

DEFINING SINGLE EXTREME WEATHER EVENTS IN A CLIMATE PERSPECTIVE

JULIEN CATTIAUX AND AURÉLIEN RIBES

A generic road map for objectively defining a single extreme weather event is proposed.

Retrospective analysis of single weather events has been for a long time a regular activity of weather services. Commenting on and explaining a remarkable episode that is occurring or that has just occurred is part of climate monitoring. Determining return periods of particular extremes (e.g., those causing large socioeconomic impacts) is part of the mandatory prevention of natural hazards (<https://public.wmo.int/en/our-mandate/focus-areas/natural-hazards-and-disaster-risk-reduction>). More recently, single weather events have also received a growing attention from a climate change perspective. Weather services need to account for nonstationarity

in the return period estimates (Katz 2010; Cooley 2013). Climatologists interested in global warming attribution aim to quantify how human activities have affected the risk of occurrence of specific events (Stott et al. 2013; Trenberth et al. 2015; Otto et al. 2016), and dedicated annual reports have been published in *BAMS* since 2012 [e.g., Peterson et al. 2012, among others (www.ametsoc.org/ams/index.cfm/publications/bulletin-of-the-american-meteorological-society-bams/explaining-extreme-events-from-a-climate-perspective/)]. The result of such attribution studies is generally provided as a risk ratio (RR) or a fraction of attributable risk (FAR) given by

$$RR = \frac{p_1}{p_0} \text{ and } FAR = \frac{p_1 - p_0}{p_1} = 1 - \frac{1}{RR}, \quad (1)$$

where p_1 is the probability of the event occurring in the factual world (transposable into a return period) and p_0 is the probability of the event occurring in a counterfactual world without anthropogenic forcings (Stott et al. 2004). Both weather services and climatologists have developed a strong expertise in answering questions related to single extreme weather events, and even the attribution issue can now be addressed within a few days (e.g., van Oldenborgh et al. 2016).

AFFILIATIONS: CATTIAUX AND RIBES—Centre National de Recherches Météorologiques, Université de Toulouse, CNRS, Météo-France, Toulouse, France

CORRESPONDING AUTHOR: Julien Cattiaux, julien.cattiaux@meteo.fr

The abstract for this article can be found in this issue, following the table of contents.

DOI:10.1175/BAMS-D-17-0281.1

In final form 13 February 2018

©2018 American Meteorological Society

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).

A key prerequisite to all these considerations is the definition of the event itself. The question can be formulated very easily—for example, What was this extreme event?—but the answer involves many nontrivial choices. Despite an impressive number of single events analyzed over recent years, no systematic road map has been proposed to define the event to analyze, as particularly highlighted in the recent report of the National Academies of Science dedicated to attribution: “It would be useful to develop a set of objective event ... definition criteria” (National Academies of Sciences 2016, p. 15). In particular, the spatiotemporal scale is chosen arbitrarily in most studies. The authors either use predefined areas (e.g., a local station, a national territory) and periods (e.g., a day, a month, a season) or use the space–time characteristics that best depict the event and/or its impacts according to their own expertise. Defining a single extreme weather event through predetermined and/or subjective criteria can be useful to weather services or stakeholders, but is questionable from a physical perspective. First, this may not faithfully portray the event: meteorological phenomena do not conform to calendar or geopolitical divisions, while a subjective definition may be biased by our perception. Second, this may bring some confusion when putting an identified extreme event in a climate perspective. For instance, attribution results have been shown to be sensitive to the spatiotemporal scale (Otto et al. 2012; Uhe et al. 2016); although supposedly designating the same extreme event, different definitions can thus lead to different conclusions.

Here, we propose a straightforward and—as far as possible—objective procedure to define a single extreme weather event. In particular, we suggest an automatic way to designate the most relevant spatiotemporal scale for a particular event. We first illustrate our methodology on the European heat wave of summer 2003 (EHW2003), a well-documented event (Stott et al. 2004; Beniston 2004; Black et al. 2004; Schär et al. 2004; Cassou et al. 2005; Trigo et al. 2005; Chase et al. 2006; Barriopedro et al. 2011; Christidis et al. 2015). We also consider a precipitation event, the Boulder, Colorado, intense rainfall of September 2013 (BIR2013), which was responsible for severe floods in Colorado (Hoerling et al. 2014; Gochis et al. 2015; Eden et al. 2016; Pall et al. 2017).

THE FOUR STEPS OF EVENT DEFINITION.

First, one must decide which variable to consider. This choice is generally well constrained as it relates to the type of the natural hazard studied (e.g., heat wave, flooding, storm) and the purpose of the analysis (e.g.,

climate oriented vs impact oriented, both being perfectly valid). For instance, climate studies would use surface atmospheric temperature for heat waves and rainfall for heavy precipitation events, while, respectively, heat stress (McGregor et al. 2010; Fischer et al. 2012) and river discharge would be more relevant for impacts studies. Multivariable approaches may also be relevant for some applications (Sippel and Otto 2014); for instance, considering both daily minimal and maximal temperatures brings additional information for heat wave impacts. Here, we use single weather variables, since (i) our purpose is climate oriented and (ii) impact variables are more likely to be influenced by nonclimatic anthropogenic factors (e.g., a change in exposure, the construction of a river dam). We thus characterize EHW2003 through the daily mean surface temperature and BIR2013 through the daily amount of precipitation. Our method can nevertheless be applied to any other variable and/or generalized to a multivariable framework, enabling it to also be used for impact-oriented studies.

Second, it must be noted that a particular event has a null probability to happen exactly as it was: for a random variable X that has a continuous distribution and a given value x_0 , $\Pr\{X = x_0\} = 0$. The computation of the occurrence probability p_1 therefore requires one to specify a class of events, a topic that has been particularly discussed in the attribution community recently (Trenberth et al. 2015; Hannart et al. 2016; Otto et al. 2016; Shepherd 2016; Harrington 2017). The traditional approach (used, e.g., for the computation of return periods in climate monitoring) is to define the class as all events equally or more intense than the observed one. This approach is consistent with risk estimation and is commonly referred to as the “risk based” approach in attribution studies. In this case, p_1 corresponds to the tail of the distribution ($\Pr\{X \geq x_0\}$), which fits into the mathematical framework of extreme value theory (Coles 2001) and is highly relevant for most climate and impacts applications. Alternatively, it has been recently proposed to define the event as accurately as possible and to identify its causal chain of contributors in a deterministic way (Shepherd 2016). Scrutinizing an event in such a “storyline” perspective is indeed helpful for both climate monitoring and physical understanding (Hoerling et al. 2013). However, from a probabilistic point of view, narrowing the class of events to those of about the same intensity as the observed one ($\Pr\{x_0 - \varepsilon \leq X \leq x_0 + \varepsilon\}$ with $\varepsilon \rightarrow 0$) is inconsistent with return period estimation and could lead to misleading attribution statements, since every single cause becomes necessary to explain an event exactly as it was. We therefore use the traditional

return period approach here ($\Pr\{X \geq x_0\}$). We choose the intensity threshold x_0 as the value observed during the event itself; using less extreme values (e.g., as in Stott et al. 2004) would ease the probability estimation but not fully reflect the event rarity.

The third question is whether or not the event must be defined conditionally to (i) the concurrent state of the climate system (e.g., sea surface temperature, atmospheric circulation, soil moisture, El Niño phase) and/or (ii) the time of year. Technically, this amounts to narrowing the class of events by making probabilities conditional: $\Pr\{X \geq x_0\}$ becomes $\Pr\{X \geq x_0 \mid Y \in \Omega\}$ with Y the conditioning variable and Ω its observed state. Explaining a particular event from the perspective of other variables in the climate system can serve physical understanding (Cattiaux et al. 2010; Trenberth et al. 2015; Vautard et al. 2016), but the conditioning generally confuses climate change questions. Some natural hazards could indeed become less frequent given a particular atmospheric circulation pattern while becoming more frequent in general. This is why we do not use any conditioning to the concurrent climate state in this paper, though the proposed procedure easily extends to conditioned event definition if p_1 and p_0 are made conditional. The question of the timing of the event in the seasonal cycle is a bit different. Climate monitoring involves the description of events with respect to both annual maximum values (especially for the most extreme events) and the seasonal context. The annual maxima approach is particularly relevant for natural hazards whose impacts are similar regardless of the season, such as the risk of dike overflow by storm surges or river floods. By contrast, the calendar conditioning makes it possible to study weather events that are not particularly unusual at the annual scale but can cause specific impacts at the time they occur (winter heat waves, summer cold spells, tropical intense rainfall during the dry season, etc.). It must however be noted that limiting the study to a particular period of the year only provides a conditional p_1 , which should not be interpreted as the formal return period of the event. Here, we contrast the two approaches (annual maxima and calendar) for EHW2003 and only consider the annual maxima approach for BIR2013.

Last, one must define the spatiotemporal scale of the event. This choice might again be directly driven by the application. For instance, after a flood event, weather services mandated to compute the return period of the peak discharge at a given location would naturally focus at the local and instantaneous scale. But for climate monitoring or attribution studies, assessing whether the associated heavy rainfall event

should be considered at an hourly local or a daily regional scale can be less trivial. We further illustrate this issue in the following section on the basis of the EHW2003 event.

THE SPATIOTEMPORAL-SCALE ISSUE ILLUSTRATED FROM THE EUROPEAN HEAT WAVE OF SUMMER 2003.

In the first event attribution study, Stott et al. (2004) defined EHW2003 as a 3-month event [June–August (JJA)] over the so-called Mediterranean (MED) Giorgi region (Giorgi and Francisco 2000) covering southern Europe and the Mediterranean basin entirely. Choosing such a large and predefined scale was motivated by the wish to prevent what they referred to as a “selection bias” (see “Discussion and conclusions” section) and ensure the reliability of their climate model in simulating the temperature distribution. Depicting EHW2003 from this space–time window is, however, debatable given (i) the synoptic nature of heat waves and (ii) the fact that most EHW2003 impacts were reported for August and western Europe (i.e., at a much smaller spatiotemporal scale and a slightly different location; World Health Organization 2003). Alternatively, for climate monitoring over the French territory, Météo-France defined EHW2003 as an 18-day event on the basis of arbitrary temperature thresholds (www.meteofrance.fr/climat-passe-et-futur/impacts-du-changement-climatique-sur-les-phenomenes-hydrometeorologiques/changement-climatique-et-canicules). Such a 2-week duration appears particularly relevant for impact studies such as those focusing on human mortality (World Health Organization 2003; Haines et al. 2006; Robine et al. 2008). But EHW2003 could also be defined as a shorter and local event as some daily temperatures of early August were very extreme at some locations (Beniston 2004; Trigo et al. 2005).

EHW2003 therefore constitutes a typical case study for which the scale selection is nontrivial. It particularly matters when putting the event in the perspective of climate change. On average, over the whole season (JJA) and a large European domain, the temperature anomaly of EHW2003—2.5 K relative to 1961–90—gets overtaken by the median of phase 5 of the Coupled Model Intercomparison Project (CMIP5) representative concentration pathway 8.5 (RCP8.5) projections by 2040 and becomes a cold extreme by 2100 (Fig. 1a; see caption for data information). But in the same climate scenario, locally in Paris, France, daily temperatures observed in early August 2003 remain unusually high by 2100 for these calendar days (i.e., reaching their 99th percentiles), while the seasonal anomaly is slightly below normal

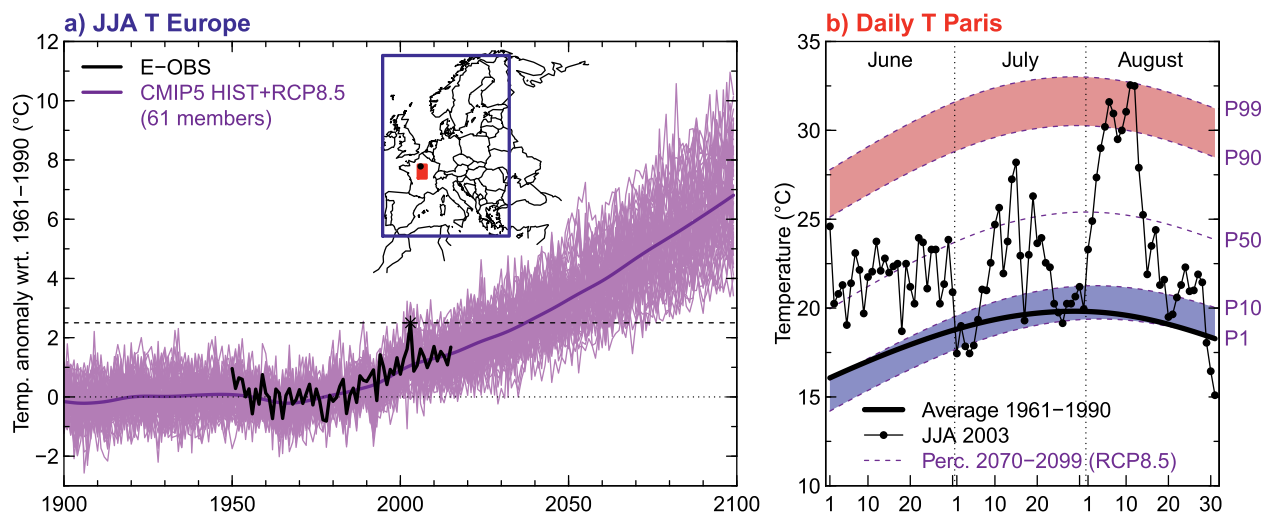


FIG. 1. (a) Observed (black) and simulated (violet) land surface temperature anomalies relative to 1961–90 averaged over JJA and the domain 35° – 70° N, 10° – 30° E (blue area on the map). Smoothed ensemble mean (thick violet) and 2003 observed anomaly (dashed black line) are added. **(b)** JJA 2003 observed daily temperature in Paris [red pixel on the map in (a)] compared with the 1961–90 observed climatology (thick black) and the 2070–99 projected ensemble-mean 1st, 10th, 50th, 90th, and 99th percentiles, respectively, noted P1, P10, P50, P90, and P99 on the right y axis [dashed violet, with blue (red) shading between P1 and P10 (P90 and P99)]. Paris observations have been provided by B. Dubuisson (Météo-France). European observations are taken from the E-OBS, version 13, gridded dataset (Haylock et al. 2008). CMIP5 projections are made of 61 concatenated historical and RCP8.5 experiments from 30 models (Taylor et al. 2012). E-OBS and CMIP5 data are both bilinearly interpolated onto a regular $2.5^{\circ} \times 2.5^{\circ}$ grid. List of CMIP5 models: ACCESS1.0 (1 member), ACCESS1.3 (1), BCC-CSM1.1 (1), BCC-CSM1.1(m) (1), BNU-ESM (1), CanESM2 (5), CCSM4 (6), CESM1(BGC) (1), CMCC-CESM (1), CMCC-CM (1), CMCC-CMS (1), CNRM-CM5 (5), CSIRO Mk3.6.0 (10), FGOALS-g2 (1), FGOALS-s2 (3), GFDL CM3 (1), GFDL-ESM2G (1), GFDL-ESM2M (1), GISS-E2-R (1), INM-CM4.0 (1), IPSL-CM5A-LR (4), IPSL-CM5A-MR (1), IPSL-CM5B-LR (1), MIROC5 (3), MIROC-ESM (1), MIROC-ESM-CHEM (1), MPI-ESM-LR (3), MPI-ESM-MR (1), MRI-CGCM3 (1) and NorESM1-M (1).

(Fig. 1b). Therefore, depending on whether EHW2003 is considered at seasonal regional, seasonal local, or daily local scale, it can, respectively, be interpreted as an extremely cold, a slightly cold, or an unusually hot event in late-century climate projections. Similarly, in the present-day climate, different probabilities of occurrence p_1 will be obtained for different scales. One thus needs to select the scale that best reflects the event.

SELECTING THE SPATIOTEMPORAL SCALE THAT MAXIMIZES THE RARITY.

Extreme events are, by nature, rare. This is precisely why they receive public or scientific attention and cause large impacts. In mathematical terms, they are thus inherently characterized by a small probability of occurrence p_1 . This probability can even be considered as the “mugshot” of the event: it portrays the event and quantifies how extreme it has been. As noted above, p_1 varies with the scale at which the extreme event is considered. Here, we argue that the value that best portrays the extreme feature of the event is the minimum p_1 . Therefore, we suggest that the last step of event definition, the selection of the

spatiotemporal scale, could be made by minimizing p_1 over all possible space–time windows. Again, this optimization is not needed for all applications, and predetermined space–time windows remain relevant for studies focusing on a particular population center, political region, or set of infrastructure.

Defining the event on the basis of p_1 translates the question of what this extreme event was into what the most extreme in this extreme event was, which, beyond the definition issue, is a legitimate question for climate monitoring. For attribution studies, our suggestion does not require additional work since estimating p_1 is requested to calculate RR or FAR. As p_1 follows a uniform distribution between 0 and 1 for each spatiotemporal scale—this is an intrinsic property¹ of an occurrence probability as defined

¹ For a random variable X , its cumulative density function F_X (assumed continuous), and $Z = F_X(X)$, the random variable corresponding to the probability of being under a given value (i.e., $1 - p_1$), one can write $\forall x, F_Z(x) = \Pr\{F_X(X) < x\} = \Pr\{X < F_X^{-1}(x)\} = F_X[F_X^{-1}(x)] = x$, which proves that Z (thus p_1) follows a uniform distribution.

above ($\Pr\{X \geq x_0\}$)—it provides a fair and unbiased comparison across scales.

Back to the EHW2003 event, our suggestion would translate into minimizing the probability that an event equally or more intense occurs in 2003; that is,

$$p_1 = \Pr\{X^{(t_1)} \geq x_{t_1}\}, \quad (2)$$

where $X^{(t_1)}$ is the random variable of daily mean temperature describing all possible realizations of temperature at time $t_1 = 2003$ and x_{t_1} is the value effectively observed that year. We show in the following that the minimum p_1 can be obtained in an automatic way. For the sake of clarity, we use here rather simple estimation methods for p_1 that are realistic enough to communicate effectively our main point but should not be blindly adopted. Our algorithm remains usable with any estimation method, including more sophisticated techniques (beyond our scope).

Let us first consider the temporal aspect only, setting the location to the Paris weather station. For each day d of JJA 2003 (92 days) and each duration n from 1 to 92 days, we compute the occurrence probability p_1 of having a temperature greater than or equal to the one observed over the n -day time window centered on d . Then we select the minimum of the 92×92 p_1 values, which provides the time window that maximizes EHW2003 rarity at Paris weather station. For computational reasons, we only consider an arbitrary subset of 24 durations. We empirically estimate p_1 by fitting a distribution on a sample of 66 yearly values x_t (corresponding to observed temperatures over 1950–2015) that we correct for climate change beforehand—indeed, the temperature observed in 1950 is not directly comparable to that of 2015 since the climate has changed in the mean time. Three different options are taken for the distribution and the sample used, corresponding to different choices of calendar conditioning [i.e., different definitions of $X^{(t_1)}$ and x_{t_1} in Eq. (2)]. As for the climate change correction, we assume that only the mean of the temperature distribution has changed (variance and shape are kept constant) and that the change is uniform throughout the summer season (the same correction is applied for all time windows throughout JJA). Again, such deliberately strong hypotheses are reasonable for our purpose but should not be blindly adopted. The detrending involves two steps:

- 1) The climate change signal x_t^* is estimated from a 10-degrees-of-freedom (df) spline smoothing (arbitrary choice) of the multimember mean of CMIP5 JJA temperatures. Each grid point is

treated separately, and trends are then averaged over the spatial domain of interest. For the grid point of Paris, the estimated change is about +1.5 K between 1950 and 2015.

- 2) The sample of 66 raw values x_t is translated into a sample $x_t^{(t_1)} = x_t - (x_t^* - x_{t_1}^*)$, which is representative of the climate of the year of the event ($t_1 = 2003$ in our case). In other words, we correct for climate change before and after t_1 , keeping x_{t_1} constant [$x_{t_1}^{(t_1)} = x_{t_1}$].

Finally, note that we always include the value corresponding to the event itself x_{t_1} in our sample; removing it could be regarded as ignoring available information and thus biasing the estimation of p_1 .

The first option—the calendar approach—consists of comparing the temperature observed in 2003 for a given time window with the temperature distribution at the exact same calendar time window. The sample x_t therefore corresponds to the 1950–2015 temperatures observed over the n -day calendar window of interest, and the distribution of $X^{(t_1)}$ is assumed to be Gaussian (Fig. 2). Within this calendar approach, which is common for climate monitoring (www.meteofrance.fr/climat-passe-et-futur/bilans-climatiques/bilan-2017/bilan-climatique-de-l-ete-2017; www.metoffice.gov.uk/climate/uk/interesting/hot-spell-june-2017), most

T Paris pdf for June 1

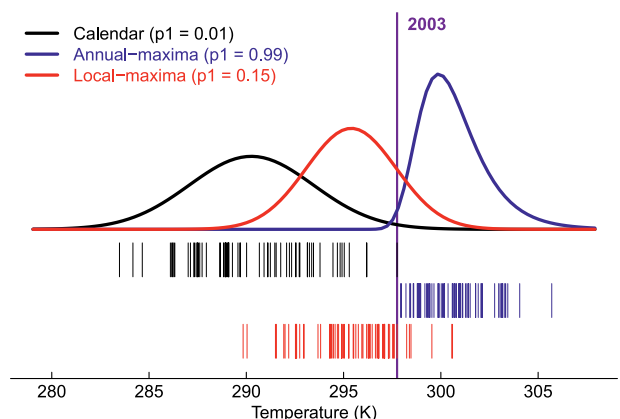


FIG. 2. Illustration of the three methods used to estimate p_1 for EHW2003 in the case of a 1-day time window (1 Jun) at the Paris location. The temperature observed in 2003 x_{t_1} (violet) is compared with the Gaussian distribution of all 1 Jun temperatures (calendar, black), the Gumbel distribution of annual 1-day temperature maxima (annual maxima, blue), and the Gaussian distribution of yearly 1-day temperature maxima within the “1 Jun \pm one week” time window (local maxima, red). All distributions (lines) are empirically fitted on the 66 detrended values of the period 1950–2015 $x_t^{(t_1)}$ (vertical bars), and estimated p_1 values are given.

of summer 2003 appears abnormally warm, with the exception of early/late July and late August (Fig. 3a). Early August is particularly extreme, and the lowest p_1 —or highest rarity—is found for the 8-day time window of 5–12 August ($p_1 = 4 \times 10^{-6}$). To a lesser extent, the entire summer (JJA), as well as both June and August months, is also found to be extremely hot ($p_1 = 5 \times 10^{-4}$), consistent with the literature (Cassou et al. 2005; Trigo et al. 2005; Barriopedro et al. 2011).

Our calendar approach, however, prevents us from interpreting p_1 in terms of return period. A 0.01 probability on 1 June means that such a warm day would occur on average once in 100 years on 1 June, but it would certainly occur much more frequently anytime in the year. A second option for estimating p_1 , the annual maxima approach, is to fit the traditional return period framework by replacing the n -day calendar Gaussian distributions with n -day annual maxima Gumbel distributions (Fig. 2). Interestingly, almost the same time window is found to minimize p_1 (4–12 August), albeit the gap with the season-scale event is reduced ($p_1 = 0.008$ vs $p_1 = 0.020$; Fig. 3b). The annual maxima approach provides conventional return periods but annihilates the rarity of events distant from the annual cycle peak; for instance, June 2003 is not unusual any more.

As a compromise, we consider a third definition for p_1 , the local maxima approach: we compare a given day d (or n -day time window centered on d) of 2003 with the Gaussian distribution of yearly maxima located in its calendar neighborhood $d \pm k$ (Fig. 2). Taking $k = 0$ equates to the calendar approach, and $k = 365/2$ is almost identical to the annual maxima; k should therefore be chosen as the typical duration over which the amplitude of the seasonal cycle remains small relative to the intraseasonal variability. Here, we arbitrarily choose $k = 7$, so a 0.01 probability on 1 June means that such a warm day would occur on average once in 100 years on 1 June \pm one week. This 2-week calendar neighborhood seems relevant for European temperatures, although a precise optimization of k is beyond our scope. With this method, the highest rarity is found for the 11-day time window of 3–13 August, and the month of June is found to be unusual even if individual days are not (Fig. 3c).

The spatial dimension can be incorporated by repeating the above procedure for an ensemble of possible spatial domains and searching for the global minimum of p_1 over all space–time windows. In our case, we consider all rectangular domains of size $m \times n$ grid points that (i) encompass Paris (48.5°N, 2.2°E), (ii) are included in the whole European domain (35°–70°N, 10°W–40°E; size 21×15), and (iii) contain at least

50% of continental grid points. For computational reasons, we limit ourselves to squared or near-squared domains. We incorporate the local Paris weather station as the 0×0 dimension. For both annual and local maxima approaches, the highest rarity is found for an early August event (11–12 days) over France and Spain (size 7×5 ; Figs. 3d–f). This p_1 minimum also arises in the calendar approach but is overtaken by a smaller-scale event. Importantly, all methods agree that EHW2003 is less extreme when considered over the whole season and domain (i.e., the space–time characteristics retained in several attribution studies; Stott et al. 2004; Christidis et al. 2015).

MAXIMIZING THE RARITY DOES NOT MAXIMIZE (OR MINIMIZE) THE ATTRIBUTABLE RISK. Searching for the spatiotemporal scale at which a single extreme weather event has been the most extreme (minimum p_1) is an academic question that is relevant for climate monitoring. It can, however, appear disconcerting for attribution studies at first sight: as p_1 is directly involved in the RR and FAR [see Eq. (1)], minimizing p_1 is likely to impact attribution results. Here, we show that selecting the scale that minimizes p_1 does have implications for attribution results but does not systematically bias the RR or FAR toward high or low values.

In our EHW2003 example, a rough estimation of the FAR can be obtained by computing p_0 as the probability that an event equally or more intense than EHW2003 occurs not in 2003 but at the beginning of our period of study ($t_0 = 1950$):

$$p_0 = \Pr\{X^{(t_0)} \geq x_t^*\}. \quad (3)$$

In practice, p_0 is estimated from the sample $x_t^{(t_0)} = x_t - (x_t^* - x_{t_0}^*)$, that is, derived from the 1950–2015 observed temperatures by applying the climate change correction described above relative to $t_0 = 1950$ rather than $t_1 = 2003$. The European climate of 1950 $X^{(t_0)}$ is colder than that of 2003 $X^{(t_1)}$, so $p_0 < p_1$ and $\text{FAR} > 0$. We obtain higher FAR values (between 75% and 100%) with calendar conditioning than without (between 30% and 90%; Figs. 3g–i). We consider that these estimates are consistent with the existing literature (Stott et al. 2004; Christidis et al. 2015) although our estimation method for p_0 and p_1 is deliberately very simple.

The interesting point is that, contrary to the rarity, the FAR of the EHW2003 event clearly increases with the spatiotemporal scale for all three approaches (Figs. 3g–i). This is consistent with Uhe et al. (2016), who show that the RR of another hot event—the

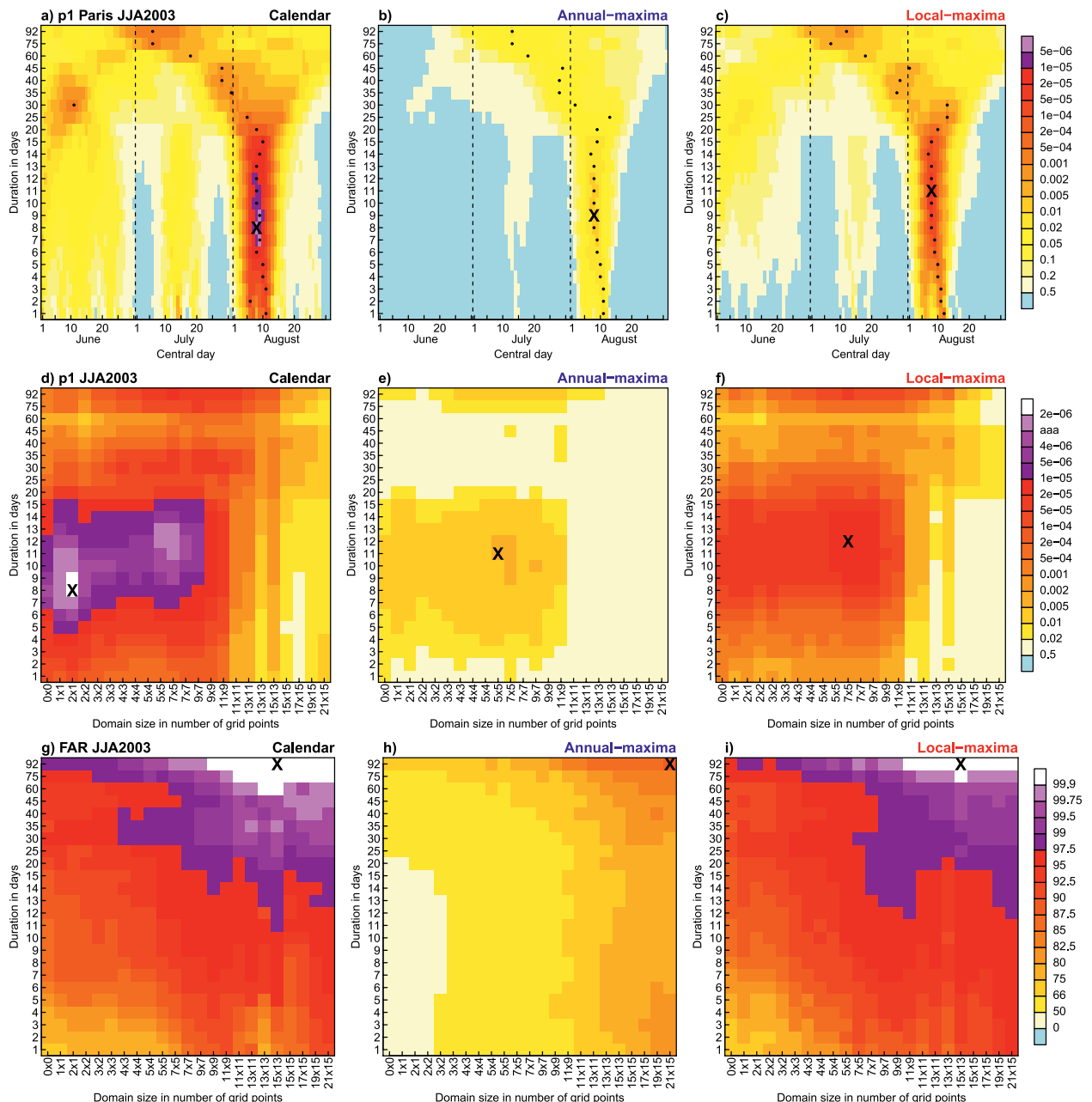


FIG. 3. (a)–(c) Estimated EHW2003 p_i at Paris location for time windows of various durations (days, y axis) centered over days from 1 Jun to 31 Aug (x axis) for the three estimation methods (calendar, annual maxima, and local maxima). Dots (crosses) indicate minima for every duration (overall minima). (d)–(f) Minimum EHW2003 p_i for various spatial sizes (number of grid points, x axis) and various temporal durations (days, y axis) for the three estimation methods. Crosses indicate overall minima: 5–12 Aug over 47.5°N, 0°–2.5°E; 3–13 Aug over 37.5°–52.5°N, 7.5°W–2.5°E; and 2–13 Aug over 40°–50°N, 7.5°W–7.5°E, respectively. Leftmost columns (size 0 × 0) correspond to Paris location [i.e., dots in (a)–(c)]. (g)–(i) FAR values associated with the p_i plotted in (d)–(f) (%). Crosses indicate overall maxima.

European record-breaking yearly temperatures of 2014—increases with the domain size. In fact, this is not surprising given that the RR or FAR directly responds to the signal-to-noise ratio of the human-induced change. For temperatures, the warming signal is rather uniform across scales: at the first

order, the whole distribution shifts toward a warmer climate, and daily local temperatures are affected similarly as seasonal regional ones. By contrast, the noise of temperature variability is highly nonuniform: it is stronger (weaker) for small (large) spatiotemporal scales. The signal-to-noise ratio of temperature events

therefore increases with the scale (as also evidenced in Fig. 1) and so does the RR or FAR.

Attribution studies providing quantitative risk-based results (e.g., RR or FAR), for a specific extreme event, can therefore be biased if they use a space–time window that does not portray the event supposedly analyzed. Here, the term *bias* does not refer to the RR or FAR calculation made in such studies, which can be correct, but to the potential mismatch between the definition used and the targeted event. In other words, all FAR values in Figs. 3g–i are scientifically valid, including those obtained for the seasonal regional event [comparable to Stott et al. (2004) and Christidis et al. (2015)]; however, we argue that not all should be interpreted as the FAR of the EHW2003 event. To take an exaggerated example, one could perfectly compute the FAR of the global and annual mean temperature in 2003, but it would be unfair to consider it as the FAR of the EHW2003 event. For temperature extremes, using a too-large space–time window to define a single event can inflate the RR or FAR ascribed to that event.

An objective definition procedure such as that proposed here may therefore guide the monitoring and attribution of single weather events. Interestingly, Fig. 3 clearly illustrates that defining an extreme event from the spatiotemporal scale that maximizes its rarity does not maximize (or minimize) its RR or FAR.

A PRECIPITATION CASE STUDY: THE BOULDER INTENSE RAINFALL OF SEPTEMBER 2013. The signal-to-noise behavior of temperatures should however not apply to all climate variables. For instance, the response of the distribution of precipitation to climate change is more complex than a shift: changes in mean precipitation are rather small and depend on the location, while extremes are expected to undergo a more robust and spatially generalized increase (Kharin et al. 2013). Therefore, contrary to temperature, both signal (climate change) and noise (variability) vary with scale for precipitation, possibly with complex consequences for measures like RR or FAR.

We illustrate this point by applying our definition road map to the BIR2013 event. All the three approaches used for EHW2003 could be applied, but since the calendar conditioning is less relevant for precipitation extremes (smaller seasonal cycle), here we only consider the annual maxima approach for simplicity. We now use Eq. (2) with $X^{(t_i)}$ being the random variable for rainfall annual maxima, $t_1 = 2013$, and x_{t_i} the value observed during BIR2013. As for EHW2003, we empirically estimate p_1 from 1901 to 2014 detrended series of observations (Fig. 4a;

see caption for data information). However, the detrending procedure differs, since climate change does not affect the precipitation distribution as it affects temperatures. We proceed in two steps. First, as for EHW2003, we estimate the local long-term warming from the CMIP5 ensemble; we obtain about 1.6 K between 1901 and 2014 at the Boulder grid point. Second, we estimate the change in n -day precipitation extremes relative to the local warming ($\% K^{-1}$) from the CMIP5 ensemble;² the scaling is assumed to be spatially uniform over Colorado and ranges from 2.5% K^{-1} for 1-day annual maxima to 0.7% K^{-1} for 92-day annual maxima. Such a small increase in precipitation extremes in this region—typically between 0 and the Clausius–Clapeyron rate—is consistent with the literature (Hoerling et al. 2014; Eden et al. 2016). Within this procedure, a stronger climate change signal is diagnosed for short-duration precipitation events. Ultimately, we compute p_1 assuming that $X^{(t_i)}$ follows a generalized extreme value (GEV) distribution with a shape parameter $\xi = 0.1$ (Fig. 4a), which is a convenient value in our case as it is consistent across all space–time windows tested for BIR2013. Using a constant ξ ensures that the probability to exceed a given amount of rainfall (mm) increases with time duration. We compute p_0 and the FAR similarly, taking $t_0 = 1901$.

Exploring all time windows within August–October 2013 and all domains from the Boulder station (40°N, 105°W) to Colorado (37°–41°N, 102°–109°W), we find that the space–time window that maximizes the rarity of the BIR2013 event is the 5-day period of 11–15 September at local scale ($p_1 = 7 \times 10^{-5}$; Figs. 4b and 4c). This might not be surprising given that for 4 of these 5 days, the observed daily amount of rainfall exceeds the median of the 1-day annual maxima distribution. This 5-day local scale happens to be the definition of BIR2013 used in existing literature (Hoerling et al. 2014; Eden et al. 2016; Pall et al. 2017). Given the small increase in precipitation extremes derived from CMIP5 models (between 0.7% and 2.5% K^{-1} , depending on the duration), the obtained FAR is only slightly positive for this event, typically between 10% and 25% (Fig. 4d). Given the large uncertainties associated with our simple computation procedure of p_1 and p_0 , we consider that this estimate is consistent with previous analysis (Hoerling et al. 2014; Eden et al.

² Scaling estimated from an exponential fit: $\log(P) = \alpha T + \beta$, with P the series of annual n -day precipitation maxima and T the smoothed series of yearly temperatures. An estimation is made for each model separately and then averaged over the ensemble.

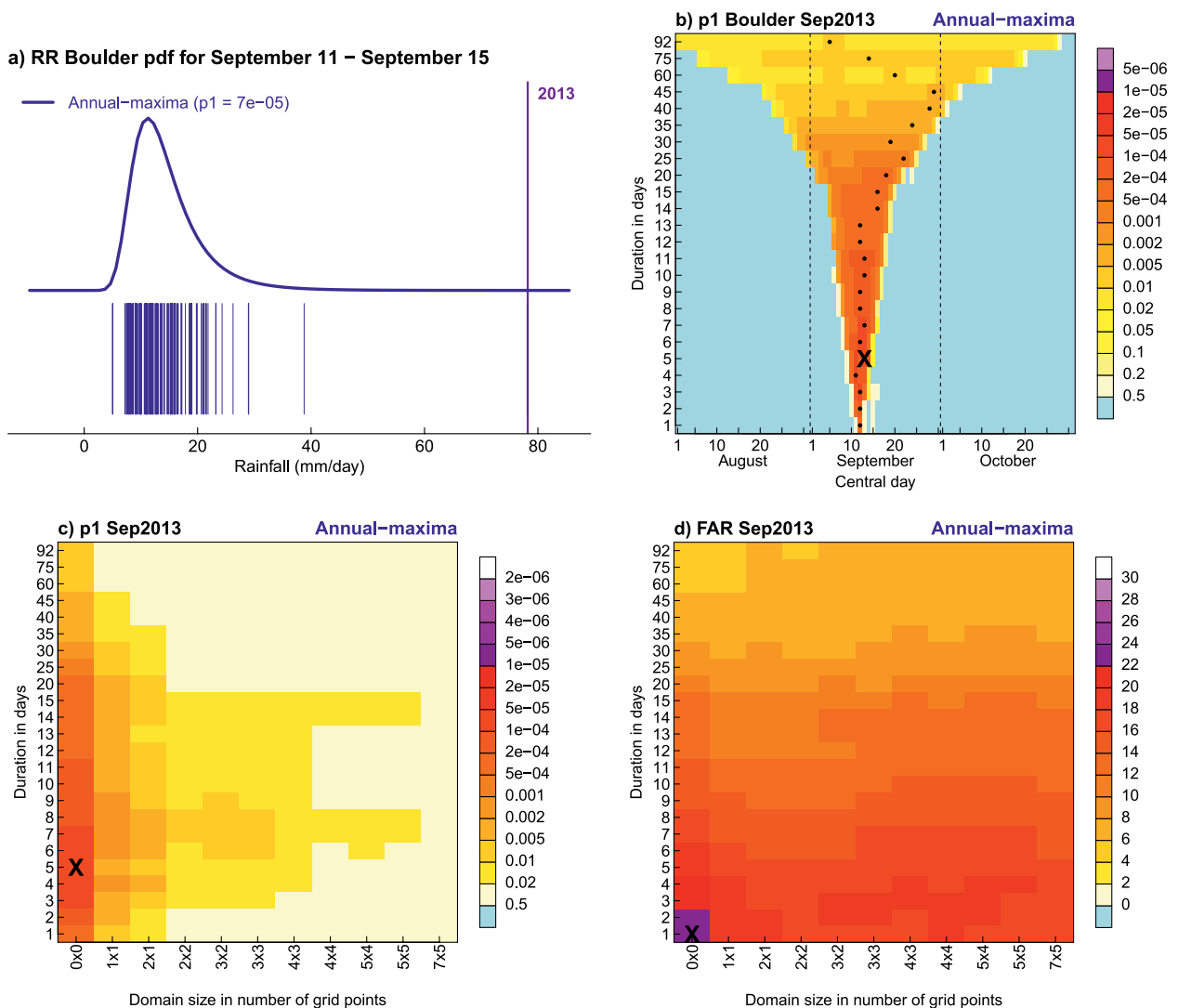


FIG. 4. As in Fig. 3, but for the BIR2013 event. (a) The rainfall observed during 11–15 Sep 2013 x_t (violet) is compared with the GEV distribution of 5-day rainfall annual maxima (blue), empirically fitted on the 114 detrended values of the period 1901–2014 $x_t^{(t)}$ (vertical bars). The estimated p_1 is given. (b) Estimated BIR2013 p_1 at Boulder location for all time windows. (c) Minimum BIR2013 p_1 for all spatial sizes and temporal durations. (d) Associated FAR (%). Boulder observations are taken from the corresponding station in the Global Historical Climatology Network (GHCN)-Daily dataset (Menne et al. 2012). Gridded data at 1° resolution have been created from GHCN-Daily and provided by M. Hoerling (Hoerling et al. 2014).

2016). Interestingly, the FAR is found to be larger for short-duration events, meaning that, in this case, the signal-to-noise ratio of climate change is dominated by the behavior of the signal. Beyond this example, our analysis shows that for precipitation events as well, the choice of the space–time characteristics can substantially affect attribution results. But again, maximizing the rarity of this event does not pull its RR or FAR toward particularly high or low values.

DISCUSSION AND CONCLUSIONS. The aim of this study is to demonstrate that an as-objective-as-possible definition road map for single extreme

weather events may guide their monitoring and attribution to climate change. Our main suggestion is to select the spatiotemporal scale in an automatic way, by minimizing the probability of occurrence p_1 of the event over all possible space–time windows. Here, we discuss some of the potential shortcomings associated with our suggestion.

Searching for the minimum p_1 can complicate the estimation of p_1 itself (and p_0 if needed), as it possibly requires venturing far in the tails of the distribution. However, dealing with small probabilities is inherent to the analysis of extreme events, and mathematical tools exist (e.g., the extreme value theory; Coles 2001)

to cope with distribution tails and enable statistical inference on rare values. Further, the most extreme events are typically those causing the largest impacts and/or drawing the most public and scientific attention. To us, it seems paradoxical to be willing to analyze very extreme events but to intentionally use a definition that makes them less extreme (e.g., through threshold or spatiotemporal scale selection) for the only reason that it makes the estimation of relevant quantities (p_1 and p_0) easier.

This decision is nevertheless regularly made in attribution studies using large ensembles of simulations to empirically estimate p_1 and p_0 [e.g., in the pioneer study of Stott et al. (2004), among many others]. In this approach, using a “too extreme” event definition can result in no simulation reproducing the event, which translates into estimated probabilities $\hat{p}_1 = \hat{p}_0 = 0$, so no conclusion can be drawn on the RR or FAR. We argue that, in this case, revising the estimation method (e.g., using extreme-value theory rather than empirical frequencies) seems more appropriate than revising the event definition. Further, we consider that the lack of conclusion due to uncertainties in RR or FAR is not necessarily an issue and that, above all, single extreme events should not be defined in order to provide a particular response to the attribution question.

Another criticism that could potentially be made to our suggestion of minimizing p_1 is related to what some authors of attribution studies refer to as a “selection bias.”³ For instance, Stott et al. (2004, p. 610) explicitly chose a predefined region to define EHW2003 “in order to minimize any bias that could result from selecting [the] region already knowing where the most extreme temperatures occurred.” In our view, there is no bias in selecting the scale that best portrays the extreme feature of an event; searching for a minimum value (here, for p_1) does not make it biased as long as the estimation is correct. Further, for attribution purposes, minimizing p_1 does not induce any systematic bias on the RR or FAR, as evidenced in Figs. 3 and 4.

³ The term “selection bias” seems to have been used in different ways in the event attribution community. It sometimes refers to the fact that events affecting developed countries are more likely to be studied than others, therefore designating a “geographical selection bias.” In the subsequent citation of Stott et al. (2004), this term is used with a different meaning, suggesting a “statistical selection bias”; the latter, however, has not been properly defined to the best of our knowledge. In particular, if the word “bias” denotes an error relative to an “expectation,” it is not clear what the expected value for p_1 should be.

In fact, this is the opposite: Fig. 3 highlights that using a too-large space–time window to depict EHW2003 can inflate the RR or FAR ascribed to that particular event and thus bias—at least quantitatively—the attribution statement.

Overall, p_1 appears as an objective way to select the spatiotemporal scale associated to a single extreme weather event. It can also help to compare the rarity of different events, for example, two heat waves occurring in different years and/or at different locations. In particular, it could be used to objectively designate the most extreme events within a year for BAMS dedicated annual reports and therefore alleviate the often-reported geographical selection bias (National Academies of Sciences 2016). In this study, we acknowledge that our estimation procedures for p_1 , p_0 , and the FAR are rather simple. More sophisticated techniques could be used, including more complex distributions and/or formal detection and attribution methods to properly estimate the climate change signal. Accounting for the uncertainties in the estimation of p_1 , p_0 , and their ratio (RR or FAR) would also constitute an important improvement; in particular, considering confidence intervals for p_1 would provide the ensemble of space–time windows for which p_1 is statistically consistent with the minima found in Figs. 3 and 4. As (i) our main objective was to advertise the relevance of p_1 to define events, and (ii) our overall procedure would be usable with any estimation method for p_1 (including confidence intervals), we leave such potential improvements for future work. Besides, while we only focus on two illustrations in this piece (EHW2003 and BIR2013), our generic road map can be applied to analyze any extreme weather event from both climate and impacts perspectives.

ACKNOWLEDGMENTS. The authors thank two anonymous reviewers for their comments that helped improve the manuscript. They are grateful to B. Dubuisson (Météo-France) for providing the Paris temperature data and M. Hoerling (NOAA) for providing the U.S. gridded rainfall data. They also acknowledge the E-OBS dataset from the ECA&D project as well as providers of the GHCN-Daily dataset. They thank the climate modeling groups involved in CMIP5 for producing and making available their simulations. They are indebted to S. Tyteca for data handling at CNRM and H. Douville (CNRM), E. Fischer (ETHZ), A. Jézéquel (LSCE), P. Naveau (LSCE), D. Nychka (NCAR), P. Pall (Berkeley), M. Schneider (Météo-France), and R. Vautard (LSCE) for helpful discussions. The figures and analyses were produced with the R software (<https://cran.r-project.org>). J. C. is supported by CNRS; A. R. is supported by Météo-France.

REFERENCES

- Barriopedro, D., E. Fischer, J. Luterbacher, R. Trigo, and R. García-Herrera, 2011: The hot summer of 2010: Redrawing the temperature record map of Europe. *Science*, **332**, 220–224, <https://doi.org/10.1126/science.1201224>.
- Beniston, M., 2004: The 2003 heat wave in Europe: A shape of things to come? An analysis based on Swiss climatological data and model simulations. *Geophys. Res. Lett.*, **31**, L02202, <https://doi.org/10.1029/2003GL018857>.
- Black, E., M. Blackburn, G. Harrison, B. 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>.
- Cassou, C., L. Terray, and A. Phillips, 2005: Tropical Atlantic influence on European heat waves. *J. Climate*, **18**, 2805–2811, <https://doi.org/10.1175/JCLI3506.1>.
- Cattiaux, J., 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>.
- Chase, T. N., K. Wolter, R. A. Pielke Sr., 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>.
- Christidis, N., G. Jones, and P. Stott, 2015: Dramatically increasing chance of extremely hot summers since the 2003 European heatwave. *Nat. Climate Change*, **5**, 46–50, <https://doi.org/10.1038/nclimate2468>.
- Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer, 208 pp.
- Cooley, D., 2013: Return periods and return levels under climate change. *Extremes in a Changing Climate: Detection, Analysis and Uncertainty*, A. AghaKouchak et al., Eds., Springer, 97–114, https://doi.org/10.1007/978-94-007-4479-0_4.
- Eden, J., K. Wolter, F. Otto, and G. van Oldenborgh, 2016: Multi-method attribution analysis of extreme precipitation in Boulder, Colorado. *Environ. Res. Lett.*, **11**, 124009, <https://doi.org/10.1088/1748-9326/11/12/124009>.
- Fischer, E., K. Oleson, and D. Lawrence, 2012: Contrasting urban and rural heat stress responses to climate change. *Geophys. Res. Lett.*, **39**, L03705, <https://doi.org/10.1029/2011GL050576>.
- Giorgi, F., and R. Francisco, 2000: Evaluating uncertainties in the prediction of regional climate change. *Geophys. Res. Lett.*, **27**, 1295–1298, <https://doi.org/10.1029/1999GL011016>.
- Gochis, D., and Coauthors, 2015: The great Colorado flood of September 2013. *Bull. Amer. Meteor. Soc.*, **96**, 1461–1487, <https://doi.org/10.1175/BAMS-D-13-00241.1>.
- Haines, A., R. Kovats, D. Campbell-Lendrum, and C. Corvalán, 2006: Climate change and human health: Impacts, vulnerability and public health. *Public Health*, **120**, 585–596, <https://doi.org/10.1016/j.puhe.2006.01.002>.
- Hannart, A., A. Carrassi, M. Bocquet, M. Ghil, P. Naveau, M. Pulido, J. Ruiz, and P. Tandeo, 2016: DADA: Data assimilation for the detection and attribution of weather and climate-related events. *Climatic Change*, **136**, 155–174, <https://doi.org/10.1007/s10584-016-1595-3>.
- Harrington, L., 2017: Investigating differences between event-as-class and probability density-based attribution statements with emerging climate change. *Climatic Change*, **141**, 641–654, <https://doi.org/10.1007/s10584-017-1906-3>.
- Haylock, M., N. Hofstra, A. Klein Tank, E. Klok, P. Jones, and M. New, 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J. Geophys. Res.*, **113**, D20119, <https://doi.org/10.1029/2008JD010201>.
- Hoerling, M., and Coauthors, 2013: Anatomy of an extreme event. *J. Climate*, **26**, 2811–2832, <https://doi.org/10.1175/JCLI-D-12-00270.1>.
- , and Coauthors, 2014: Northeast Colorado extreme rains interpreted in a climate change context [in “Explaining Extreme Events of 2013 from a Climate Perspective”]. *Bull. Amer. Meteor. Soc.*, **95** (9), S15–S18, <https://doi.org/10.1175/1520-0477-95.9.S1.1>.
- Katz, R., 2010: Statistics of extremes in climate change. *Climatic Change*, **100**, 71–76, <https://doi.org/10.1007/s10584-010-9834-5>.
- Kharin, V., F. Zwiers, X. Zhang, and M. Wehner, 2013: Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, **119**, 345–357, <https://doi.org/10.1007/s10584-013-0705-8>.
- McGregor, G., P. Bessemoulin, K. Ebi, and B. Menne, Eds., 2010: Heatwaves and health: Guidance on warning system development. WMO Rep. 1142, 114 pp., www.who.int/globalchange/publications/WMO_WHO_Heat_Health_Guidance_2015.pdf.
- Menne, M., I. Durre, R. Vose, B. Gleason, and T. Houston, 2012: An overview of the global historical climatology network-daily database. *J. Atmos. Oceanic Technol.*, **29**, 897–910, <https://doi.org/10.1175/JTECH-D-11-00103.1>.
- National Academies of Sciences, 2016: *Attribution of Extreme Weather Events in the Context of Climate Change*. The National Academies Press, 186 pp., <https://doi.org/10.17226/21852>.
- Otto, F., N. Massey, G. van Oldenborgh, R. Jones, and M. Allen, 2012: Reconciling two approaches to

- attribution of the 2010 Russian heat wave. *Geophys. Res. Lett.*, **39**, L04702, <https://doi.org/10.1029/2011GL050422>.
- , G. van Oldenborgh, J. Eden, P. Stott, D. Karoly, and M. Allen, 2016: The attribution question. *Nat. Climate Change*, **6**, 813–816, <https://doi.org/10.1038/nclimate3089>.
- Pall, P., C. M. Patricola, M. F. Wehner, D. A. Stone, C. J. Paciorek, and W. D. Collins, 2017: Diagnosing conditional anthropogenic contributions to heavy Colorado rainfall in September 2013. *Wea. Climate Extremes*, **17**, 1–6, <https://doi.org/10.1016/j.wace.2017.03.004>.
- Peterson, T., P. Stott, and S. Herring, 2012: Explaining extreme events of 2011 from a climate perspective. *Bull. Amer. Meteor. Soc.*, **93**, 1041–1067, <https://doi.org/10.1175/BAMS-D-12-00021.1>.
- Robine, J., S. Cheung, S. Le Roy, H. van Oyen, C. Griffiths, J. Michel, and F. 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>.
- Schär, C., P. Vidale, D. Luethi, C. Frei, C. Haerberli, M. 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>.
- Shepherd, T., 2016: A common framework for approaches to extreme event attribution. *Curr. Climate Change Rep.*, **2**, 28–38, <https://doi.org/10.1007/s40641-016-0033-y>.
- Sippel, S., and F. Otto, 2014: Beyond climatological extremes—Assessing how the odds of hydrometeorological extreme events in south-east Europe change in a warming climate. *Climatic Change*, **125**, 381–398, <https://doi.org/10.1007/s10584-014-1153-9>.
- Stott, P., D. Stone, and M. Allen, 2004: Human contribution to the European heatwave of 2003. *Nature*, **432**, 610–614, <https://doi.org/10.1038/nature03089>.
- , and Coauthors, 2013: Attribution of weather and climate-related events. *Climate Science for Serving Society: Research, Modeling and Prediction Priorities*, A. Ghassem and J. W. Hurrell, Eds., 307–337, https://doi.org/10.1007/978-94-007-6692-1_12.
- Taylor, K., R. Stouffer, and G. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Trenberth, K., J. Fasullo, and T. Shepherd, 2015: Attribution of climate extreme events. *Nat. Climate Change*, **5**, 725–730, <https://doi.org/10.1038/nclimate2657>.
- Trigo, R., R. García-Herrera, J. Díaz, I. Trigo, and M. Valente, 2005: How exceptional was the early August 2003 heatwave in France? *Geophys. Res. Lett.*, **32**, L10701, <https://doi.org/10.1029/2005GL022410>.
- Uhe, P., F. Otto, K. Haustein, G. van Oldenborgh, A. King, D. Wallom, M. Allen, and H. Cullen, 2016: Comparison of methods: Attributing the 2014 record European temperatures to human influences. *Geophys. Res. Lett.*, **43**, 8685–8693, <https://doi.org/10.1002/2016GL069568>.
- van Oldenborgh, G., and Coauthors, 2016: Rapid attribution of the May/June 2016 flood-inducing precipitation in France and Germany to climate change. *Hydrol. Earth Syst. Sci. Discuss.*, <https://doi.org/10.5194/hess-2016-308>.
- Vautard, R., P. Yiou, F. Otto, P. Stott, N. Christidis, G. van Oldenborgh, and N. Schaller, 2016: Attribution of human-induced dynamical and thermodynamical contributions in extreme weather events. *Environ. Res. Lett.*, **11**, 114009, <https://doi.org/10.1088/1748-9326/11/11/114009>.
- World Health Organization, 2003: The health impacts of 2003 summer heat waves. Briefing Note for the Delegations of the Fifty-Third Session of the WHO Regional Committee for Europe, 12 pp.