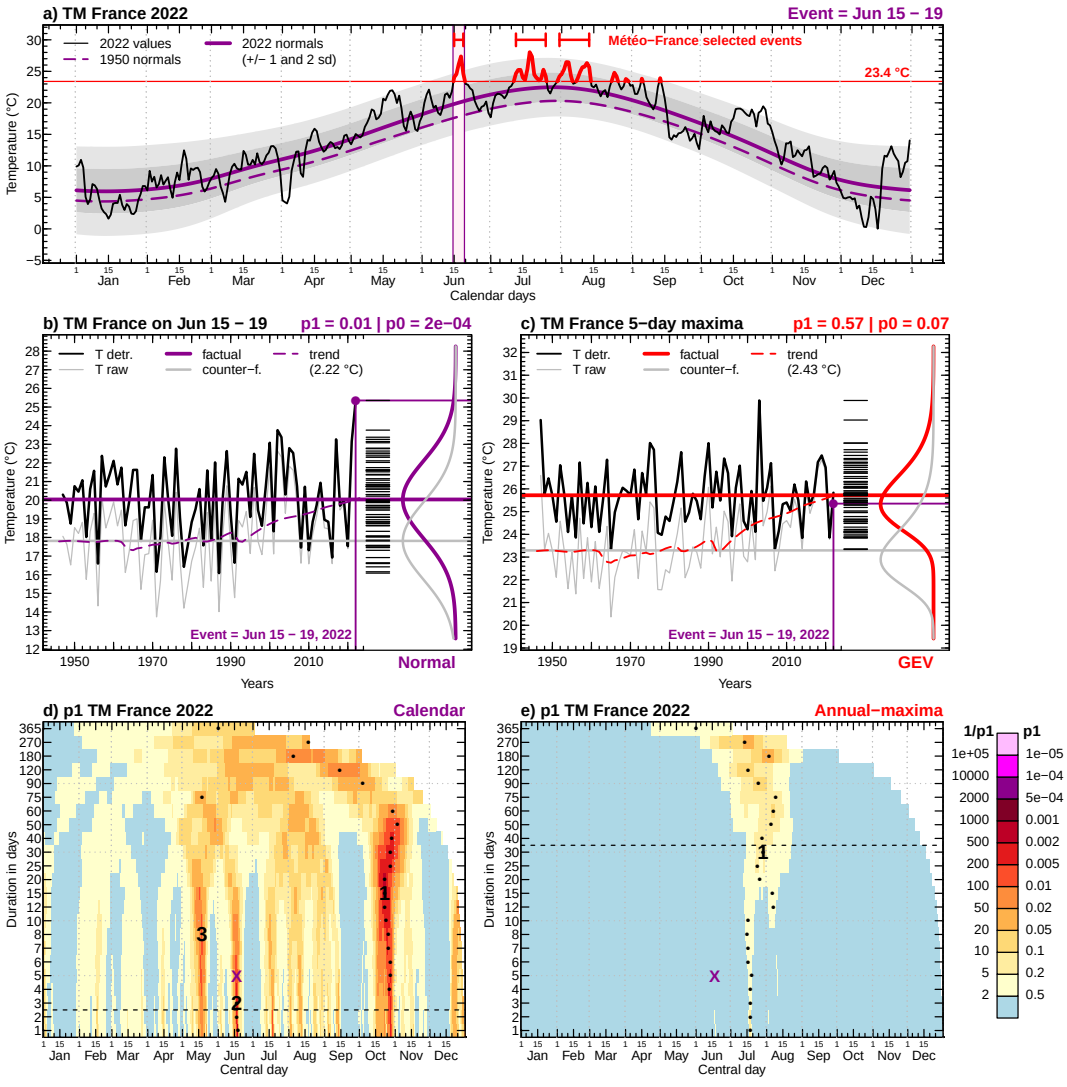

FIG. 1. **(a)** Daily-mean temperatures of 2022 (black). 2022 normals (solid violet), associated ±1 and ±2 s.d. (gray shading), and 1950 normals (dashed violet). Heatwaves reported by Météo-France (i.e. with 3 days above 23.4 °C) in red, with June 15–19 hightlighted. **(b)** Computing calendar p_1: raw (thin gray) and detrended (wrt. 2022, thick black) time series of June 15–19 temperatures. Long-term trend (dashed colored) and its levels in 2022 (horizontal colored, factual climate) and 1950 (horizontal gray, counterfactual climate). Barcode: detrended (factual) sample, with the fitted normal distribution (violet). Counterfactual distribution in gray. **(c)** Computing annual-maxima p_1: same as (b) except that the time series is Tx5day and the fits are GEV distributions. **(d)** Calendar and **(e)** annual-maxima p_1 as function of the time of the year (x-axis) and the temporal averaging (y-axis), with our selections of events indicated and ranked.

on the detrended samples to derive p_1 and p_0. Our detrending procedure is described in detail in Appendix A; in short, the shape of the trend is estimated from a multi-model ensemble of historical simulations and the amplitude from observations alone, with a dependency on the time of year and the temporal duration n. Two different options are then taken for the sample x_t and the fitted distribution to describe the event conditionally to its calendar context or not.

The first — calendar conditioning — consists of comparing the n-day mean temperature of the event with the multi-year sample of temperatures observed on the exact same dates, and using a Gaussian distribution fit after detrending (Figure 1b). In this case minimizing p_1 is equivalent to searching for the largest normalized anomalies, i.e. departures from the mean in numbers of standard deviations (s.d.)[6]. For the June 2022 event we find a 5-day anomaly of 5.2 K, (2.6 s.d), which gives a calendar p_1 of 0.005; the interpretation is that an equally or hotter heatwave had a 0.5 % chance to occur *at these exact dates* in 2022.

However a 5-day average temperature of 25.4 K is not that exceptional in France in the climate of 2022: it is actually close to the expectance of the annual maxima of 5-day mean temperatures (hereafter Tx5day) for that year (estimate of 25.8 K, Figure 1c). Thus the second option — no conditioning — consists of comparing the n-day event with the sample of Tx$n$day and using a Generalized Extreme Value (GEV) distribution fit after detrending. In this case $1/p_1$ is interpretable as a formal return period. For the June 2022 event, the annual-maxima p_1 is found to be 0.57, which means that a heatwave at least as hot had more than 50 % chance to occur *in any dates* in 2022, and translates into an estimated return period of $1/p_1 \sim 1.8$ years.

For both calendar and annual-maxima approaches, p_0 is estimated by detrending the respective x_t sample relative to $t_0 = 1950$ instead of $t_1 = 2022$ (gray distributions in Figure 1b,c). As climate has warmed — estimates of 2.22 °C for June 15–19 and 2.43 °C for Tx5day — p_0 is unsurprisingly smaller than p_1. Finally, it must be noted that:

- the formalism of Equations 1 and 2 and the use of annual "maxima" is well suited for hot events; to apply it to cold events one has to consider that X is the temperature multiplied by -1 and to consider annual "minima";
- there is a debate in the community as to whether the value of the event should be included in or excluded from the sample when estimating its probability of occurrence (Philip et al. 2020;

[6]For $X \sim N(\mu, \sigma)$, p_1 is a monotonic function of $(X - \mu)/\sigma$.

van Oldenborgh et al. 2021b). The question especially arises when the time series is stopped because the last event is particularly extreme and we want to study it. Solutions to properly formulate stopping rules have been proposed (e.g., Barlow et al. 2020). Here we decide to systematically include the value of the event in the sample, because (i) we scan all the events of the whole period so we have no particular stopping rule and (ii) GEV distributions for annual maxima of temperature are often right-bounded (negative shape parameters) and in case of a very rare event, excluding its value for the fit can lead to an estimated upper bound lower than the event effectively observed (as in Philip et al. 2022). In such a case, it would be nonsense to conclude that the event has $p_1 = 0$ while we *know* that it has occurred.

The scanning procedure then consists of computing calendar and annual-maxima (-minima) p_1 for all 27740 days of 1947–2022 (February 29s are omitted) and 23 n-day temporal extents, and search for the local minima of p_1 — the most extreme events — within these 27740×23 matrices. To do so, we select the absolute minimum of p_1, then mask all overlapping calendar windows (we consider that they correspond to the same event) and iterate while there remains p_1 smaller than a given threshold: we choose 0.01 for the calendar method and 0.25 for annual-maxima and -minima methods. In addition, we impose a minimal duration of 3 days for the calendar approach for readability and a maximal duration of 30 days for annual-maxima (-minima) approaches for the quality of GEV fits.

This procedure is first illustrated on the complete year 2022 (Figure 1d,e). Three hot events are selected in the calendar approach; the June event mentioned above ranks 2nd and is found to be the most extreme at the 3-day scale. In the annual-maxima approach, a unique 30-day event is selected that overlaps the two episodes reported by Météo-France in July and August. Its estimated return period is small ($1/p1 = 10$ years), reflecting the fact that exceeding the reference threshold of 23.4 °C has become likely in the climate of 2022 (see also Figure 1a). (The 180-day event from May to October has a lower p_1 but the use of annual blocks and GEV is questionable for such long durations).

Now expanding to the entire period, the calendar approach enables quick identification of abnormally hot events throughout the whole year, which is useful for routine climate monitoring and can be relevant for impacts studies (Figure 2a). Only the heatwaves occurring near the peak of the annual cycle are retained by the annual-maxima approach (Figure 2b). Both approaches

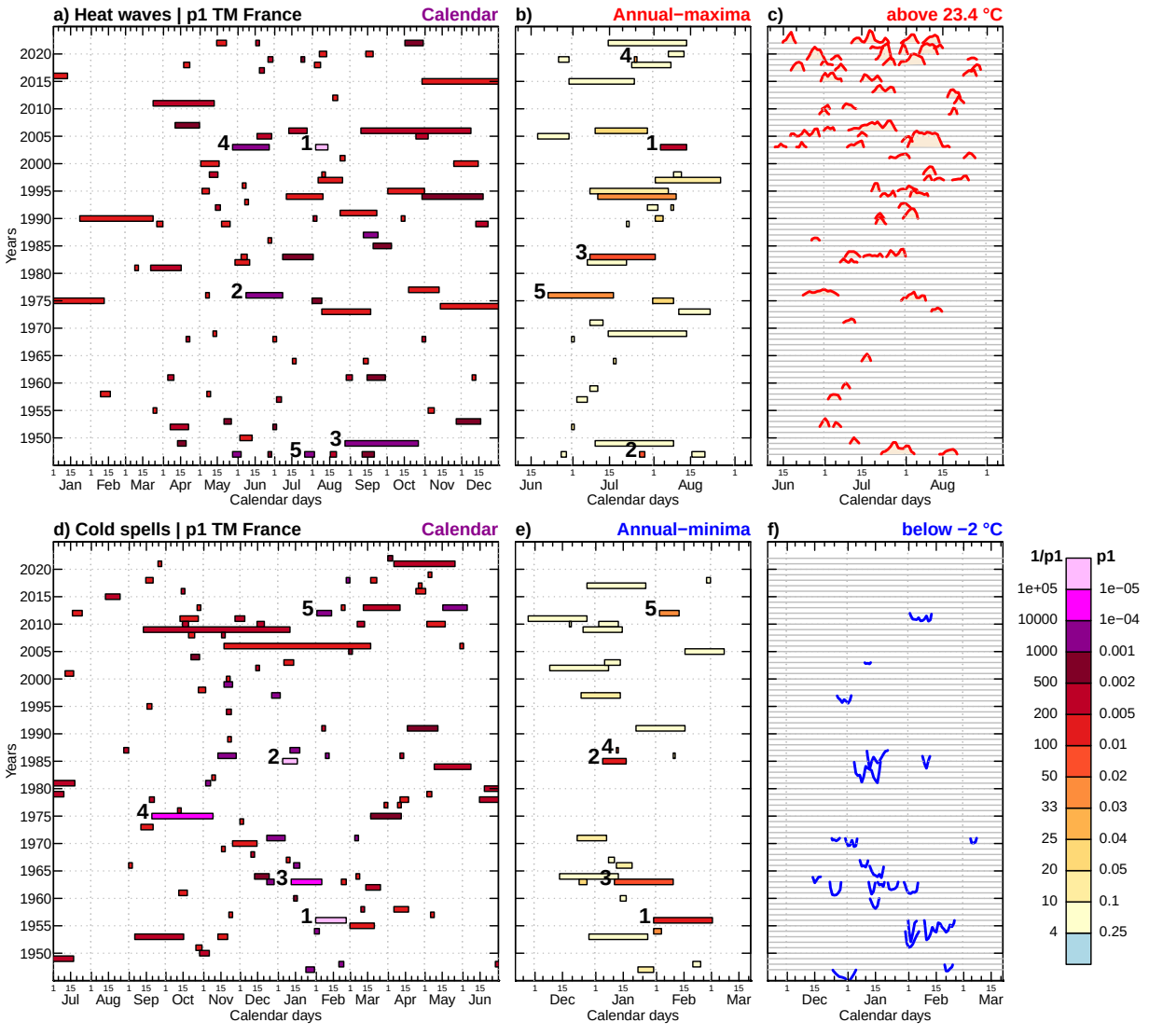

FIG. 2. **(a–c)** Selection of the hottest events with the (a) calendar and (b) annual-maxima approaches, compared with (c) Météo-France threshold-based selection (3 days above 23.4 °C, source of graphics: Les Décodeurs, Le Monde). **(d–f)** Same for cold events, with annual-minima in (e) and exceedances below -2 °C in (f). Rectangles and pieces of time series indicate the dates (x-axis) and years (y-axis) of the events, with a restricted x-axis on middle and right columns. Rectangle colors indicate p_1 with the five most extreme events highlighted. A minimal duration of 3 days and a maximal p_1 of 0.01 are imposed to calendar events. A maximal duration of 30 days and a maximal p_1 of 0.25 are imposed to annual-maxima and -minima events. p_1 values are lower in the calendar method due to the condition that the event occurs at these exact dates.

TABLE 1. The five hottest and coldest events selected by the annual-maxima and -minima approaches in France, ordered by increasing p_1 (see also Figure 2). Associated p_0 and p_1/p_0 are indicated, as well as average temperature and normalized anomaly. Brackets indicate 90%-level confidence intervals computed by bootstrapping years in the Tx*n*day or Tn*n*day samples; as random resampling removes the value of the year of interest from the N-year sample with a probability $(1-\frac{1}{N})^N \xrightarrow[N\gg1]{} e^{-1} \sim 0.37$, the upper bound of the confidence interval of $1/p_1$ can be found to be infinite whereas the event actually occurred.

#	Year	Dates	T (°C)	s.d.	$1/p_1$	$1/p_0$	p_1/p_0
Heatwaves							
1.	2003	Aug 4 - 13	28.6	4.4	410 [93 to Inf]	3200 [250 to Inf]	7.6 [2.6 to Inf]
2.	1947	Jul 27 - 28	27.7	3.3	97 [37 to 3200]	95 [37 to 3000]	0.98 [0.94 to 0.99]
3.	1983	Jul 8 - Aug 1	24.3	3	64 [31 to 350]	62 [30 to 320]	0.96 [0.92 to 0.97]
4.	2019	Jul 25	29.4	3	35 [20 to 130]	1100 [160 to Inf]	32 [7.1 to Inf]
5.	1976	Jun 22 - Jul 16	23.7	3.3	34 [20 to 100]	28 [17 to 77]	0.84 [0.75 to 0.89]
Cold spells							
1.	1956	Jan 31 - Mar 1	-3.2	-4.3	170 [56 to 8700]	180 [58 to 9700]	1 [1 to 1.1]
2.	1985	Jan 5 - 16	-6.9	-4.4	110 [44 to 880]	110 [43 to 850]	0.99 [0.97 to 0.99]
3.	1963	Jan 11 - Feb 9	-2.8	-3.7	96 [36 to 4100]	120 [42 to 11000]	1.2 [1.1 to 1.9]
4.	1987	Jan 12	-9	-3.8	50 [25 to 220]	45 [23 to 180]	0.89 [0.8 to 0.92]
5.	2012	Feb 3 - 12	-4.3	-3.7	46 [25 to 150]	18 [12 to 35]	0.38 [0.21 to 0.49]

agree that the greatest hot event for France is the heatwave of early August 2003, considered as a 10-day event (August 4–13, consistent with CR18). It corresponds to a normalized anomaly of 4.4 s.d. and an estimated return period of 410 [95 to Inf] years (Table 1, see caption for details on the confidence interval). Other major heatwaves are found in summers 1947, 1976, 1983 and 2019, with durations of respectively 2, 25, 25 and 1 days. This is overall consistent with the events selected by Météo-France on the basis of fixed thresholds (Figure 2c). As for cold spells, the most extreme events are found in winters 1956, 1963, 1985, 1987 and 2012 (Figure 2d,e and Table 1), which is also consistent with Météo-France reporting (Figure 2f). The events of February 1956 and January 1985 have normalized anomalies of −4.4 s.d. (Table 1), i.e. are equally unusual as the August 2003 heatwave from a calendar perspective; however, their formal return periods estimated from the annual-minima approach are shorter (respectively 170 and 110 years).

The right panels in Figure 2 illustrate the limitations of climate monitoring of extremes using fixed threshold approaches. On the one hand, using the same threshold for the whole year makes it impossible to describe certain events conditionally to their calendar context, e.g. late cold snaps in

early spring, which have an impact on vegetation (Vautard et al. 2022). On the other hand, global warming leads to heatwave thresholds being exceeded almost every summer (we can no longer speak of "extreme" events), while cold spell thresholds are almost never reached (we can no longer speak of events at all). Our method provides complementary and useful information for national weather services on both these calendar and non-stationary aspects. Lastly, note that an interactive application associated to this paper enables to reproduce the analyses carried out in this section and retrace Figure 1 for various choices of years and dates and Figure 2 for different thresholds of p_1 and duration (details in Appendix C).

4. Scanning worldwide

To extend the scanning procedure to the global scale, we use daily-mean 2-meter temperatures provided by the ERA5 reanalysis over the 1959–2022 period (Hersbach et al. 2020). The advantage of ERA5 is that it is a physically consistent dataset globally (model simulation) with a sufficiently fine resolution for our purpose and without missing values, which is particularly convenient for the calculation of annual maxima or minima. The main disadvantage is that it is not temporally homogenized (assimilated observations evolve in time), which can cause problems for both our detrending and p_1 calculations. Nevertheless, we have verified that the main results of this section, i.e. the selections of extremes, do not suffer from obvious heterogeneity problems. To our knowledge, there exists no observational dataset of global daily temperature with time homogeneity, long history, and no missing values.

The exploration of spatial domains is performed with the hierarchical collection of economic and political regions defined by Stone (2019) within the "Weather Risk Attribution Forecast" (WRAF) framework. This collection provides successive divisions of land areas into regions of 10, 5, 2, and 0.5 million km^2 (hereafter Mkm^2), each division being nested in the previous one, with only a few gaps in the global coverage (for instance Europe appears only from 5 Mkm^2). The set of 0.5 Mkm^2 regions was also used by Thompson et al. (2022, 2023). The full range of spatial sizes covers the vast majority of event definitions in recent attribution studies, with the exception of studies at global scale or at the local scale (in this case we could refine within a 0.5 Mkm^2 region). Scanning the full period at global scale thus involves considering 64 years $\times$ 365 days $\times$ 347 WRAF regions $\times$ 23 time durations $\sim$ 186 million temperature events.

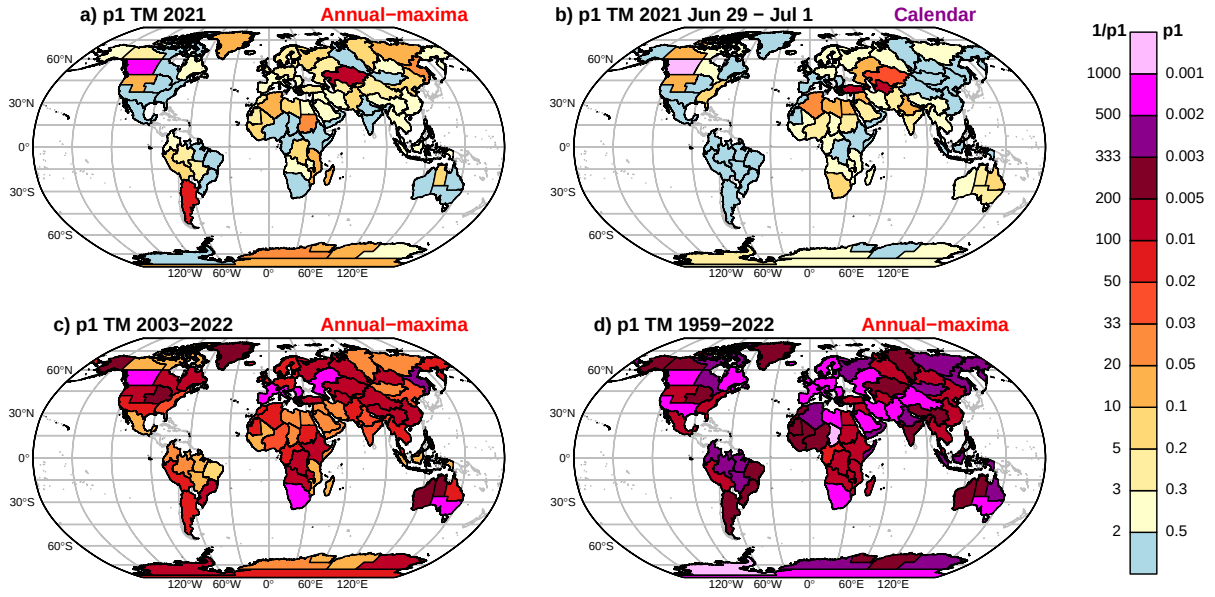

FIG. 3. Scanning heatwaves worldwide. **(a)** For each 2 Mkm^2 region, the minimum $p1$ of all 2021 events in the annual-maxima approach is shown. **(b)** Calendar p_1 of a warm event occurring on June 29 – July 1, 2021, i.e. coinciding with the absolute minimum of p_1 in (a). **(c–d)** Same as (a) for (c) 2003–2022 and (d) 1959–2022.

Firstly, we scan a single year to identify the most extreme events of that year, as this can be useful for annual reports of climate monitoring and event attribution. Figures 3a and 4a provide the minima of p_1 at each 2 Mkm^2 region for respectively the annual-maxima approach (heatwaves) in 2021 and the annual-minima approach (cold spells) in 2022. We have chosen these two examples because outstandingly, west Canada holds the strongest event in both cases, i.e. the 3-day heatwave of June 29 – July 1, 2021 — the famous "heat dome" analyzed in numerous studies (e.g., Bercos-Hickey et al. 2022; Philip et al. 2022; Schumacher et al. 2022; Thompson et al. 2022; Terray 2023; White et al. 2023) — and the 5-day cold spell of December 19–23, 2022, related to winter storm "Elliott". We estimate that these two events have remarkably low values of p_1, corresponding to estimated return periods of 810 years for the heatwave and 550 years for the cold spell (Table 2). There is certainly no meteorological relationship between these two events, suggesting that this region has, by chance, experienced particularly extreme temperatures over the past two years. For a closer look, the equivalent of Figure 1 for these two events has been plotted in Figures S1 and S2. The top 3 heatwaves in 2021 then include a 7-day July event in Kazakhstan and a 30-day event in December 2021 – January 2022 in Argentina, while a 30-day November event in Congo

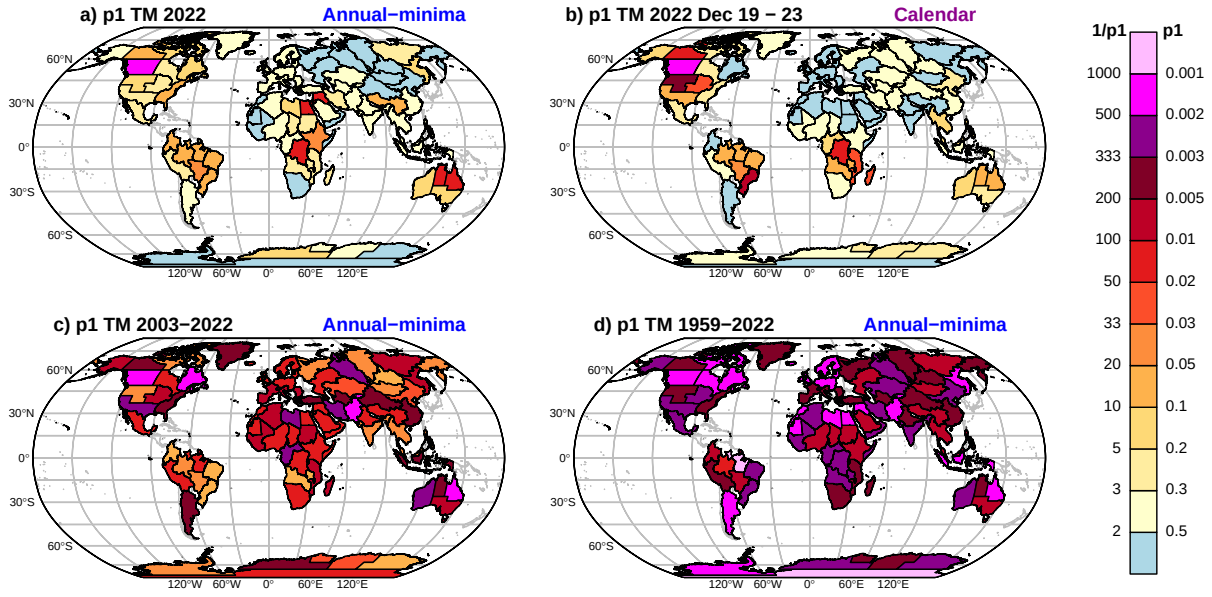

FIG. 4. Scanning cold spells worldwide. **(a)** For each 2 Mkm^2 region, the minimum $p1$ of all 2022 events in the annual-minima approach is shown. **(b)** Calendar p_1 of a cold event occurring on 2022, December 19–23, 2022, i.e. coinciding with the absolute minimum of p_1 in (a). **(c–d)** Same as (a) for (c) 2003–2022 and (d) 1959–2022.

and a 12-day July event in Queensland complete the podium of cold spells 2022. By construction, about half of the regions have a minimum p_1 above 0.5 in Figures 3a and 4a; for a given temporal duration, the distribution of yearly minimum p_1 across regions is indeed uniform (Appendix B).

Figures 3b and 4b show the calendar p_1 associated with the 2021 and 2022 west Canadian events. We find normalized calendar anomalies of respectively 5.0 and 3.0 s.d. over this region, which highlights the extremeness of these two events (this gives a lower p_1 than in the annual-maxima approach since here it is conditioned by the time of the year). The global picture of calendar p_1 can be useful for routine climate monitoring: for example here we see that the 2021 heatwave was restricted to west Canada while the 2022 cold spell also spreads over the northwest USA region (2.7 s.d.), and to a lesser extent over NWT & Yukon (2.2 s.d.) and midwest USA (1.9 s.d.). This is consistent with the characteristic spatial size of atmospheric patterns responsible for summer vs. winter extremes in mid-latitudes. Figures 3a also shows that the 2021 Kazakhstan heatwave occurred concurrently with the west Canadian one, which could be a coincidence or result from an amplified circumglobal wave pattern. Note that for given dates, the distribution of calendar p_1

TABLE 2. The ten hottest and coldest worldwide events selected by the annual-maxima and -minima approaches for 2 Mkm^2 regions, ordered by increasing p_1 (see also Figure 3 and 4). Brackets indicate 90%-level confidence intervals computed as in Table 1 with the same remark about the possibly infinite upper bound of $1/p_1$. Region ID and names from Stone (2019).

#	Reg ID	Region name	Year	Dates	$1/p_1$	$1/p_0$	p_1/p_0
Heatwaves							
1.	1.2.1	west Canada	2021	Jun 29 - Jul 1	810 [120 to Inf]	39000 [630 to Inf]	48 [4.4 to 540]
2.	7.1.2	southwest Russia	2010	Jul 21 - Aug 14	780 [110 to Inf]	46000 [520 to Inf]	59 [3.9 to 1700]
3.	X.2.2	south EEA	2003	Aug 3 - 14	740 [120 to Inf]	Inf [1800 to Inf]	Inf [11 to Inf]
4.	11.1.2	southeast Australia	2019	Jan 14 - 17	670 [110 to Inf]	Inf [44000 to Inf]	Inf [72 to Inf]
5.	6.2.2	south SADC	2016	Jan 1 - 7	630 [100 to Inf]	1700 [160 to Inf]	2.7 [1.5 to 3.9]
6.	8.2.2	south Pacific Russia	2011	Jul 17 - 28	390 [83 to Inf]	99000 [610 to Inf]	250 [6 to Inf]
7.	2.1.3	Midwest USA	2012	Jun 27 - Jul 26	320 [75 to Inf]	7500 [240 to Inf]	24 [2.8 to Inf]
8.	11.1.1	Western Australia	2019	Dec 22 - 29	320 [77 to Inf]	37000 [330 to Inf]	120 [3.4 to Inf]
9.	2.1.1	Alaska (USA)	2019	Jul 6 - 10	310 [71 to Inf]	33000 [360 to Inf]	110 [4.3 to Inf]
10.	11.2.1	Northern Territory (Australia)	2019	Dec 17 - 26	270 [70 to Inf]	560 [110 to Inf]	2.1 [1.5 to 380]
Cold spells							
1.	1.2.3	east Canada	2014	Jan 2 - 3	760 [120 to Inf]	12 [8.3 to Inf]	0.016 [0 to 0.085]
2.	11.2.2	Queensland (Australia)	2007	Jun 20 - Jul 19	620 [110 to Inf]	84 [35 to Inf]	0.14 [0 to 0.37]
3.	9.2.2	southeast ECO	2008	Jan 18 - 27	560 [97 to Inf]	100 [38 to Inf]	0.19 [0 to 0.42]
4.	1.2.1	west Canada	2022	Dec 19 - 23	550 [110 to Inf]	6.2 [4.7 to 13]	0.011 [0 to 0.051]
5.	5.2.2	central and south CEMAC	2015	Jan 11 - 12	420 [94 to Inf]	5.1 [4 to 8.1]	0.012 [0 to 0.047]
6.	11.1.1	Western Australia	2006	Jul 12 - 17	400 [87 to Inf]	47 [24 to Inf]	0.12 [0.071 to 0.35]
7.	4.2.1	Libya	2017	Dec 31 - Jan 11	360 [82 to Inf]	20 [12 to Inf]	0.054 [0 to 0.19]
8.	7.2.1	West Siberia (Russia)	2006	Jan 11	340 [75 to Inf]	240 [63 to Inf]	0.7 [0.58 to 0.87]
9.	2.2.1	southwest USA	2013	Dec 25 - Jan 18	340 [74 to Inf]	5.8 [4.2 to 9.7]	0.017 [0 to 0.077]
10.	9.1.2	Iran	2008	Jan 8 - 17	330 [78 to Inf]	44 [21 to 7000]	0.13 [0 to 0.29]

CEMAC = Economic and Monetary Community of Central Africa, ECO = Economic Cooperation Organization, EEA = European Economic Area, SADC = Southern Africa Development Community, USA = United States of America.

across regions is expected to be uniform in our procedure, so that about half of the regions are expected to have p_1 above 0.5 in such maps (Appendix B).

We now extend the scan to two time periods: the last 20 years (2003–2022, Figures 3c and 4c and Table 2) over which event attribution has developed, and the whole period of study (1959–2022, Figures 3d and 4d) — the interactive application enables to further visualize the results of any year or period (Appendix C). The podium of heatwaves of the last 20 years is occupied by events well known in the scientific community (Figure 3c): west Canada 2021, southwest Russia 2010, and south Europe 2003 (the same as in France in Figure 2). This first one is consistent with Thompson

et al. (2022) while the last two are in Russo et al. (2015). Other major events include heatwaves that have been well documented in the scientific literature (e.g. January and December 2019 in Australia and associated bushfires, e.g. van Oldenborgh et al. (2021a)), and others for which no attribution study exist (e.g. South Africa 2016, south Pacific Russia 2011, see Table 2), illustrating the geographic selection bias. Over the whole period, only two events of size 2 Mkm^2 (May 1998 in Chad and January 1993 in west Antarctica, Figure 3d) appear to exceed the 2021 west Canadian heatwave, with estimated return periods of 2200 and 1100 years respectively.

On the cold spell side, a 4-day event in July 1975 in central North Brazil wins the contest in the 2 Mkm^2 category, with January 2014 in east Canada being the most extreme event of the last 20 years (Figure 4c,d and Table 2). The December 2022 "Elliott" storm in west Canada ranks 4th over 2003–2022 and 12th over the whole period, which again highlights its extremeness in the climate of 2022.

Lastly, the hierarchical character of the collection of WRAF regions is used to refine the selection of events among spatial scales. Figure 5 shows the minimum p_1 associated with the three major heatwaves identified in Figure 3 and Table 2 on the geographical division of the 10, 5, 2 and 0.5 Mkm^2 regions. The 2021 Canadian event is found to be the most extreme for a 0.5 Mkm^2 region (Alberta), the 2010 Russian event for a 5 Mkm^2 region (European Russia), and the 2003 European event for the 2 Mkm^2 region as in Figure 3.

More generally, when minimizing p_1 both over time and the entire collection of WRAF regions for the period 2003–2022, most of the events of Table 2 stand out, sometimes on parent or child domains (Table 3). Notably, two Chinese heatwaves slip into the trio of heatwaves described above: a 3-day event in 2003 and a 25-day event in 2022, with estimated return periods similar to the 2021 Canadian heatwave. These two events were reported in the media[7] but seem to have received less scientific attention. Extending the scan to the entire period 1959–2022 can be done in the interactive application (Appendix C): other events join the top 10, but sometimes with obvious data problems (Appendix B). The top events remain the heatwave of May 1998 in Chad (at size 2 Mkm^2) and the cold spell of July 1975 in Brazil (at size 0.5 Mkm^2, south Para).

Using the collection of WRAF regions can thus be helpful to both (i) identify the most extreme events of a given period worldwide, and (ii) give a rough idea of their spatial size. However extreme

[7] eg. https://www.lemonde.fr/en/asia-and-pacific/article/2022/08/24/china-s-record-heatwave-causes-severe-threat-to-crop-production_5994568_153.html

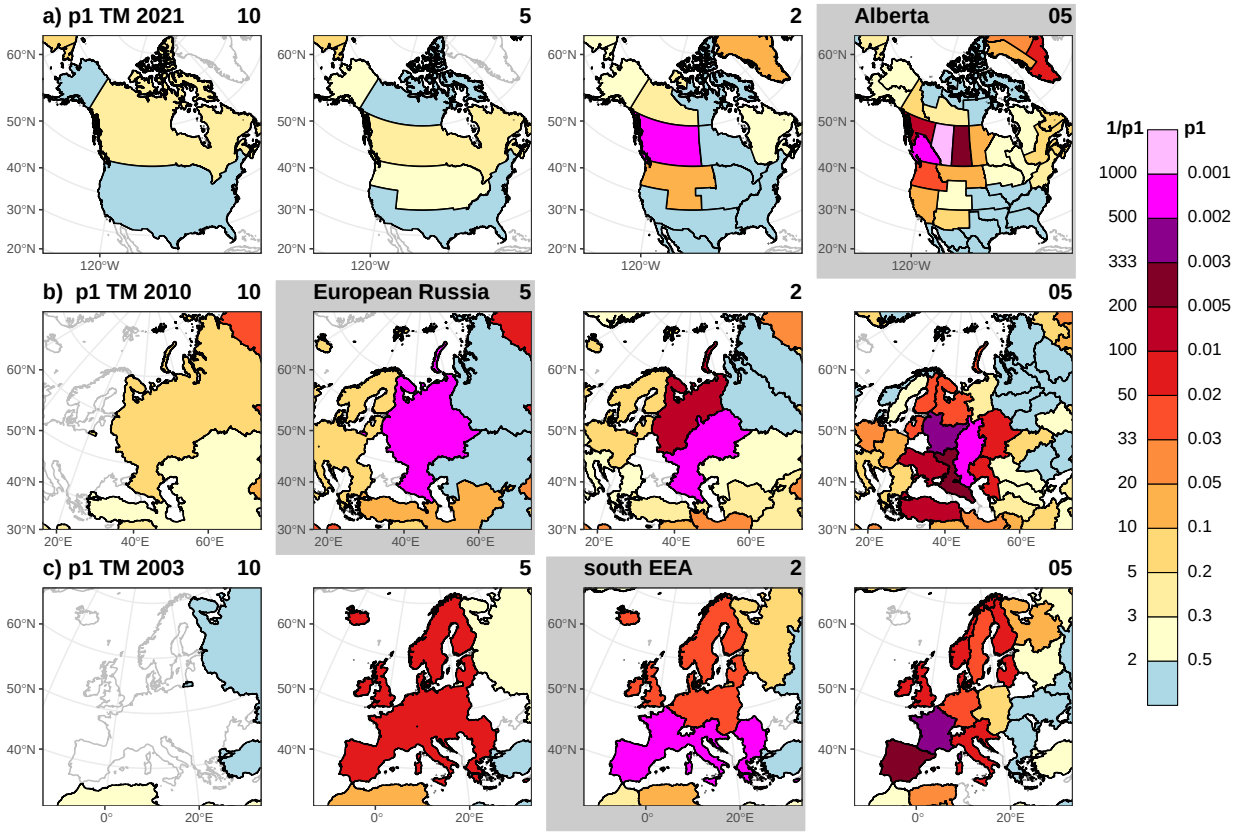

FIG. 5. Zooming over a region. **(a)** For each 10, 5, 2 and 0.5 Mkm^2 region (from left to right), the minimum $p1$ of all 2021 events in the annual-maxima approach is shown, which coincides with the 2021 Canadian heatwave listed as 1st in Table 2 and 2nd in Table 3. **(b)** Same as (a) for the 2010 Russian heatwave (2nd and 4th). **(c)** Same as (a) for the 2003 European heatwave (3rd and 5th). For each event, the region with the lowest p_1 is explicitely named and its panel (i.e. region size) is highlighted in gray.

events do not stop at administrative borders, and once an event has been selected for a study, its precise definition may be decided through a more systematic exploration of various spatial domains and sizes (e.g., CR18), and possibly on a more suitable dataset than ERA5 for the area of interest.

5. Discussion and conclusions

We have objectively identified and ranked the most extreme temperature events over a large spatio-temporal range, using the probability of event occurrence in the factual climate (p_1) as a universal metric. We produce a list of the greatest hot and cold extremes over the recent past, together with their spatio-temporal scale. Applied locally, such as on a national level, this

TABLE 3. Same as Table 2 but for the entire collection of WRAF regions. The number of characters in the region ID indicates its size: X for 10 Mkm^2, X.X for 5 Mkm^2, X.X.X for 2 Mkm^2 and X.X.X.X for 0.5 Mkm^2. Region ID and names from Stone (2019).

#	Reg ID	Region name	Year	Dates	$1/p_1$	$1/p_0$	p_1/p_0
Heatwaves							
1.	10.2.2.3	Fujian, Hunan, and Jiangxi (China)	2003	Jul 30 - Aug 1	1100 [130 to Inf]	50000 [410 to Inf]	47 [2.7 to 2200]
2.	1.2.1.3	Alberta (Canada)	2021	Jun 29 - Jul 1	1000 [120 to Inf]	5900 [240 to Inf]	5.7 [1.9 to 12]
3.	10.2.1.3	Chongqing and Sichuan (China)	2022	Aug 3 - 27	1000 [130 to Inf]	Inf [2200 to Inf]	Inf [11 to Inf]
4.	7.1	European Russia	2010	Jul 23 - Aug 11	790 [120 to Inf]	31000 [450 to Inf]	40 [3.4 to Inf]
5.	X.2.2	south EEA	2003	Aug 3 - 14	740 [120 to Inf]	Inf [2000 to Inf]	Inf [11 to Inf]
6.	3.2.3.2	southern Brazil	2014	Jan 19 - Feb 12	730 [98 to Inf]	110000 [560 to Inf]	150 [5 to Inf]
7.	6.2.2.4	east South Africa	2016	Jan 6 - 7	690 [110 to Inf]	4700 [250 to Inf]	6.8 [2.2 to 140]
8.	11.1.2	southeast Australia	2019	Jan 14 - 17	670 [120 to Inf]	Inf [31000 to Inf]	Inf [77 to Inf]
9.	X.X.1.2	Northeast Greenland National Park	2012	Jul 10 - Aug 8	640 [110 to Inf]	Inf [150000 to Inf]	Inf [260 to Inf]
10.	2.1.1.2	south Alaska (USA)	2019	Jul 5 - 9	640 [90 to Inf]	34000 [350 to Inf]	53 [3.3 to 2000]
Cold spells							
1.	1.2.3.2	south Québec (Canada)	2014	Jan 2 - 3	870 [130 to Inf]	59 [28 to Inf]	0.068 [0 to 0.23]
2.	6.2.2.3	west South Africa	2014	Jul 6 - 10	740 [120 to Inf]	42 [22 to Inf]	0.057 [0.038 to 0.23]
3.	1.2.3.1	Nord-du-Québec (Canada)	2015	Jan 31 - Feb 19	730 [100 to Inf]	14 [9.1 to 70]	0.019 [0 to 0.12]
4.	10.2	south China	2008	Jan 22 - Feb 5	700 [110 to Inf]	43 [21 to Inf]	0.061 [0.039 to 0.23]
5.	4.1.2.2	west Algeria	2005	Jan 28 - Feb 1	690 [110 to Inf]	18 [11 to Inf]	0.026 [0 to 0.12]
6.	11.2.2.2	central Queensland (Australia)	2007	Jun 7 - Jul 1	690 [110 to Inf]	130 [44 to Inf]	0.18 [0.14 to 0.44]
7.	7.2.1.1	Yamalo-Nenets (Russia)	2006	Jan 11	680 [120 to Inf]	37 [20 to Inf]	0.055 [0.037 to 0.2]
8.	8.1.2.2	northeast Sakha (Russia)	2013	Jan 29 - Feb 27	650 [94 to Inf]	17 [10 to 130]	0.027 [0 to 0.12]
9.	11.2.2.3	south Queensland (Australia)	2007	Jul 13 - 20	650 [120 to Inf]	120 [49 to Inf]	0.19 [0.14 to 0.44]
10.	6.1.1.3	west DR Congo	2022	Nov 25 - Dec 24	630 [110 to Inf]	1 [1 to 1]	0.0016 [0 to 0.0094]

DR Congo = Democratic Republic of Congo, EEA = European Economic Area, USA = United States of America.

procedure can provide additional method for the climate monitoring of weather events in a warming world. Applied globally, it ensures extreme events are identified objectively, independently of the populations concerned or the media coverage. This can help the event attribution community both by providing objective lists of events to be studied in dedicated annual reports, without selection bias, and by building reference databases of extreme events on which climate models and statistical attribution methods can be confronted.

Our procedure is not without limitations, particularly because applying it in such a generalized and exhaustive way requires use of simple and fast methods. A first remark concerns the long-term trend, which is crucial when comparing extreme events of various years. As its amplitude is

here estimated solely from the observations (Appendix A and Figure S3), it may contain traces of decadal internal variability, and this can locally affect p_1 estimates. In regions where internal variability causes recent years to be persistently warmer than the forced response — as suspected e.g. in France (Ribes et al. 2022) —, our procedure could overestimate the trend and underestimate the rarity of recent hot events. Using a long-term trend estimated from a combination of observed and modeled data would be more reliable but for future work. Importantly, the histogram of years of selected events or the time series of yearly minimum p_1 do not suggest any collective bias related to trend estimates (Appendix B and Figure S4.).

A second remark concerns the estimation procedure for p_1, more precisely the fit of parametric laws onto observed samples. To compare various temporal durations and spatial domains, it is assumed here that the same, simple, distributions apply to all seasons, regions, and a wide range of time and space scales (with different parameters). For the calendar approach, we use the normal distribution, which is classical for temperature, but does not account for potential asymmetric behaviors (e.g., McKinnon and Simpson 2022). For the annual-maxima or -minima approaches, we use GEV distributions to fit Tx$n$day or Tn$n$day samples, which is well suitable for small n (here we have limited to $n = 30$) but less for large n (typically n=365, ie. full year). In both cases, we find pretty uniform histograms of the p_1 of all events in all regions, suggesting that the estimation procedure is "collectively" correct (Appendix B and Figure S3.). Ideally, one would need a family of distributions that allows a slow transition from skewed or GEV distributions for small scales to normal distributions when the law of large numbers starts to apply. This would affect the values of p_1, but without changing the philosophy of our work (the idea of focusing on p_1) nor even the ranking of events in a dramatic way.

Lastly, exploring extreme events in a generalized way over so many years and regions requires a high quality data set of temperature observations with temporal homogenization. Here we use ERA5 which performs well over many regions where the data assimilation and the numerical model are efficient. For France we have verified that ERA5 and the Météo-France thermal index are very close. However there are some regions where one may question the data quality (Appendix B and Figure S4) as previously noted by Thompson et al. (2022) in their ranking of heatwaves. Repeating the scanning exercise with other datasets could allow to quantify how observational uncertainties reverberate in our results.

Overall, we show that p_1 is an appropriate metric to identify the most extreme weather events in recent history. We have opted for a non-stationary view of climate monitoring — p_1 is the probability of occurrence in the factual, evolving, climate — but minimizing p_0 instead of p_1 can provide a stationary view. The method can be extended to other meteorological variables, such as precipitation or surface winds; yet the estimation of p_1 over various space and time scales is more delicate than for temperature. Following van der Wiel et al. (2020), it can also be adapted to select extreme events in terms of impacts, not weather, by replacing the meteorological variable (e.g. temperature) with an impact-oriented variable (e.g heat stress). Writing p_1 as a joint probability would also allow to consider compound events. Ultimately, the range of applications for such an exhaustive spatio-temporal scan goes beyond the area of extreme events: for example, replacing p1 with an appropriate evaluation metric could enable the identification of regions and periods where discrepancies between observations and climate models are the strongest, which could make this approach useful for a much wider community.

Acknowledgments. The authors are grateful to M. Wehner (LBNL) and two anonymous reviewers for their careful proofreading. They thank Y. Robin and M. Schneider (Météo-France) for providing the France thermal index and the associated selections of events and D. Stone (National Institute of Water & Atmospheric Research) for the WRAF regions files. They acknowledge the ERA5 dataset provided by the ECMWF and climate modeling groups involved in CMIP6 for producing and making available their simulations. J.C. particularly thanks S. Qasmi and R. Samacoïts (Météo-France) for their great support on spatial and interactive visualization with R and the shiny package (https://cran.r-project.org) and M. Goar (Le Monde).

Data availability statement. Input data and scanning results are available on a Zenodo archive: https://doi.org/10.5281/zenodo.7966559. This includes Météo-France thermal index, ERA5 data averaged over WRAF regions, CMIP data used for the detrending, WRAF data used for the analysis, and date vs. duration matrices of estimated p_1 and p_0 for the three methods (calendar, annual-maxima and -minima) and all regions. R scripts for computing the scanning procedure and visualizing the results are available on GitHub: https://github.com/jlncttx/CRT23/. Figures and Tables can also be reproduced on the interactive application: https://jlncttx.shinyapps.io/CRT23-app/.

APPENDIX A

Details on the detrending procedure

The statistical procedure used to estimate p_1 involves (i) detrending the time series (making it stationary) and (ii) using standard statistical methods for stationary data. It is equivalent to empirical methods using non-stationary fits on observations (Philip et al. 2020) but we prefer to do a two-step process to ensure consistency of trends between consecutive calendar dates or n-day durations. Here we detail how we correct a (non-stationary) yearly temperature time series with respect to a reference year to make all values representative of the (stationary) climate of that year. We do that by fitting observations onto a multi-model estimate of climate change (forced response F); this is illustrated with the example of France in Figure S3.

First, following Ribes et al. (2020), we estimate the pattern of long-term warming at the location of interest by isolating natural and anthropogenic forced responses in a multi-model average of annual mean temperatures. We use an ensemble of 227 historical + SSP5-8.5 members from 46

models[8] of the Phase 6 of the Coupled Model Intercomparison Project and write the multi-model multi-member mean of yearly temperatures $\mu(y)$ as:

$$\mu(y) = \underbrace{\mu_{nat}(y) + \mu_{ant}(y)}_{F(y)} + \varepsilon(y) \ , \tag{A1}$$

where μ_{nat} the natural forced response, μ_{ant} the anthropogenic forced response, $F = \mu_{nat} + \mu_{ant}$ the total forced response and ε the residual internal variability. This separation is done through a Generalized Additive Model fit, considering that μ_{nat} is proportional to the response of an Energy Budget Model to natural forcings (Geoffroy et al. 2013) and μ_{ant} is a smooth spline with 10 degrees of freedom (Ribes et al. 2020; Robin and Ribes 2020).

Second, we subtract from the sample of observations its linear fit onto $F(y)$. For the calendar approach, the detrending is done on a daily basis, prior to the computation of n-day averages, to account for potential changes in the annual cycle (e.g. a greater warming in summer than in winter). Following Rigal et al. (2019), we decompose the daily temperature T of day d and year y:

$$\mathrm{T}(d,y) = \mathrm{T}_{cyc}(d) + \beta_{\mathrm{T}}(d) \times F(y) + \mathrm{T}'(d,y) \ , \tag{A2}$$

with T_{cyc} the average annual cycle (equals to $\frac{1}{n_y}\sum_y \mathrm{T}(d,y)$), β_{T} the scaling factor of each calendar day onto the forced response F (centered so that $\sum_y F(y) = 0$) and T' the residual or daily anomaly (verifies $\Sigma_y \mathrm{T}'(d,y) = 0$). Note that T_{cyc} and β_{T} are smoothed using periodic splines of respectively 12 and 6 degrees of freedom, and that $\mathrm{T}_{cyc} + \beta_{\mathrm{T}} \times F$ provides an estimate of the daily non-stationary normals (i.e. the average annual cycle plus a deformation associated with the long-term trend). Then the detrending of $\mathrm{T}(d,y)$ with respect to the year y_1 writes:

$$\mathrm{T}^{(y_1)}(d,y) = \mathrm{T}(d,y) - \beta_{\mathrm{T}}(d) \times (F(y) - F(y_1)) \ . \tag{A3}$$

[8]ACCESS-CM2 (3 members), ACCESS-ESM1-5 (10), AWI-CM-1-1-MR (1), BCC-CSM2-MR (1), CAMS-CSM1-0 (2), CanESM5-CanOE (3), CanESM5 (50), CAS-ESM2-0 (2), CESM2 (3), CESM2-WACCM (3), CIESM (1), CMCC-CM2-SR5 (1), CMCC-ESM2 (1), CNRM-CM6-1-HR (1), CNRM-CM6-1 (6), CNRM-ESM2-1 (5), E3SM-1-1 (1), EC-Earth3-CC (1), EC-Earth3 (3), EC-Earth3-Veg-LR (3), EC-Earth3-Veg (4), FGOALS-f3-L (1), FGOALS-g3 (4), FIO-ESM-2-0 (3), GFDL-CM4 (1), GFDL-ESM4 (1), GISS-E2-1-G (7), HadGEM3-GC31-LL (4), HadGEM3-GC31-MM (4), IITM-ESM (1), INM-CM4-8 (1), INM-CM5-0 (1), IPSL-CM6A-LR (7), KACE-1-0-G (3), KIOST-ESM (1), MCM-UA-1-0 (1), MIROC6 (49), MIROC-ES2L (9), MPI-ESM1-2-HR (2), MPI-ESM1-2-LR (10), MRI-ESM2-0 (2), NESM3 (2), NorESM2-LM (1), NorESM2-MM (1), TaiESM1 (1), UKESM1-0-LL (5).

For the annual-maxima and -minima approaches, the detrending is done on a yearly basis, but separately for all n-day averages, to account for different warming rates in n-day temperature extremes (e.g. a greater warming in Tx1day than in Tx90day):

$$\begin{aligned} \mathrm{Tx}n\mathrm{day}^{(y_1)}(y) &= \mathrm{Tx}n\mathrm{day}(y) - \beta_{\mathrm{Tx}n} \times (F(y) - F(y_1)) \\ \mathrm{Tn}n\mathrm{day}^{(y_1)}(y) &= \mathrm{Tn}n\mathrm{day}(y) - \beta_{\mathrm{Tn}n} \times (F(y) - F(y_1)) \end{aligned} \tag{A4}$$

with $\beta_{\mathrm{Tx}n}$ and $\beta_{\mathrm{Tn}n}$ scaling factors of Tx$n$day and Tn$n$day onto the forced response.

Two important remarks on this simple detrending procedure must be made. First, it only corrects for potential changes in the mean, not in the variance or higher moments; this is not a major problem for temperature because changes in the mean dominate the total signal, but a more sophisticated correction could be used, especially to extend the method to other variables. Second, as we allow the scaling factors β_T to be different from 1, we correct for possible mismatches between observations and models on the magnitude of long-term trends; in other words, our correction $\beta_{\mathrm{T}} \times F$ has the shape of the multi-model mean forced response but the amplitude of the observed trend.

APPENDIX B

Validation of the p_1 estimation procedure

Here we provide further elements of validation of the p_1 estimation procedure. First, we verify that the selected events are evenly distributed throughout the scanning period, and that no clear long-term trend is visible in the yearly minimum p_1 (Figure S4). This suggests that the long-term trend is reasonably accounted for in all regions. Note that El Nino (La Nina) years coincide with peaks in the number of regions having a record heat-wave (cold spell), as expected since a large number of regions are dominated by this mode of variability.

Second, we verify that the distributions of our p_1 estimates are close to uniform. Figure S5 shows the histograms of p_1 estimates for all events of all 2 Mkm^2 regions, for various temporal durations and the three (calendar, annual-maxima, annual-minima) methods. In all cases p_1 is rather uniformly distributed, suggesting that fitting normal distributions to calendar temperatures and GEV to annual-maxima and -minima is reasonable. In the calendar approach, slight departures from the uniform arise at the edges of the [0,1] interval, which could result for asymetric behaviors.

Finally, we show examples of problems encountered in a few cases when scanning the full period and all the regions. Importantly, there is no obvious evidence that the events listed in the main text or tables are concerned by such issues.

Figure S6a shows a time series of Tx20day in the 5 Mkm^2 region of east and south Brazil. The first 5 years of the series (1959–1963) appear well above the rest of the detrended series, which clearly points to a homogeneity problem. This event of October 1963 is selected by our procedure in this region and several child regions with pretty high rankings, but it is at least partly for wrong reasons.

Figure S6b illustrates another type of problem: the time series is Tn3day in Pacific Russia, for which the GEV fit gives a very negative shape parameter ($\xi = -0.7$). This very sharp "upper" tail — right side of the distribution since the GEV is reversed for annual minima — results in a very low p_1 associated with this January 1960 event, although its value does not look that extreme. This is partly because several years have "not so low" values, and encourages to increase block size (i.e. more than one year) to help estimating ξ and ensure a better GEV fit.

APPENDIX C

An interactive webpage to further explore the results

We have deployed a R shiny application to reproduce the Figures and Tables of the paper for all methods, years, and other variables (daily-maximum and -minimum temperatures): https://jlncttx.shinyapps.io/CRT23-app/.

The "Local scan" tab enables the user to explore the France data and redo Figures 1, 2 and Table 1 for any selection of variable, method and year.

The "Global scan" tab enables the user to explore results of the ERA5 scan and redo Figures 3 and 4 and Tables 2 for any selection of variable, method, year(s) and size of WRAF regions.

References

Angélil, O., D. A. Stone, M. Tadross, F. Tummon, M. Wehner, and R. Knutti, 2014: Attribution of extreme weather to anthropogenic greenhouse gas emissions: sensitivity to spatial and temporal scales. *Geophys. Res. Lett.*, **41 (6)**, 2150–2155, https://doi.org/10.1002/2014GL059234.

Angélil, O., and Coauthors, 2018: On the nonlinearity of spatial scales in extreme weather attribution statements. *Clim. Dyn.*, **50 (7)**, 2739–2752, https://doi.org/10.1007/s00382-017-3768-9.

Barlow, A. M., C. Sherlock, and J. Tawn, 2020: Inference for extreme values under threshold-based stopping rules. *Journal of the Royal Statistical Society Series C: Applied Statistics*, **69 (4)**, 765–789, https://doi.org/10.1111/rssc.12420.

Bercos-Hickey, E., T. A. O'Brien, M. F. Wehner, L. Zhang, C. M. Patricola, H. Huang, and M. D. Risser, 2022: Anthropogenic contributions to the 2021 Pacific Northwest heatwave. *Geophys. Res. Lett.*, e2022GL099396, https://doi.org/10.1029/2022GL099396.

Boettcher, M., M. Röthlisberger, R. Attinger, J. Rieder, and H. Wernli, 2023: The ERA5 extreme seasons explorer as a basis for research at the weather and climate interface. *Bull. Am. Meterol. Soc.*, **104 (3)**, E631–E644, https://doi.org/10.1175/BAMS-D-21-0348.1.

Cattiaux, J., and A. Ribes, 2018: Defining single extreme weather events in a climate perspective. *Bull. Am. Meterol. Soc.*, **99 (8)**, 1557–1568, https://doi.org/10.1175/BAMS-D-17-0281.1.

Geoffroy, O., D. Saint-Martin, D. J. L. Olivié, A. Voldoire, G. Bellon, and S. Tytéca, 2013: Transient climate response in a two-layer energy-balance model. Part I: Analytical solution and parameter calibration using CMIP5 AOGCM experiments. *J. Clim.*, **26 (6)**, 1841–1857, https://doi.org/10.1175/JCLI-D-12-00195.1.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.*, **146 (730)**, 1999–2049, https://doi.org/10.1002/qj.3803.

Kirchmeier-Young, M. C., H. Wan, X. Zhang, and S. I. Seneviratne, 2019: Importance of framing for extreme event attribution: The role of spatial and temporal scales. *Earth's Future*, **7 (10)**, 1192–1204, https://doi.org/10.1029/2019EF001253.

McKinnon, K. A., and I. Simpson, 2022: How Unexpected Was the 2021 Pacific Northwest Heatwave? *Geophys. Res. Lett.*, **49**, https://doi.org/10.1029/2022GL100380.

Otto, F. E. L., and Coauthors, 2020: Challenges to understanding extreme weather changes in lower income countries. *Bull. Am. Meterol. Soc.*, **101 (10)**, E1851–E1860, https://doi.org/10.1175/BAMS-D-19-0317.1.

Peterson, T. C., P. A. Stott, and S. Herring, 2012: Explaining extreme events of 2011 from a climate perspective. *Bull. Am. Meterol. Soc.*, **93 (7)**, 1041–1067, https://doi.org/10.1175/BAMS-D-12-00021.1.

Philip, S., and Coauthors, 2020: A protocol for probabilistic extreme event attribution analyses. *Adv. Stat. Clim. Meteorol. Oceanogr.*, **6 (2)**, 177–203, https://doi.org/10.5194/ascmo-6-177-2020.

Philip, S. Y., and Coauthors, 2022: Rapid attribution analysis of the extraordinary heat wave on the Pacific coast of the US and Canada in June 2021. *Earth Syst. Dynam.*, **13 (4)**, 1689–1713, https://doi.org/10.5194/esd-13-1689-2022.

Ribes, A., J. Boé, S. Qasmi, B. Dubuisson, H. Douville, and L. Terray, 2022: An updated assessment of past and future warming over France based on a regional observational constraint. *Earth Syst. Dynam.*, **13 (4)**, 1397–1415, https://doi.org/10.5194/esd-13-1397-2022.

Ribes, A., S. Thao, and J. Cattiaux, 2020: Describing the relationship between a weather event and climate change: a new statistical approach. *J. Clim.*, **33 (15)**, 6297–6314, https://doi.org/10.1175/JCLI-D-19-0217.1.

Rigal, A., J.-M. Azaïs, and A. Ribes, 2019: Estimating daily climatological normals in a changing climate. *Clim. Dyn.*, **53 (1)**, 275–286, https://doi.org/10.1007/s00382-018-4584-6.

Robin, Y., and A. Ribes, 2020: Nonstationary extreme value analysis for event attribution combining climate models and observations. *Adv. Stat. Clim. Meteorol. Oceanogr.*, **6 (2)**, 205–221, https://doi.org/10.5194/ascmo-6-205-2020.

Russo, S., J. Sillmann, and E. M. Fischer, 2015: Top ten European heatwaves since 1950 and their occurrence in the coming decades. *Environ. Res. Lett.*, **10 (12)**, 124 003, https://doi.org/10.1088/1748-9326/10/12/124003.

Röthlisberger, M., M. Hermann, C. Frei, F. Lehner, E. M. Fischer, R. Knutti, and H. Wernli, 2021: A new framework for identifying and investigating seasonal climate extremes. *J. Clim*, **34 (19)**, 7761–7782, https://doi.org/10.1175/JCLI-D-20-0953.1.

Schumacher, D. L., M. Hauser, and S. I. Seneviratne, 2022: Drivers and mechanisms of the 2021 Pacific Northwest heatwave. *Earth's Future*, **10 (12)**, e2022EF002 967, https://doi.org/10.1029/2022EF002967.

Seneviratne, S., and Coauthors, 2021: Weather and Climate Extreme Events in a Changing Climate. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, journal=Cambridge University Press pages=1513-–1766, doi=10.1017/9781009157896.013.

Stone, D. A., 2019: A hierarchical collection of political/economic regions for analysis of climate extremes. *Climatic Change*, **155 (4)**, 639–656, https://doi.org/10.1007/s10584-019-02479-6.

Stott, P. A., D. A. Stone, and M. R. Allen, 2004: Human contribution to the European heatwave of 2003. *Nature*, **432 (7017)**, 559–60, https://doi.org/10.1038/nature03089.

Terray, L., 2023: A storyline approach to the June 2021 northwestern North American heatwave. *Geophys. Res. Lett.*, **50 (5)**, e2022GL101 640, https://doi.org/10.1029/2022GL101640.

Thompson, V., D. Mitchell, G. C. Hegerl, M. Collins, N. J. Leach, and J. M. Slingo, 2023: The most at-risk regions in the world for high-impact heatwaves. *Nat. Commun.*, **14 (1)**, 2152, https://doi.org/10.1038/s41467-023-37554-1n.

Thompson, V., and Coauthors, 2022: The 2021 western North America heat wave among the most extreme events ever recorded globally. *Sci. Adv.*, **8 (18)**, eabm6860, https://doi.org/10.1126/sciadv.abm6860.

van der Wiel, K., F. M. Selten, R. Bintanja, R. Blackport, and J. Screen, 2020: Ensemble climate-impact modelling: extreme impacts from moderate meteorological conditions. *Environ. Res. Lett.*, **15**, 034 050, https://doi.org/10.1088/1748-9326/ab7668.

van Oldenborgh, G. J., and Coauthors, 2021a: Attribution of the Australian bushfire risk to anthropogenic climate change. *Natural Hazards and Earth System Sciences*, **21 (3)**, 941–960, https://doi.org/10.5194/nhess-21-941-2021.

van Oldenborgh, G. J., and Coauthors, 2021b: Pathways and pitfalls in extreme event attribution. *Climatic Change*, **166 (1-2)**, 13, https://doi.org/10.1007/s10584-021-03071-7.

Vautard, R., and Coauthors, 2022: Human influence on growing-period frosts like the early April 2021 in Central France. *Nat. Hazards Earth Syst. Sci.*, **2022**, 1–25, https://doi.org/10.5194/nhess-23-1045-2023.

White, R. H., and Coauthors, 2023: The unprecedented Pacific Northwest heatwave of June 2021. *Nat. Commun.*, **14 (1)**, 727, https://doi.org/10.1038/s41467-023-36289-3.