Weather and Climate Extremes 47 (2025) 100743

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

Data-driven upper bounds and event attribution for unprecedented heatwaves

Mark D. Risser [a,*], Likun Zhang [b], Michael F. Wehner [c]

[a] Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, United States of America
[b] Department of Statistics, University of Missouri, Columbia, MO, 65211, United States of America
[c] Applied Mathematics and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, United States of America

ARTICLE INFO

Keywords:
Extreme temperature
Upper bounds
Uncertainty quantification
Spatial statistics
Extreme value analysis
Bayesian analysis

ABSTRACT

The last decade has seen numerous record-shattering heatwaves in all corners of the globe. In the aftermath of these devastating events, there is interest in identifying worst-case thresholds or upper bounds that quantify just how hot temperatures can become. Generalized Extreme Value theory provides a data-driven estimate of extreme thresholds; however, upper bounds may be exceeded by future events, which undermines attribution and planning for heatwave impacts. Here, we show how the occurrence and relative probability of observed yet unprecedented events that exceed *a priori* upper bound estimates, so-called "impossible" temperatures, has changed over time. We find that many unprecedented events are actually within data-driven upper bounds, but only when using modern spatial statistical methods. Furthermore, there are clear connections between anthropogenic forcing and the "impossibility" of the most extreme temperatures. Robust understanding of heatwave thresholds provides critical information about future record-breaking events and how their extremity relates to historical measurements.

1. Introduction

In recent years, our planet has experienced a growing number of record-breaking heatwaves that have a devastating impact on human health and infrastructure. Western Russia experienced temperatures in June, 2010, that were unprecedented since at least the 19th century (Rahmstorf and Coumou, 2011), which contributed to significant loss of human life (Dole et al., 2011). A deadly heatwave impacted much of western Europe in June and July, 2019, breaking previous records in metropolitan France by nearly 2 °C and directly causing hundreds of excess deaths (Mitchell et al., 2018; Vautard et al., 2020). In late June, 2021, an unprecedented heatwave struck the United States Pacific Northwest and western Canada that broke all-time records by more than 15 °C (Bercos-Hickey et al., 2022) and was among the most extreme events ever recorded globally (Thompson et al., 2022), causing over 500 deaths (Popovich and Choi-Schagrin, 2021) and significant agricultural losses (Baker and Sergio, 2021). These and many other similarly devastating heatwaves are termed "low likelihood, high-impact" (LLHI) weather events by the sixth Intergovernmental Panel on Climate Change Report, which furthermore states that we currently have low confidence in current and future projections of LLHIs (Seneviratne et al., 2021). It is now clear that large swaths of the globe are vulnerable to LLHI heatwaves (Thompson et al., 2023), which poses serious problems for adaptation and impacts planning.

In the aftermath of a devastating heatwave event, there is considerable interest in quantifying worst-case upper bounds on extreme hot temperatures as well as their frequency or return interval. Atmospheric theory provides physical upper bounds on extreme temperatures (see, e.g., Zhang and Boos, 2023), but these generally correspond to idealized conditions that may rarely occur in reality. Generalized extreme value (GEV) theory provides a data-driven estimate of thresholds for extremes: when the GEV shape parameter is negative, the distribution has a finite upper bound (Coles, 2001). When the shape parameter is less than −0.3, the distribution is very sharply bounded. However, for the aforementioned heatwaves in Russia, France, and British Columbia, the hottest temperatures experienced during the event exceeded *a priori* GEV-based upper bounds. This leads us to define a so-called "impossible" temperature: measurements that are extreme enough relative to more typical extrema that they exceed what was previously thought to be the hottest possible temperatures. We use "impossible" informally since, of course, the temperatures actually occurred in Nature. Impossible temperatures defined this way pose serious problems to planning for the impacts of extreme heatwaves: the best estimate is that these events have a zero probability, meaning we cannot assess their rarity

* Corresponding author.
E-mail address: mdrisser@lbl.gov (M.D. Risser).

https://doi.org/10.1016/j.wace.2025.100743
Received 9 August 2024; Received in revised form 13 January 2025; Accepted 23 January 2025
Available online 30 January 2025

or return interval much less determine the extent to which human activities affected the statistics of the event.

In many cases, the data-driven upper bounds used to qualify a temperature as "impossible" are derived using traditional extreme value methods: non-stationary GEV analysis with a monotonically increasing covariate such as time or greenhouse gas concentrations to account for climate change, applied independently to the records at each weather station (see, e.g, Van Oldenborgh et al., 2019; Philip et al., 2021; Bercos-Hickey et al., 2022). It has been shown that augmenting time trends with additional physically-based covariates can often help anticipate the most extreme temperatures (Zeder and Fischer, 2023). While the traditional approach (even when augmented with additional covariates) is relatively straightforward to implement and hence broadly used, it ignores an obvious source of information: measurements of temperature extremes from nearby locations. One solution to address this limitation leverages a relatively old idea, wherein one "trades space for time" as is done in regional frequency analysis (Hosking and Wallis, 1993). This is particularly important given that the relatively short observational record can lead to real-world events being deemed impossible (Zeder et al., 2023). The broad statistical literature on novel extreme value techniques (see, e.g., Huser and Wadsworth, 2019; Zhang et al., 2021, 2023) allows us to use information from spatially-nearby sites and provides a path forward for assigning non-zero probabilities to the most extreme temperatures, even from a data-driven perspective.

In this paper, we explore how methodological choices impact data-driven upper bound thresholds for the most extreme temperatures, what we refer to as *unprecedented* heatwaves. We specifically focus on one-day temperature extremes, since their underlying analyses commonly form the basis for climate change adaptation, mitigation, and attribution. "Unprecedented" is defined as annual maximum daily maximum temperatures that are at least 4σ events relative to other annual maxima (see Section 2.2). Our approach leverages in situ records and is distinct from dynamical model-based studies to assess statistics of impossible temperatures (e.g., McKinnon and Simpson, 2022; Fischer et al., 2023). We explicitly demonstrate that, using the same input data, it is *very likely* that state-of-the-art methods can explain 69.5% more of the impossible temperatures from the historical record relative to the traditional approach. Equipped with quantitatively non-zero probabilities for these events, we then revisit the attribution problem and propose a metric to quantify the relative rarity of the next unprecedented heatwave event.

2. Materials and methods

2.1. Data sources

We analyze measurements of daily maximum temperature (°C) from the Global Historical Climate Network-Daily (GHCN-D) database (Menne et al., 2012) over the historical record, defined as 1901 to 2022. We identify a high-quality set of records based on a minimum threshold of non-missing daily measurements as follows: first, we define annual "blocks" as January–December for stations in the Northern Hemisphere and July–June for stations in the Southern Hemisphere. Next, we calculate and store the maximum daily maximum temperature (denoted "TXx") in each block-year so long as that block-year had at least 66.7% non-missing daily measurements. We also require that the TXx occurs in the warm season, i.e., April–September for the Northern Hemisphere and October–March for the Southern Hemisphere. We then select stations that have at least 50 years of non-missing TXx measurements over 1901–2022. We acknowledge that allowing for as many as one-third of missing measurements in a given year is anti-conservative; this choice was made to ensure we have as many measurements for analysis as possible. Finally, we remove stations that have less than one other station within approximately 500 km, since this prevents us from leveraging spatially-nearby measurements (this step removes 121 stations from the 8113 records with at least 50 years of data). Ultimately, this yields $N = 7992$ gauged locations for analysis, denoted S, the geographic distribution of which is shown in Supplemental Figure S3. Denote the TXx measurements as $\{Y(s,t)\}$ in year $t = 1901, \ldots 2022$ for station $\mathbf{s} \in S$. Across all stations and years, this yields $n = 612,735$ non-missing TXx measurements for analysis. It is important to note that the temperature records we analyze are of differing length: of the 7992 records, only around 2000 have non-missing TXx measurements in the early 1900s and only around 4000 in 2022; the highest density of non-missing records was in the 1960s (see Fig. 1a). The geospatial distribution of the overall length of temperature records and for each decade is shown in Supplemental Figure S4.

Physical information about the Earth system is a useful tool for describing spatial and spatio-temporal variability in the behavior of weather extremes (see, e.g., Zhang et al., 2010; Risser et al., 2021; Zeder and Fischer, 2023; Zhang et al., 2024). First, we use five covariates to describe year-to-year variability and secular trends in the TXx climatology: a time series of the radiative forcing from greenhouse gases (GHGs) to describe human-induced secular trends (following seminal work from Arrhenius, 1897); the ENSO Longitude Index (ELI, Williams and Patricola, 2018) to account for the El Niño-Southern Oscillation (ENSO); the Standardized Precipitation Evapotranspiration Index (SPEI; Vicente-Serrano et al., 2010) to account for the effect of evapotranspirative cooling from the surface soil moisture content and local vegetation (Domeisen et al., 2023); and the Pacific-North American (PNA) teleconnection pattern and the North Atlantic Oscillation (NAO) to account for large-scale modes of climate variability (Kenyon and Hegerl, 2008). While the SPEI covariate is temperature dependent, in general there is a separation of time scales between long-term drought metrics and short term heatwaves. Second, we utilize elevation (meters above sea level) to describe orographically-driven heterogeneity in the statistical parameters that define the climatological distributions. For more information, see Section 1.2 of the Supporting Information and Supplemental Figure S5.

2.2. Selecting a test set of unprecedented temperatures

To assess the efficacy of data-driven upper bounds derived using statistical methods, we need a set of test events that are excluded from the analysis – those that are very extreme (and hence "unprecedented") even relative to more typical annual maxima. (As an aside, note that our use of the term "unprecedented" does not necessarily imply that the temperatures of interest have never before occurred, but instead is used to represent the most extreme events.) For our analysis, it is particularly important to avoid the selection bias associated with so-called "trigger events" (Miralles and Davison, 2023), which can lead to errors in return level estimates of the most extreme events. In selecting a set of test events from gauge-based records, however, it is critical to ensure that the selected events arereal and do not correspond to measurement errors. We therefore propose a threshold-based algorithm to identify the unprecedented events used to test our statistical methods; see Section 2.1 of the Supporting Information for complete details. In short, the method requires that the selected TXx measurements are larger than the 95th percentile of all TXx at the station, at least 4σ larger than typical extreme temperatures at the gauged location, and have at least one neighboring spatial or temporal measurement that is at least a 1σ extreme. We explored various combinations of these subjective thresholds and found that our results were insensitive to these specific choices (see Section 2.1 of the Supporting Information). This approach yields a total of 1692 candidate events out of the more than 600,000 measurements from all 7992 stations. Our algorithm further determines that $n_{\text{err}} = 147$ are measurement error events (which are discarded from the analysis altogether) leaving $n_{\text{real}} = 1545$ real unprecedented events (representing less than 0.3% of all TXx measurements). The unprecedented events are withheld from our various statistical analyses and used as out-of-sample test data. The real vs. measurement error

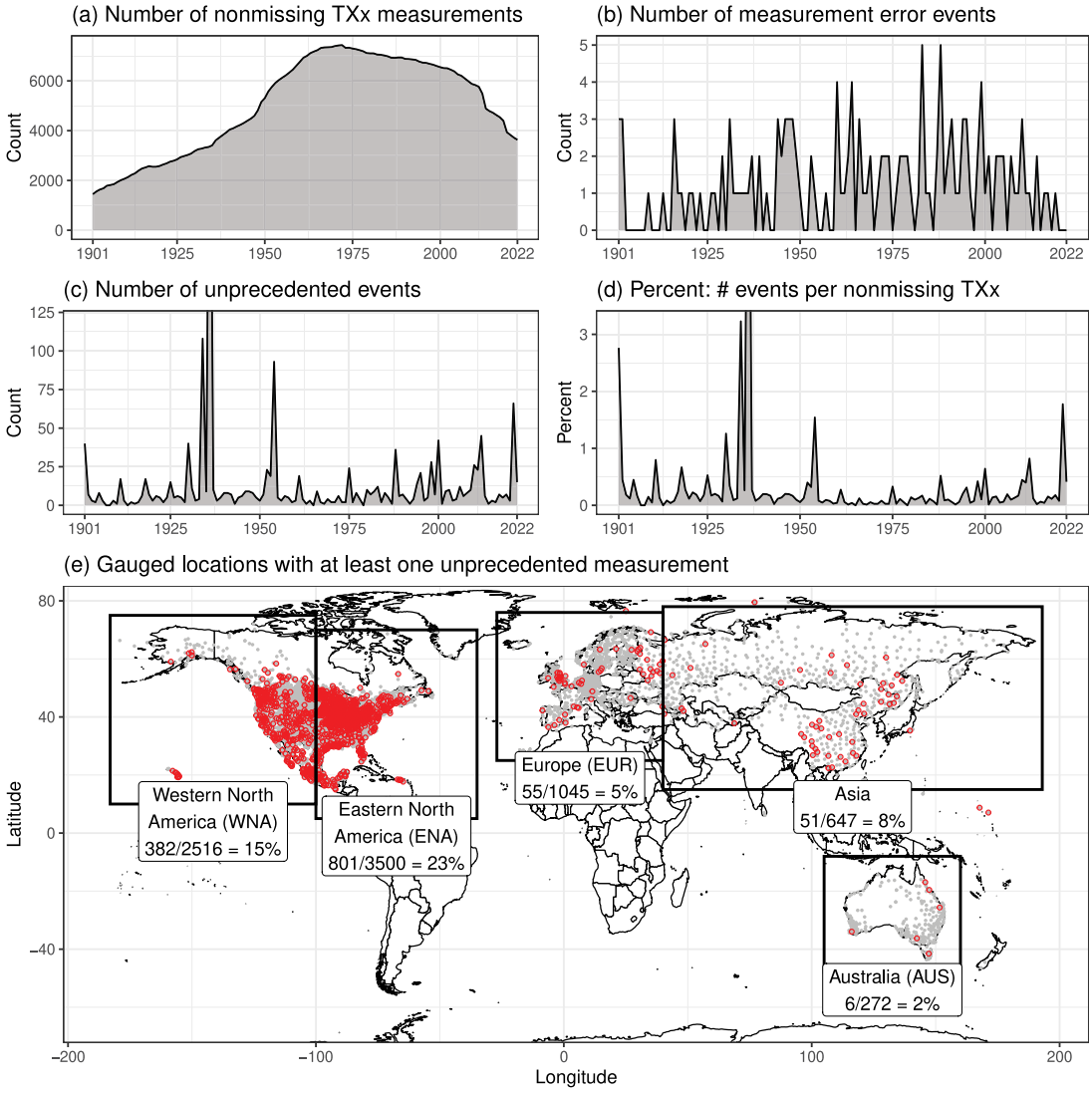

Fig. 1. The number of unprecedented measurements and measurement errors in each year (panels a. and b., respectively), along with the geographic distribution of GHCN-D gauged locations (gray points, all records) with at least one unprecedented event (red circles; panel c.). Recall here that "unprecedented" refers to annual maximum daily maximum temperatures that are at least 4σ events relative to other annual maxima (see Section 2.2 for more details). Boundaries for five continental summaries are also shown in panel (c) along with the fraction of stations in each region with at least one unprecedented event. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

events are tallied for each year in Fig. 1(b)–(c); the geographic distribution of where the unprecedented events occur are shown in Fig. 1(e); Supplemental Table S1 categorizes the number of stations with real and measurement error events. In light of the differing record lengths discussed in Section 2.1 and shown in Fig. 1(a), we also show the occurrence rates of unprecedented events each year, normalized by the total number of non-missing TXx records; see Fig. 1(d). Note that both the number and relative rate of unprecedented events appears to be increasing since about 1960, with trends earlier in the record obscured by large spikes in 1934 (108 events), 1936 (342 events), and 1954 (93 events). Furthermore, the unprecedented events occur across all global land areas that are sampled by the GHCN-D database.

2.3. Extreme value analysis

As described in Section 1, our hypothesis is that many extreme temperatures are deemed "impossible" largely because of methodological choices. Specifically, the traditional approach ignores (at least) three important sources of information: (1) year-to-year variability in temperature extremes, (2) nearby locations will experience the same types of heatwaves (climatological dependence, over long time scales), and (3) nearby locations will also experience the same heatwave events (weather dependence, over short time scales). We briefly describe how we account for each of these sources of information.

First, we suppose the TXx measurements in year t at gauged location $\boldsymbol{s}$, denoted $Y(\boldsymbol{s},t)$, arise from a Generalized Extreme Value (GEV) distribution whose parameters depend on space- and time-varying covariates. The cumulative distribution function for $Y(\boldsymbol{s},t)$ is

$$\mathbb{P}(Y(\boldsymbol{s},t) \leq y) = \exp\left\{-\left[1+\xi(\boldsymbol{s},t)\left(\frac{y-\mu(\boldsymbol{s},t)}{\sigma(\boldsymbol{s},t)}\right)\right]^{-1/\xi(\boldsymbol{s},t)}\right\}$$

(Coles, 2001, Theorem 3.1.1, page 48), defined for $\{y : 1+\xi(\boldsymbol{s},t)(y-\mu(\boldsymbol{s},t))/\sigma(\boldsymbol{s},t) > 0\}$. Following, e.g., Zhang et al. (2010), Sillmann et al. (2011), and Risser et al. (2024), we utilize covariates to describe year-to-year changes in different aspects of the extreme value distribution, assuming

$$\begin{aligned} \mu(\boldsymbol{s},t) &= \mu_0(\boldsymbol{s}) + \mu_1(\boldsymbol{s})\text{GHG}_t + \mu_2(\boldsymbol{s})\text{ELI}_t + \mu_3(\boldsymbol{s})\text{SPEI}(\boldsymbol{s},t) + \\ &\quad \mu_4(\boldsymbol{s})\text{PNA}_t + \mu_5(\boldsymbol{s})\text{NAO}_t \\ \log\sigma(\boldsymbol{s},t) &= \phi_0(\boldsymbol{s}) + \phi_1(\boldsymbol{s})\text{GHG}_t \\ \xi(\boldsymbol{s},t) &\equiv \xi(\boldsymbol{s}) \end{aligned} \tag{1}$$

In other words, the center of the GEV distribution $\mu(s,t)$ is modeled statistically as a linear function of GHG forcing, the ENSO longitude index (ELI), the Standardized Precipitation Evaporation Index (SPEI), the Pacific-North American teleconnection pattern (PNA), and the North Atlantic Oscillation (NAO). The natural logarithm of the year-to-year variability $\log\sigma(s,t)$ is modeled statistically as a linear function of GHG forcing. The shape parameter, $\xi(s)$, governs the upper tail behavior and varies across space but is otherwise time-invariant (as is standard practice for heatwaves; see, e.g., Philip et al., 2020). The GEV formalism allows us to quantify extreme heatwaves using three quantities:

1. Data-driven upper bounds: $b(s,t) = \mu(s,t) - \sigma(s,t)/\xi(s)$. This quantity is well-defined when $\xi(s) < 0$. Note that the estimated upper bound will always be larger than the largest TXx measurement, reiterating the importance of treating the unprecedented measurements as out-of-sample when fitting GEV distributions.
2. Risk probabilities, denoted $p(s,t;u)$, which quantify the likelihood of exceeding a given temperature threshold u at a location s and year t. Alternatively, the risk probability can be defined as the inverse of the return interval for u.
3. The σ-event threshold $\tau(s) = [b(s,t) - m(s,t)]/\sigma(s,t) = -\Gamma(1-\xi(s))/\xi(s)$, where $\Gamma(\cdot)$ is the gamma function and $m(s,t) = \mu(s,t)+\sigma(s,t)\left[\Gamma(1-\xi(s))-1\right]/\xi(s)$ is the GEV mean. This threshold quantifies how extreme a temperature measurement must be (number of σ's) relative to the GEV mean to be considered "impossible". Note that $\tau(s)$ is time-invariant because the shape parameter $\xi(s)$ is time-invariant.

Second, we propose a statistical framework to account for the fact that nearby stations will experience similar heatwave climatologies: in other words, the spatially-varying quantities on the right-hand side of Eq. (1) should be spatially coherent (i.e., spatially dependent). This idea leverages the intuition of Tobler's first law of geography: "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). We thus develop an approach related to conditional independence methods (see, e.g., Cooley et al., 2007; Risser et al., 2019) wherein the GEV coefficients are a linear combination of spatially-coherent, compactly supported basis functions. The centroid of each basis function corresponds to the center of an equal-area hexagonal global grid (each cell with area approximately 200,000 km^2) with nominal spacing of approximately 500 km; 356 of the cells have at least one station. Our approach mimics that of regional frequency analysis (Hosking and Wallis, 1993), which borrows strength across a group of (presumably nearby) sites to estimate extreme statistics. Unlike regional frequency analysis, however, our approach does not require one to identify homogeneous regions and instead simply restricts the GEV coefficients to vary smoothly according to geospatial coordinates (longitude, latitude, and elevation). This restriction ensures that nearby stations will have similar climatological properties.

Finally, we account for the fact that nearby stations will experience the same heatwave events (i.e., accounting for weather dependence) using modern techniques from the spatial extremes literature (Zhang et al., 2023). This approach accounts for spatial structure in the realized TXx values in a given year using a flexible copula model. The copula can account for spatially- and temporally-varying dependence in extreme events (e.g., heatwaves may have different spatial structures in the tropics versus the midlatitudes, and also the beginning versus the end of the record) while also allowing the dependence to vary as a function of how extreme the measurements are. The copula is the embedded within a variational autoencoder (an unsupervised learning technique; see, e.g. Doersch, 2016) to enable scalability to the large, global data set that we set out to analyze in this paper.

Ultimately, we consider a set of statistical models for estimating upper bound thresholds, starting with the traditional approach with only a single covariate, here GHG forcing (denoted "M1"), and sequentially increasing the complexity: adding additional physical covariates (denoted "M2"), accounting for climatological dependence only (denoted "M3"), and furthermore accounting for weather dependence (denoted "M4"); see Supplemental Table S2. For each statistical model, all components will be integrated within a Bayesian framework for uncertainty quantification. For each of the statistical models we use proper but non-informative prior distributions for all statistical parameters. The one exception is the GEV shape parameter, for which we use the maximal data information (MDI) reference prior (Northrop and Attalides, 2016; Zhang and Shaby, 2024) which is the optimal "noninformative" prior distribution. We note that the MDI prior is an "improper" prior, meaning that it is not a valid probability distribution; however, the underlying theory guarantees that it will yield a valid posterior distribution (Zhang and Shaby, 2024). Our expectation is that M4 will perform the best because it captures the most known structure in the data; however, the hierarchy of models allows us to explicitly assess the relative importance of each innovation. Note that we could have considered other combinations of methodologies, e.g., accounting for weather and climate dependence but only using the GHG covariate. However, the spatial aspects of M3 and M4 impose regularization on the covariate coefficients (as with statistical learning methods; see, e.g., Gareth et al., 2013), wherein the regression coefficients can be essentially zeroed out in a data-driven manner when they do not improve the fit of the statistical model. More details on all aspects of the statistical methods are provided in Section 2.2 of the Supporting Information.

2.4. Upper bound uncertainty and containment of unprecedented events

Recall that our main objective is to assess the extent to which statistical methodology impacts the "impossibility" of each unprecedented event identified in Section 2.2. In other words, we want to compare the data-driven upper bound estimates with the temperature thresholds experienced in the unprecedented events and determine whether or not the upper bound is larger than the actual observed temperatures. Given a finite amount of data (e.g., roughly 120 years of TXx measurements), estimates of the statistical parameters of a GEV distribution are uncertain, such that each GEV parameter μ, σ, and ξ are not perfectly known. The implication, then, is that functions of these parameters (e.g., risk probabilities and upper bounds) will also be uncertain. This concept is illustrated in Fig. 2(a), where we show how the TXx measurements imply a statistical fit with uncertainty and how uncertain GEV parameters imply uncertain upper bounds.

Hence, answering the question "is event E contained by upper bound U?" must account for uncertainty innate to the data-driven upper bound. In a Bayesian sense, all uncertainty is summarized by the *posterior distribution* of the upper bound, which is used to calculate best estimates, uncertainty measures, and credible intervals (the Bayesian version of a confidence interval). For the purposes of this analysis, we use the posterior distribution to calculate the probability that a unprecedented event is contained by a given upper bound estimate. We subsequently map the estimated probabilities onto the IPCC likelihood scale (Mastrandrea et al., 2010) in order to assign each probability a qualitative label. Throughout this manuscript, we use the IPCC likelihood language as a convenient way to map probabilities into categories for ease of aggregation across many events and to avoid the "*P*-value trap" (Wasserstein and Lazar, 2016) associated with dichotomous conclusions based on fixed significance levels. We note that this is most certainly not equivalent to carefully pondered and peer-reviewed IPCC assessments. Also, we note that there is alternate yet complementary way to frame quantifying uncertainty in the upper bounds based on upper confidence limits (see Supporting Information Section 2.3); our results are the same regardless of which perspective is taken.

This approach is illustrated in Fig. 2(b), where we plot the posterior distributions of the GEV upper bound from statistical models M1-M4 for the year 2021 at a gauged location near Nanaimo, BC, Canada, that experienced a unprecedented event of 38.0 °C during the Pacific

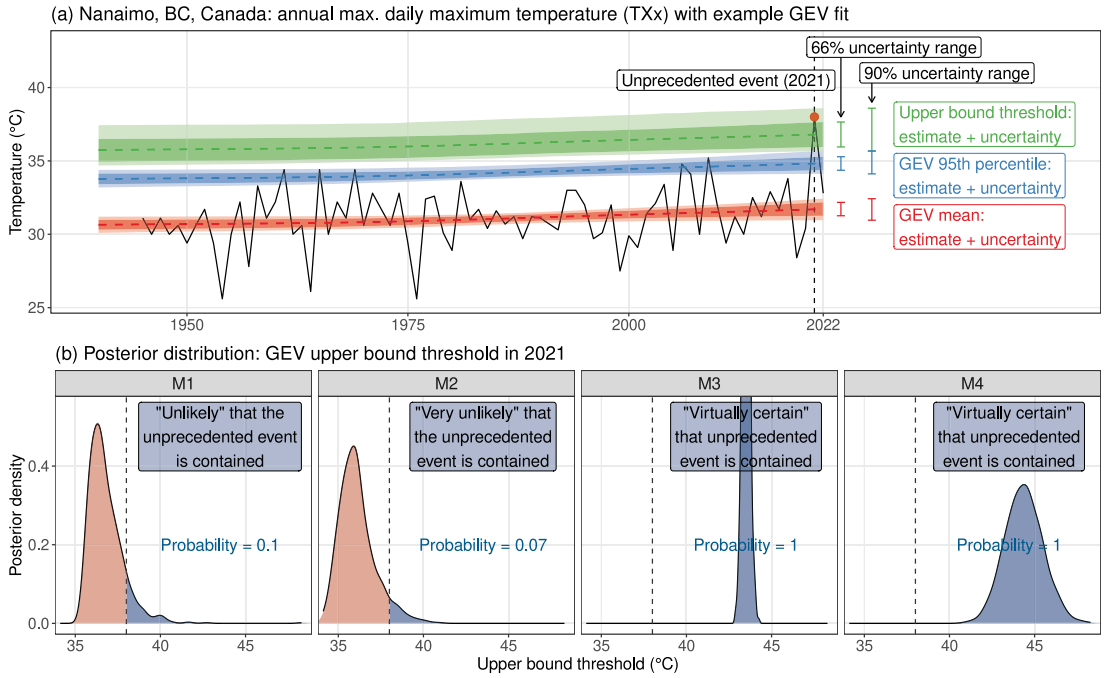

Fig. 2. Demonstration of our approach to account for GEV upper bound uncertainty in determining whether a unprecedented temperature is contained by a given statistical method. Panel (a) shows TXx records from a gauged location near Nanaimo, BC, Canada that experienced a unprecedented event of 38.0 °C in 2021 (red dot), as well as the best estimate (colored dashed lines) and uncertainty intervals (shaded bands) for the mean, 95th percentile, and upper bound threshold of the fitted GEV distribution from M1. Panel (b) shows posterior distributions of the year-2021 upper bound for each statistical method M1-M4 with the corresponding probability that the unprecedented event is contained. Panel (b) also shows the associated likelihood category (Mastrandrea et al., 2010). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Northwest and British Columbia heatwave in the same year (shown in Fig. 2a and denoted by the vertical dashed line in Fig. 2b). The right side of each panel quantifies the posterior probability that the event is contained by the GEV upper bound, which is the area of the posterior density to the right of the unprecedented event threshold. These probabilities are mapped onto the likelihood categories to summarize our statistical confidence that the event is contained, which is shown in the blue boxes in the top right of each panel. We return to this framework in Sections 3.2 and 3.3 to tally the evidence for containment using each statistical model across all unprecedented events.

2.5. Statistical counterfactuals for event attribution

Using covariates to describe year-to-year variability and long-term trends in extreme temperatures yields space- and time-varying estimates of the risk probabilities, such that we can use the fitted statistical models to isolate the human influence on extreme temperatures. This approach follows the "statistical counterfactual" methodology proposed in Risser and Wehner (2017) to make Granger-causal (Granger, 1969) attribution statements: calculate risk probability estimates using a desired combination of GHG forcing, ELI, SPEI, PNA, and NAO. Here, we use the GHG forcing time series as a proxy for anthropogenic influence and ELI, SPEI, PNA, and NAO to describe "natural" or background conditions associated with extreme temperatures. For each unprecedented event, we calculate risk probabilities for two climate "scenarios" described by specific combinations of the physical covariates: *pre-industrial*, with natural conditions (ELI, SPEI, PNA, and NAO) from the year of occurrence and 1901 anthropogenic GHG forcing levels, and *present-day*, with natural conditions from the year of occurrence and 2022 GHG forcing levels. Both scenarios are counterfactual in the sense that they correspond to climate conditions that did not occur in reality.

The risk probabilities are then used to quantify the effect of human-induced GHG forcing on the most extreme temperatures. Specifically, we conduct extreme event attribution (EEA; Attribution of Extreme Weather Events in the Context of Climate Change, 2016) systematically across all unprecedented events to compare the present-day probability of experiencing temperatures at least as extreme as the observed TXx with corresponding pre-industrial probabilities via the "risk ratio" RR (Paciorek et al., 2018). $RR > 1$ implies that increases to GHG forcing cause temperatures at least as large as the observed TXx to become more common, while $RR < 1$ implies the opposite. Three other cases are possible, all involving cases where the risk probabilities are zero (and hence the event is "impossible"): $RR = \infty$ means that the TXx measurement is impossible *without* climate change (i.e., it has non-zero probability under present-day conditions but zero probability under pre-industrial conditions); $RR = 0$ means that the TXx measurement is made impossible by climate change (i.e., it has zero probability under present-day conditions but non-zero probability under pre-industrial conditions); and finally $RR = 0/0$ wherein the risk ratio is mathematically undefined (i.e., the event is impossible in either climate).

3. Results

3.1. Checking GEV assumptions and quality of fitted distributions

The fundamental challenge in modeling the far upper tail and whether a statistical model can adequately represent the behavior of unobserved extremes beyond the support of the data rests in determining whether the underlying postulates under which the distribution of block maxima may converge to the GEV distribution hold in nature. It is therefore critical to ensure that the GEV assumptions are reasonably satisfied (i.e., that the GEV distribution provides a suitable fit to the data) by the various statistical models proposed in Section 2.3. First, in order to safely extrapolate beyond the support of the data, it is important to ensure that our sampling of extreme temperatures maintains max-stability, i.e., that a block size of one year is large enough. The convergence rate of block maxima to a GEV distribution

Table 1
In-sample predictive information criteria and out-of-sample log scores for each fitted statistical model. The WAIC, an in-sample metric, is calculated as two times the sum of the log pointwise predictive density (lppd), which summarizes how well the model fits the data, and the effective number of parameters (p_{WAIC}) to guard against for overfitting (Gelman et al., 2013); smaller scores indicate a better fit. The out-of-sample log pointwise predictive density (lppd-out) summarizes predictive skill for data held out from the process of fitting, here the unprecedented events; larger lppd indicates a better fit. For each metric, the best model is highlighted with bold text.

Model	# of GEV parameters	lppd (in-sample)	p_{WAIC}	WAIC	lppd-out (out-of-sample)
M1	31,968	−1,022,407	20,090.3	**2,084,995**	−10,090.5
M2	71,928	−1,003,502	49,878.7	2,206,761	−9470.3
M3	6052	−1,117,785	9409.2	2,254,388	−8097.5
M4	6052	−1,491,363	214,437.1	3,411,600	**−6821.4**

is determined by the proximity of the daily weather distribution to the GEV domain of attraction (see Chapter 2 of de Haan and Ferreira, 2006). Recent work (Dombry and Ferreira, 2019) derives a closed-form expression for the rate of convergence to GEV based on the number of blocks (here, $T = 122$ years) and the block size (here, the length of the warm season); we argue that these are more than sufficient for the asymptotic behavior to manifest, particularly since we find that in almost all cases extreme temperatures are bounded (see Supplemental Figure S13). Additionally, we use quantile–quantile (Q–Q) plots to check the goodness of fit for a statistical distribution to an empirical sample. Q–Q plots compare the sample quantiles of the data with the corresponding "theoretical" quantiles of the fitted distribution. Points falling along the 45 ° (1–1) line (and within the statistical uncertainty) are evidence of a good fit, i.e., that the sample quantiles are statistically indistinguishable from the theoretical quantiles. Results shown in Section 2.2.5 of the Supporting Information provide strong evidence that the GEV assumptions are satisfied for all four statistical models, and we can be confident in using the fitted distributions to assess the upper tail behavior of extreme daily temperature measurements.

Next, we assess the relative quality of the fitted GEV distributions for statistical models M1-M4. We use two quantitative metrics for this evaluation: the Watanabe–Akaike information criterion (WAIC; Watanabe and Opper, 2010; Gelman et al., 2013) to assess the in-sample performance of each statistical model; and the out-of-sample log pointwise predictive density (lppd-out). The WAIC is defined as

$$\text{WAIC} = -2\text{lppd} + 2p_{\text{WAIC}},$$

where "lppd" is the log pointwise predictive density (summarizing the fit of the statistical distribution to the TXx measurements) and p_{WAIC} is the "effective" number of statistical parameters (used to penalize more flexible statistical models as a way to guard against overfitting). Both quantities are calculated using the fitted posterior distribution and aggregated over all space–time measurements; see Gelman et al. (2013) and Vehtari et al. (2017) for further details. Smaller WAIC indicates a better model fit. WAIC is commonly used for Bayesian model comparison because it fully captures posterior uncertainty, and it is particular helpful in examining the goodness-of-fit for the entire distribution (including the far upper tail). Therefore, it has been adopted extensively in spatial extremes literature; e.g., `R-INLA` (Rue et al., 2017), skew-t process (Hazra et al., 2020), conditional extremes modeling (Simpson et al., 2023), to name just a few. Since the WAIC evaluates in-sample performance, we also calculate the out-of-sample lppd (i.e., the lppd calculated for the held-out unprecedented events); larger lppd-out indicates better model fit.

WAIC and lppd-out results are given in Table 1. For statistical models M1-M4, Table 1 shows the number of statistical parameters in the marginal GEV model, the WAIC and its components (lppd and p_{WAIC}), and the out-of-sample lppd. The best model is highlighted in bold text for WAIC and lppd-out. First, comparing M1 and M2, note that the lppd for M2 is actually better (larger) than that of M1, which is a direct consequence of M2 having more than twice the number of GEV parameters (71,928 versus 31,968). However, the improvement in the lppd is not enough to offset the increased complexity of the model, such that the WAIC metric prefers M1 over M2. Somewhat surprisingly, the WAIC for M3 and M4 are worse than M1, which indicates that the simplest model provides the best in-sample fit to the data. It is noteworthy, however, that the lppd for M3 is not that much worse than M1 with only about 1/5 as many GEV parameters. When assessing the out-of-sample fit, it is clear that M3 and M4 significantly outperform M1 and M2, with M4 being the best model by a wide margin.

In summary, the statistical models that account for climate (M3) and weather (M4) dependence yield lppd and WAIC scores that are not too much worse than the baseline model M1 while providing a much better fit to the far upper tail of the fitted GEV distributions. As such, we argue that M3 and M4 have important benefits relative to simpler approaches and provide the best statistical fit for the most extreme temperature events.

3.2. Case study events

While our final results involve a systematic assessment of extreme events, to demonstrate our methodology we assess three unprecedented temperatures recorded in the historical record corresponding to the severe heatwave events described in Section 1; see Table 2. Historical TXx measurements and the unprecedented event of interest are shown in Fig. 3(a). In this section, we focus on the event-year upper bound estimates because, as mentioned above, they correspond to the real conditions present during each event.

Following the framework outlined in Section 2.4 and visualized in Fig. 2(b), we tally the the probability that the GEV upper bound contains the unprecedented TXx in Fig. 3(b). The bar height in Fig. 3(b) corresponds to the posterior density to the right of the unprecedented TXx (blue shading) in Fig. 2(b). The plotted colors in Fig. 3(b) map the probability that events are contained by the upper bound onto the IPCC likelihood categories, where darker colors indicate better performance (i.e., higher probability of containment) of the statistical method. Generally speaking, we can see that increasing complexity of the statistical model results in more probable containment of the unprecedented TXx. Moving from left (more traditional, M1 and M2) to right (more novel, M3 and M4) across the x-axis, the probability of containment for each unprecedented TXx increases. For all three events, M1 and M2 have low containment probabilities (*unlikely* or *very unlikely*). Notably, M4 contains all three unprecedented temperatures with a probability of 1. M3 has similarly strong performance except for the Russian heatwave, for which it fails to contain the observed TXx. Finally, it is notable that for the French and BC heatwaves, M1 outperforms M2; in other words, including more covariates results in a *smaller* probability of containment. This suggests that a spatial analysis is required when introducing multiple covariates: without the "borrowing of strength" enabled by trading space for time (à la Hosking and Wallis, 1993), the loss of degrees of freedom caused by adding covariates results in additional statistical noise.

Why are the upper bounds from statistical methods M3 and M4 so much more likely to contain the unprecedented events? To explore this more carefully, Supplemental Figure S14 shows how the GEV location μ, scale σ, shape ξ, and upper bound b depend on methodology for the three case study events as well as three other unprecedented events from Western Australia (1933), the Central U.S. (1936), and Mexico (1971). There is no systematic way in which the GEV parameters change to increase the upper bounds for M3 and M4 relative to M1 and M2: in some cases it is due to a larger location parameter (e.g., the events in Mexico and Kostomuksha, RUS); in other cases a larger scale (e.g., the events in Australia and Montpellier, FRA); in yet others a larger shape (e.g., the events in Central U.S. and Nanaimo, CAN). In all cases, these changes result in larger upper bound thresholds and larger return levels (return level curves for each of these events are

Table 2
Three unprecedented temperatures selected as case studies to demonstrate our methodology. The event refers to the annual maximum daily maximum temperature (TXx) recorded at each GHCN gauged measurement site. For reference we also provide the nearest large city to each GHCN site.

Date	GHCN ID	Coordinates	Elevation	Nearest city	Event
29 Jul 2010	RSM00022602	(30.82°E, 63.82°N)	180.0 m	Kostomuksha, RUS	35.5 °C
28 Jun 2019	FRE00106207	(3.96°E, 43.58°N)	2.0 m	Montpellier, FRA	43.5 °C
27 Jun 2021	CA001021830	(124.9°W, 49.72°N)	26.0 m	Nanaimo, CAN	38.0 °C

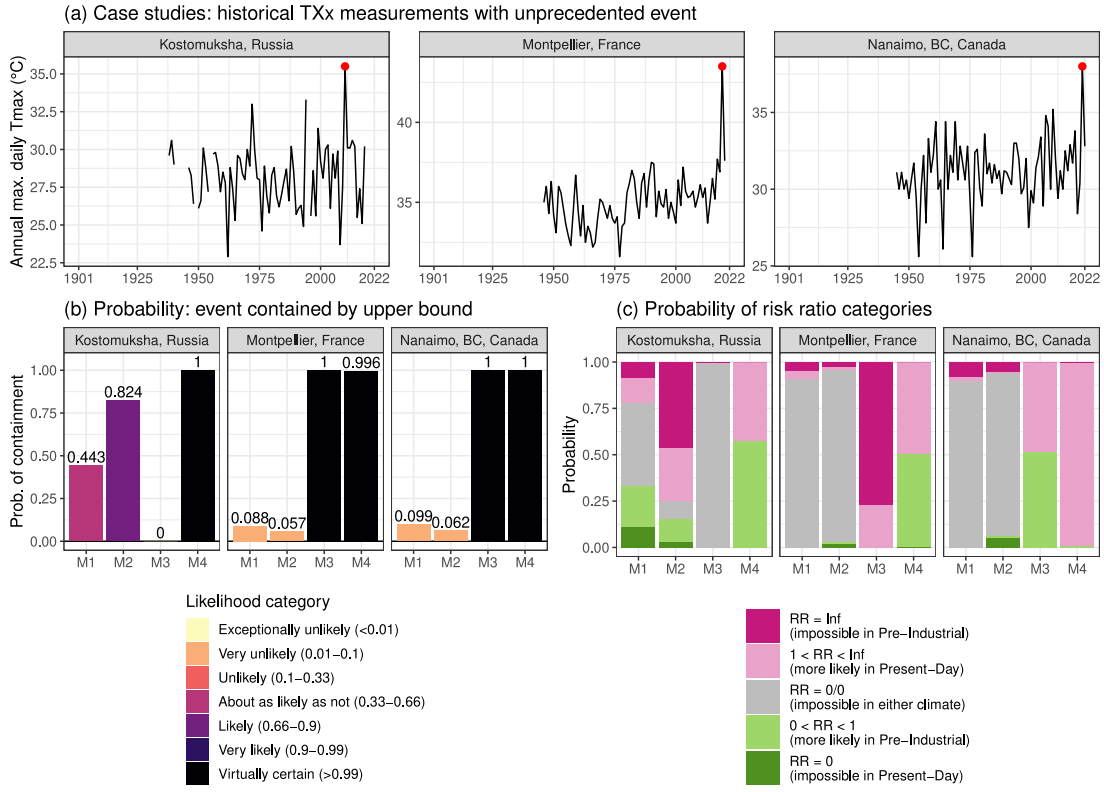

Fig. 3. Summary of results for the three unprecedented events in Table 2. Panel (a) shows the historical TXx measurements with the unprecedented event indicated by a red dot. Panel (b) shows the probability that the unprecedented event is contained by the GEV upper bound, where the plotted color maps the probability of containment onto IPCC likelihood categories. Darker colors indicate better performance. Panel (c) tallies the posterior probability that the risk ratio is in each of a set of categories. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

shown in Supplemental Figure S15). These examples illustrate how the GEV upper bound can be influenced by any one of the three statistical parameters of which it is a function of, and how accounting for dependence (weather and climate) can influence the GEV parameters in unpredictable ways.

Finally, equipped with more robust estimates of data-driven upper bounds, we return to the attribution question and assess the extent to which anthropogenic climate change (via increases to GHG forcing) influences the probability of experiencing events that are at least as extreme as the unprecedented temperatures. Since we have posterior distributions of the GEV parameters, we can similarly obtain posterior distributions of the risk probabilities and hence risk ratios, from which we can calculate posterior probabilities that the risk ratio is in one of five non-overlapping categories: $RR = 0/0$, $RR = 0$, $0 < RR < 1$, $1 < RR < \infty$, and $RR = \infty$. As with upper bounds, there is a big impact of statistical methodology on risk ratio estimates. In particular, M3 and M4 completely side-step the challenge of the risk ratio being undefined ($RR = 0/0$), except for M3 and the Russian heatwave. Furthermore, M4 is able to estimate non-zero risk probabilities for both scenarios, avoiding even the $RR = 0$ and $RR = \infty$ outcomes. Focusing in on M4, the best estimates of the risk ratios (posterior median) are 0.87, 0.96, and 6.39 for the three events, respectively, although in each case there is a large probability that $RR > 1$ (0.42, 0.495, and 0.992, respectively). We reiterate that these event attribution statements are conditional on the background conditions (e.g., ENSO, SPEI, etc.), which may explain the fact that the best estimates of the risk ratio are less than one for the France and Russian events.

3.3. Upper bound thresholds for all unprecedented events

Next, we step back and assess the extent to which all $n_{\text{real}} = 1545$ unprecedented temperatures are contained by the upper bound estimates from each statistical model (M1-M4) following the framework proposed in Section 2.4. Fig. 4 tallies the percentage of events overall and in each continental subregion that are contained for each IPCC likelihood category, determined from probability that the unprecedented TXx is less than the GEV upper bound. Darker colors indicate better performance of the statistical methodology, wherein the unprecedented events are contained with higher probability.

Three important results emerge from our analysis. First, as with the case studies in Section 3.2, it is clear that increasing the complexity of the statistical model results in better containment of the unprecedented TXx and therefore fewer "impossible" events. Overall, the percentage of events for which the unprecedented temperatures are *very likely* contained goes from just 16.2% for M1 to 85.7% for M4. However, adding complexity does not always improve performance: in many cases the

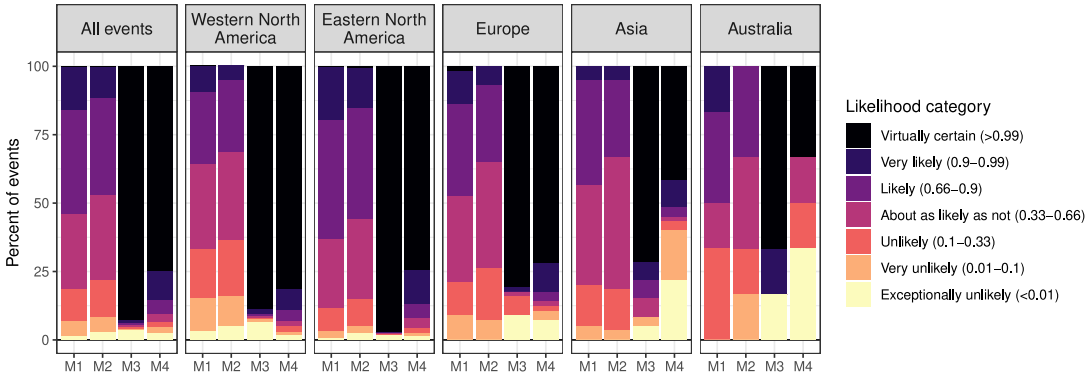

Fig. 4. The impact of statistical methodology on the probability of containment for unprecedented temperatures using data-driven GEV upper bound thresholds, aggregated globally and for the continental subregions shown in Supplemental Figure S1. Darker colors indicate better performance (i.e., larger containment probabilities) of the statistical methodology. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

single-station analysis with multiple covariates but no spatial statistics (M2) does not present an improvement over the traditional single-station analysis with only a time trend (M1). For example, the percent of events that are *very likely* contained decreases from M1 (16.2%) to M2 (11.8%). It is, however, always true that M3 and M4 outperform M1 and M2. This suggests that a spatial analysis is required when introducing multiple covariates: without the "borrowing of strength" enabled by trading space for time (à la Hosking and Wallis, 1993), the loss of degrees of freedom caused by adding covariates results in additional statistical noise.

Second, the effect of statistical methodology is the same regardless of what probability threshold is chosen (0.66, 0.9, etc.). For example, the overall percentage of events that are contained depends quite a lot on choice of IPCC confidence category, but for a given category M3 and M4 always represent an improvement over M1 and M2. The largest improvements emerge as one moves from M2 to M3: the percent of events that are (at least) *very likely* contained goes from just 11.8% for M2 to 93.9% for M3. This demonstrates that even accounting for climatological dependence only allows us to explain many more events that is otherwise possible.

Third, our general conclusions hold regardless of what corner of the globe we are considering. For each of the continental subregions shown in Fig. 4, as one moves from left (more traditional, M1 and M2) to right (more novel, M3 and M4) across the x-axis the colors darken, indicating that more events are contained with less certainty and a higher probability. Furthermore, in all subregions the addition of multiple covariates (M2) degrades performance relative to excluding the drivers of large scale climate variability (as in M1). The unprecedented events are the most "explainable" in Eastern North America, where 94.4% of events are "about as likely as not contained" (at least) under M4. On the other hand, unprecedented are events are less "explainable" in Australia, where just 50% of events are "about as likely as not contained" (at least) under M4.

In conclusion, it is clear that whether or not a given historical event is deemed "impossible" is largely a function of what statistical methods are used: it is *very likely* that 69.5% more events are contained by data-driven upper bounds when accounting for both climatological and weather dependence, relative to the traditional approach (16.2% for M1, compared to 85.7% for M4).

3.4. Changes in the likelihood of unprecedented events

Next, for all $n_{\text{real}} = 1545$ unprecedented events we calculate best estimates of the risk ratios, here taken to be the posterior mode. The posterior mode is a useful summary when we have "non-numeric" outcomes such as $RR = \infty$ and $RR = 0/0$; intuitively, the posterior mode is the risk ratio value that is most probable. We then categorize the best estimates into the same non-overlapping categories used in Section 3.2: $RR = 0/0$, $RR = 0$, $0 < RR < 1$, $1 < RR < \infty$, and $RR = \infty$. When considering all of the events globally, we both aggregate the individual events (i.e., with no averaging, ignoring the geographic sampling of events) while also area-averaging the category probabilities into 200,000 km^2 equal-area hexagonal cells. The latter summary accounts for the non-uniform sampling of the unprecedented events (see panel e. of Fig. 1). The percent of risk ratio best estimates in each category is shown in Fig. 5(a) for all unprecedented events (with and without area averaging), while Fig. 5(b) shows the events tallied separately in each continental subregion. Similar to our results in Section 3.3, there is a clear effect of methodology on the resulting attribution statements: statistical models M3 and M4 yield uniformly larger numbers of well-defined risk ratios (i.e., not the indeterminate $0/0$). This is a direct consequence of the fact that these methods account for weather and climatological dependence and are hence much more likely to yield non-zero risk probability estimates for pre-industrial and present-day conditions. Also notable is the fact that the dark green $RR = 0$ and dark pink $RR = \infty$ categories are generally smaller for models M3 and (especially) M4, again implying that more sophisticated statistical methods allow us to obviate many of the (potentially) hyperbolic statements about climate change making certain events possible or impossible. Similar conclusions hold when considering probabilities of risk ratio categories instead of single-number best estimates; see Supplemental Figure S12.

Interestingly, a large number of risk ratio best-estimates in Fig. 5 are less than one, i.e., many of the unprecedented events are more likely in a pre-industrial climate. Risk ratios of less than 1 are due to decreasing trend estimates in GEV statistics, particularly in Eastern North America where the GHCN records are densely sampled. A major reason for this has to do with the geographic sampling of the unprecedented events: for M4, the fraction of events with $RR < 1$ drops from 65% without area-averaging to less than 37% when accounting for the irregular sampling (see Fig. 5a). This sensitivity of attribution results to regional aggregation has been observed elsewhere; see, e.g., Mindlin et al. (2023). Nonetheless, the surprisingly large fraction of risk ratio estimates less than one is robust across statistical models and has been verified in different analyses that use independent methods and data sets, see e.g., Zhang and Boos (2023). Negative trends in the GEV distribution parameters imply negative trends in the risk probabilities for a fixed threshold, which implies that the unprecedented events are more probable with the lower 1901 levels of GHG forcing (particularly those towards the beginning of the record). Negative trends in temperature extremes in Eastern North America are in part due to the well-documented "global warming hole" in this part of the world (Mascioli et al., 2017). The causes of this local cooling are actively debated as to whether it is a manifestation of internal climate

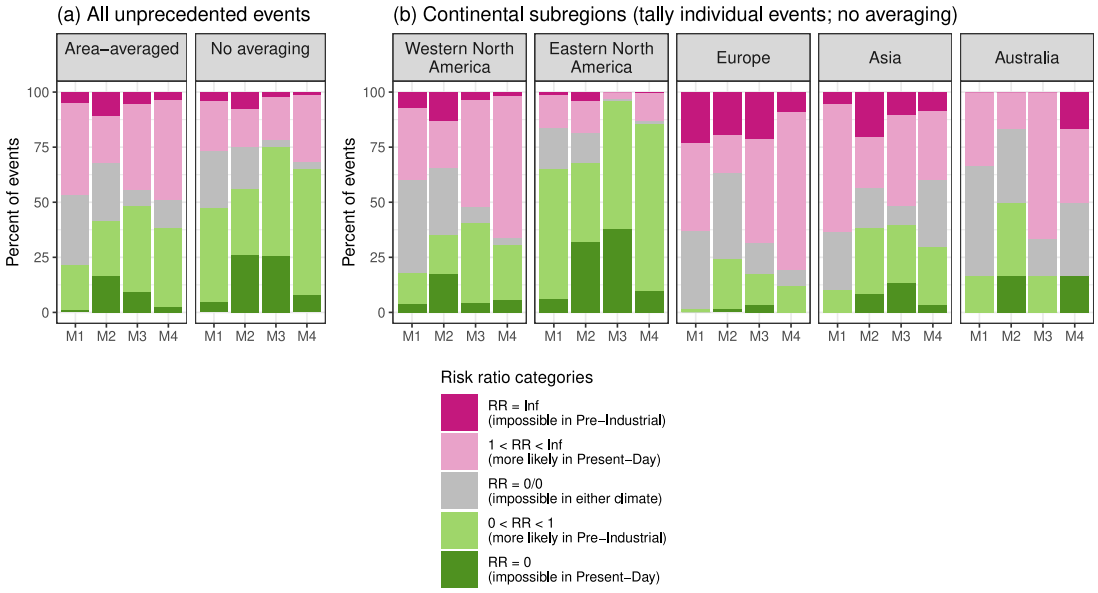

Fig. 5. The impact of statistical methodology on risk ratio best-estimates (posterior mode) for unprecedented events for five risk ratio categories. Panel (a) summarizes all unprecedented events globally, with and without area-weighted averaging (area averaging accounts for the non-uniform sampling of events). Panel (b) tallies the events separately for each continental subregion shown in Fig. 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

variability (Deser et al., 2014; Sun et al., 2022; Kumar et al., 2013) or an externally forced signal (Keil et al., 2020; Chemke et al., 2020; Qasmi, 2023). The contrary appears to be occurring in Europe where very large increases in extreme temperatures have recently been attributed to anthropogenically-driven local circulation changes (Vautard et al., 2023). Hence, a large fraction of European events are deemed impossible without climate change (see Fig. 5). Note that while a large majority of the Eastern North American unprecedented events occur in the first half of the record (see Supplemental Figure S10), the large fraction of risk ratios of less than one in this region persists whether or not we separately consider events from the full record (as in Fig. 5) or from the first half versus second half of the record (see Supplemental Figure S13).

Ultimately, these regions illustrate both the strength and weakness of Granger causal attribution. In our case, the causal influence of forced local circulation changes are only indirectly represented by our anthropogenic covariate (here, GHG forcing, but equivalently global mean temperature). Granger attribution alerts us to complex changes but does not explain them. On the other hand, with some *a priori* understanding and confidence that the anthropogenic covariates adequately represent the forced changes, Granger causal attribution allows defensible estimation of event probabilities and risk ratios.

In summary, using the best statistical model (M4), the best-estimate of the risk ratio is ∞ for 21 of all 1545 unprecedented events, meaning that only about 1.4% of the unprecedented events are, in fact, "impossible" in a pre-industrial climate. However, when we consider only the 314 unprecedented events from the 21st century, when anthropogenic climate change is at its maximum, this percent more than doubles (3.5% or 11 of the 314 events). Even applying our best statistical methodology, there are still 51 unprecedented events that have a best estimate of the risk ratio that is undefined, meaning the events are "impossible" (i.e., have a probability of zero) in either a pre-industrial or present-day climate. However, this is a significant improvement over the 406 unprecedented events with $RR = 0/0$ under the traditional approach (M1).

3.5. Relative extremity of the most severe temperatures

In light of our results on the changing probabilities of unprecedented heatwaves, a final question is: just how extreme may the next record-breaking heatwave be? This is clearly a challenging question from the annual perspective taken in this paper, e.g., the dependence of heat extremes on climate change, large-scale modes of climate variability, and drought conditions — not to mention the complicated meteorological conditions associated with the most extreme temperatures (see, e.g., McKinnon and Simpson, 2022; Wang et al., 2022; Mo et al., 2022) which are not considered in our analyses. To obviate aspects of this challenge, we return to the σ-event threshold, denoted $\tau(\boldsymbol{s})$, which defines a time-invariant relative threshold for the hottest temperatures based on how many "σ's" that temperature is from a typical extreme (see Section 2.3). This metric is already utilized in the literature on the most extreme temperatures: for example, McKinnon and Simpson (2022) found that the 2021 Pacific Northwest heatwave was a 4.5σ event and that such events are represented in a large climate model ensemble. As a metric, $\tau(\boldsymbol{s})$ is useful from an adaptation perspective: together with $\tau(\boldsymbol{s})$, knowledge of a typical extreme (a proxy for $m(\boldsymbol{s}, t)$, the GEV mean) and the year-to-year variability in extreme temperatures (a proxy for $\sigma(\boldsymbol{s}, t)$) provides an approximate estimate of how hot temperatures may become.

We now shift our focus from the stations with at least one of the $n_{\text{real}} = 1545$ unprecedented events back to all $N = 7992$ gauged locations from the GHCN-D database in order to provide estimates of the relative extremity of the most severe temperatures for the global land regions represented by these stations (note, however, that the outlier and real events are still excluded from the analysis). The spatial distribution of the σ-event threshold $\tau(\boldsymbol{s})$ calculated using statistical model M4 is shown in Fig. 6(a), with regional boxplots of $\tau(\boldsymbol{s})$ shown for all statistical models in Fig. 6(b). Our best estimates from M4 show that most often an "impossible" temperature is approximately a 4–5σ event. The largest values of $\tau(\boldsymbol{s})$ generally occur inland, e.g., the central United States, eastern Australia, and parts of central Asia, where our method predicts up to 7–8σ events. On the other hand, coastal regions have relatively smaller sigma-thresholds, where we estimate that temperatures may only approach 4σ events (see, e.g., Western North America, coastal Australia, and northern Europe), reflecting the moderating effect of the oceans on both internal variability and externally-forced changes. The utility of the statistical methods that leverage spatial dependencies (M3 and M4) is further emphasized by the regional summaries in Fig. 6(b): the single-station analyses of M1 and M2 suggest that events in excess of 20σ are possible, which is physically implausible. These

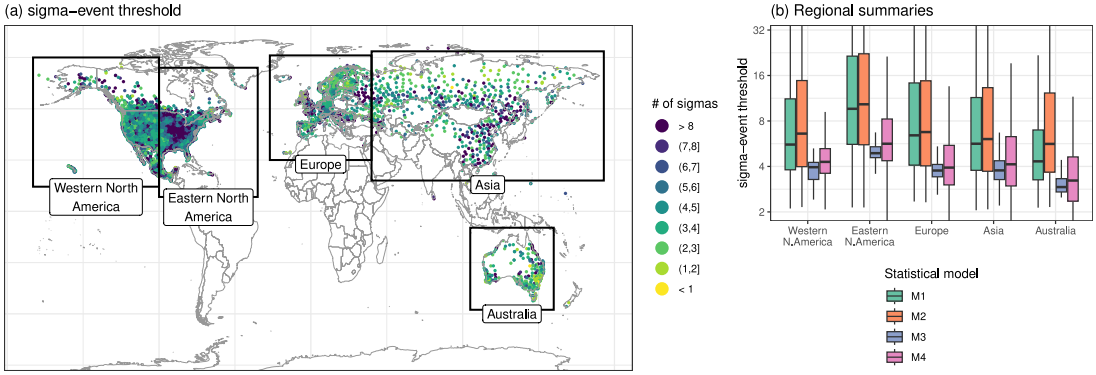

Fig. 6. The spatial distribution of sigma-event thresholds $\tau(s)$, calculated using statistical model M4, which describes a relative upper bound threshold for how many "σ's" an extreme temperature might reach (panel a.; plotted color indicates the best-estimate posterior median). Panel (b) shows regional boxplots for the five boxes drawn on the map in panel (a) for all four statistical models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

results reiterate that single-station analyses involve a large degree of uncertainty and hence introduce statistical noise when estimating upper bound thresholds and probabilities associated with the most extreme temperatures.

4. Discussion

In summary, we have clearly demonstrated how the choice of statistical methodology impacts whether the temperatures experienced during the most extreme heatwaves are deemed statistically impossible, as well as the impact on corresponding assessments of the anthropogenic influence on these events. Using the best available statistical tools allows us to anticipate a much larger fraction of the unprecedented temperatures while minimizing indeterminate risk ratios (i.e., $RR = 0/0$). Unfortunately, standard statistical software is only available for the more traditional methods (M1 and M2). In the future, we plan to develop open-source extreme value analysis tools based on M3 and M4 for use by the broader extreme event attribution community.

The other clear message of this paper is that the only way to obtain robust estimates of heatwave probabilities and return intervals is by leveraging the weather and climatological dependence innate to measurements of extreme temperatures. This is particularly important for observational analysis in light of limitations imposed by the relatively short time period from which we have in situ measurements (Zeder et al., 2023). As previously mentioned, while the methods applied here (particularly M4) are cutting-edge even in the statistics literature, the underlying intuition of "trading space for time" is a relatively old idea (Hosking and Wallis, 1993) that has clear utility for analysis of the most extreme heatwaves.

As mentioned in Section 1, a systematic and objective evaluation of extreme heatwaves that considers all global land regions is needed to address outstanding uncertainties regarding low-likelihood, high impact (LLHI) extreme events such as heatwaves (Seneviratne et al., 2021). Our results address these uncertainties in two ways: first, the unprecedented test events we studied occur throughout the 20th century and across the globe, allowing us to avoid the selection bias associated with focusing on more recent events that occur in primarily Western countries. Second, providing non-zero probability estimates for a large majority of the unprecedented events ensures that we can more robustly study the statistics of the most extreme heatwaves that have occurred in the historical record. Such observational results are critical for evaluating and improving the dynamical models used to develop projections of future climate and the LLHI events that have not yet occurred.

Finally, to accompany this article, we have developed an online graphical user interface (GUI) that allows readers to assess the statistical properties and containment probabilities of an arbitrary unprecedented event. A link to the GUI is provided in the Open Research section below. The GUI has functionality that allows the user to select a unprecedented event based on customized longitude-latitude bounds and a given date range. Then, there are several tabs that show all summaries in Sections 2.4 and 3.2: GEV upper bounds and containment probabilities, IPCC likelihood statements regarding containment, return level curves, GEV parameters for the year in which the event occurred, risk probabilities, and posterior probabilities that the risk ratio is in the five categories used in Section 3.2. While this manuscript presents a summary of all unprecedented events via aggregation, a reader might be interested in the properties of a specific event beyond those presented in Section 3.2. Our GUI is designed to provide this specific information.

CRediT authorship contribution statement

Mark D. Risser: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Likun Zhang:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Michael F. Wehner:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization.

Open research

The in situ temperature records supporting this article are based on publicly available measurements from the National Centers for Environmental Information (https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily).

A graphical user interface to visualize the results of our analysis is provided at https://mark-risser.shinyapps.io/impossible-temperatures/.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Mark D. Risser reports financial support was provided by E O Lawrence Berkeley National Laboratory. Michael Wehner reports financial support was provided by E O Lawrence Berkeley National Laboratory. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy under the Regional and Global Model Analysis program and the Calibrated and Systematic Characterization, Attribution, and Detection of Extremes (CASCADE) Scientific Focus Area (Contract No. DE-AC02-05CH11231) and used resources of the National Energy Research Scientific Computing Center (NERSC), also supported by the Office of Science of the U.S. Department of Energy, under Contract No. DE-AC02-05CH11231.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.wace.2025.100743.

Data availability

Data will be made available on request.

References

Arrhenius, S., 1897. On the influence of carbonic acid in the air upon the temperature of the earth. Publ. Astron. Soc. Pac. 9, 14.

2016. Attribution of Extreme Weather Events in the Context of Climate Change. National Academies Press, http://dx.doi.org/10.17226/21852, Retrieved from http://dx.doi.org/10.17226/21852.

Baker, M., Sergio, O., 2021. The Pacific northwest, built for mild summers, is scorching yet again. N. Y. Times.

Bercos-Hickey, E., O'Brien, T.A., Wehner, M.F., Zhang, L., Patricola, C.M., Huang, H., Risser, M.D., 2022. Anthropogenic contributions to the 2021 Pacific Northwest heatwave. Geophys. Res. Lett. 49 (23), e2022GL099396. http://dx.doi.org/10.1029/2022GL099396, Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022GL099396.

Chemke, R., Zanna, L., Polvani, L.M., 2020. Identifying a human signal in the North Atlantic warming hole. Nat. Commun. 11 (1), http://dx.doi.org/10.1038/s41467-020-15285-x, Retrieved from http://dx.doi.org/10.1038/s41467-020-15285-x.

Coles, S., 2001. An introduction to statistical modeling of extreme values. Lecture Notes in Control and Information Sciences, Springer, Retrieved from https://books.google.com/books?id=2nugUEaKqFEC.

Cooley, D., Nychka, D., Naveau, P., 2007. Bayesian spatial modeling of extreme precipitation return levels. J. Amer. Statist. Assoc. 102 (479), 824–840.

de Haan, L., Ferreira, A., 2006. Extreme Value Theory: an Introduction, vol. 3, Springer.

Deser, C., Phillips, A.S., Alexander, M.A., Smoliak, B.V., 2014. Projecting North American climate over the next 50 years: Uncertainty due to internal variability. J. Clim. 27 (6), 2271–2296. http://dx.doi.org/10.1175/jcli-d-13-00451.1, Retrieved from http://dx.doi.org/10.1175/JCLI-D-13-00451.1.

Doersch, C., 2016. Tutorial on variational autoencoders. arXiv preprint arXiv:1606.05908.

Dole, R., Hoerling, M., Perlwitz, J., Eischeid, J., Pegion, P., Zhang, T., Quan, X.-W., Xu, T., Murray, D., 2011. Was there a basis for anticipating the 2010 Russian heat wave? Geophys. Res. Lett. 38 (6), http://dx.doi.org/10.1029/2010gl046582, Retrieved from http://dx.doi.org/10.1029/2010GL046582.

Dombry, C., Ferreira, A., 2019. Maximum likelihood estimators based on the block maxima method. Bernoulli 25 (3), http://dx.doi.org/10.3150/18-bej1032, Retrieved from http://dx.doi.org/10.3150/18-BEJ1032.

Domeisen, D.I.V., Eltahir, E.A.B., Fischer, E.M., Knutti, R., Perkins-Kirkpatrick, S.E., Schär, C., Seneviratne, S.I., Weisheimer, A., Wernli, H., 2023. Prediction and projection of heatwaves. Nat. Rev. Earth Environ. 4 (1), 36–50. http://dx.doi.org/10.1038/s43017-022-00371-z, Retrieved from https://doi.org/10.1038/s43017-022-00371-z.

Fischer, E.M., Beyerle, U., Bloin-Wibe, L., Gessner, C., Humphrey, V., Lehner, F., Pendergrass, A.G., Sippel, S., Zeder, J., Knutti, R., 2023. Storylines for unprecedented heatwaves based on ensemble boosting. Nat. Commun. 14 (1), http://dx.doi.org/10.1038/s41467-023-40112-4, Retrieved from http://dx.doi.org/10.1038/s41467-023-40112-4.

Gareth, J., Daniela, W., Trevor, H., Robert, T., 2013. An Introduction to Statistical Learning: with Applications in R. Spinger.

Gelman, A., Hwang, J., Vehtari, A., 2013. Understanding predictive information criteria for Bayesian models. Stat. Comput. 24 (6), 997–1016. http://dx.doi.org/10.1007/s11222-013-9416-2, Retrieved from http://dx.doi.org/10.1007/s11222-013-9416-2.

Granger, C.W., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 424–438.

Hazra, A., Reich, B.J., Staicu, A.-M., 2020. A multivariate spatial skew-t process for joint modeling of extreme precipitation indexes. Environmetrics 31 (3), e2602.

Hosking, J.R.M., Wallis, J.R., 1993. Some statistics useful in regional frequency analysis. Water Resour. Res. 29 (2), 271–281. http://dx.doi.org/10.1029/92wr01980, Retrieved from http://dx.doi.org/10.1029/92WR01980.

Huser, R., Wadsworth, J.L., 2019. Modeling spatial processes with unknown extremal dependence class. J. Amer. Statist. Assoc. 114 (525), 434–444.

Keil, P., Mauritsen, T., Jungclaus, J., Hedemann, C., Olonscheck, D., Ghosh, R., 2020. Multiple drivers of the north atlantic warming hole. Nature Clim. Change 10 (7), 667–671. http://dx.doi.org/10.1038/s41558-020-0819-8, Retrieved from http://dx.doi.org/10.1038/s41558-020-0819-8.

Kenyon, J., Hegerl, G.C., 2008. Influence of modes of climate variability on global temperature extremes. J. Clim. 21 (15), 3872–3889. http://dx.doi.org/10.1175/2008jcli2125.1, Retrieved from http://dx.doi.org/10.1175/2008JCLI2125.1.

Kumar, S., Kinter, J., Dirmeyer, P.A., Pan, Z., Adams, J., 2013. Multidecadal climate variability and the "warming hole" in north america: Results from CMIP5 twentieth- and twenty-first-century climate simulations. J. Clim. 26 (11), 3511–3527. http://dx.doi.org/10.1175/jcli-d-12-00535.1, Retrieved from http://dx.doi.org/10.1175/JCLI-D-12-00535.1.

Mascioli, N.R., Previdi, M., Fiore, A.M., Ting, M., 2017. Timing and seasonality of the United States 'warming hole'. Environ. Res. Lett. 12 (3), 034008. http://dx.doi.org/10.1088/1748-9326/aa5ef4, Retrieved from http://dx.doi.org/10.1088/1748-9326/aa5ef4.

Mastrandrea, M.D., Field, C.B., Stocker, T.F., Edenhofer, O., Ebi, K.L., Frame, D.J., Held, H., Kriegler, E., Mach, K.J., Matschoss, P.R., et al., 2010. Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties.

McKinnon, K.A., Simpson, I.R., 2022. How unexpected was the 2021 Pacific northwest heatwave? Geophys. Res. Lett. 49 (18), http://dx.doi.org/10.1029/2022gl100380, Retrieved from http://dx.doi.org/10.1029/2022GL100380.

Menne, M.J., Durre, I., Vose, R.S., Gleason, B.E., Houston, T.G., 2012. An overview of the global historical climatology network-daily database. J. Atmos. Ocean. Technol. 29 (7), 897–910.

Mindlin, J., Vera, C.S., Shepherd, T.G., Osman, M., 2023. Plausible drying and wetting scenarios for summer in southeastern south america. J. Clim. 36 (22), 7973–7991. http://dx.doi.org/10.1175/jcli-d-23-0134.1, Retrieved from http://dx.doi.org/10.1175/JCLI-D-23-0134.1.

Miralles, O., Davison, A.C., 2023. Timing and spatial selection bias in rapid extreme event attribution. Weather. Clim. Extrem. 41, 100584. http://dx.doi.org/10.1016/j.wace.2023.100584, Retrieved from http://dx.doi.org/10.1016/j.wace.2023.100584.

Mitchell, D., Heaviside, C., Schaller, N., Allen, M., Ebi, K.L., Fischer, E.M., Gasparrini, A., Harrington, L., Kharin, V., Shiogama, H., Sillmann, J., Sippel, S., Vardoulakis, S., 2018. Extreme heat-related mortality avoided under Paris agreement goals. Nature Clim. Change 8 (7), 551–553. http://dx.doi.org/10.1038/s41558-018-0210-1, Retrieved from http://dx.doi.org/10.1038/s41558-018-0210-1.

Mo, R., Lin, H., Vitart, F., 2022. An anomalous warm-season trans-Pacific atmospheric river linked to the 2021 western north america heatwave. Commun. Earth Environ. 3 (1), 127. http://dx.doi.org/10.1038/s43247-022-00459-w, Retrieved from https://doi.org/10.1038/s43247-022-00459-w.

Northrop, P.J., Attalides, N., 2016. Posterior propriety in Bayesian extreme value analyses using reference priors. Statist. Sinica 721–743.

Paciorek, C.J., Stone, D.A., Wehner, M.F., 2018. Quantifying statistical uncertainty in the attribution of human influence on severe weather. Weather. Clim. Extrem. 20, 69–80. http://dx.doi.org/10.1016/j.wace.2018.01.002, Retrieved from https://doi.org/10.1016/j.wace.2018.01.002.

Philip, S.Y., Kew, S.F., van Oldenborgh, G.J., Anslow, F.S., Seneviratne, S.I., Vautard, R., Coumou, D., Ebi, K.L., Arrighi, J., Singh, R., van Aalst, M., Pereira Marghidan, C., Wehner, M., Yang, W., Li, S., Schumacher, D.L., Hauser, M., Bonnet, R., Luu, L.N., Lehner, F., Gillett, N., Tradowsky, J., Vecchi, G.A., Rodell, C., Stull, R.B., Howard, R., Otto, F.E.L., 2021. Rapid attribution analysis of the extraordinary heatwave on the Pacific coast of the US and Canada june 2021. Earth Syst. Dyn. Discuss. 2021, 1–34. http://dx.doi.org/10.5194/esd-2021-90, Retrieved from https://esd.copernicus.org/preprints/esd-2021-90/.

Philip, S., Kew, S., van Oldenborgh, G.J., Otto, F., Vautard, R., van der Wiel, K., King, A., Lott, F., Arrighi, J., Singh, R., van Aalst, M., 2020. A protocol for probabilistic extreme event attribution analyses. Adv. Stat. Clim. Meteorol. Ocean. 6 (2), 177–203. http://dx.doi.org/10.5194/ascmo-6-177-2020, Retrieved from http://dx.doi.org/10.5194/ascmo-6-177-2020.

Popovich, N., Choi-Schagrin, W., 2021. Hidden toll of the northwest heat wave: hundreds of extra deaths. N.Y. Times 11.

Qasmi, S.d., 2023. Past and future response of the north atlantic warming hole to anthropogenic forcing. Earth Syst. Dyn. 14 (3), 685–695. http://dx.doi.org/10.5194/esd-14-685-2023, Retrieved from http://dx.doi.org/10.5194/esd-14-685-2023.

Rahmstorf, S., Coumou, D., 2011. Increase of extreme events in a warming world. Proc. Natl. Acad. Sci. 108 (44), 17905–17909. http://dx.doi.org/10.1073/pnas.1101766108, Retrieved from http://dx.doi.org/10.1073/pnas.1101766108.

Risser, M.D., Collins, W.D., Wehner, M.F., O'Brien, T.A., Huang, H., Ullrich, P.A., 2024. Anthropogenic aerosols mask increases in US rainfall by greenhouse gases. Nat. Commun. 15 (1), http://dx.doi.org/10.1038/s41467-024-45504-8, Retrieved from http://dx.doi.org/10.1038/s41467-024-45504-8.

Risser, M.D., Paciorek, C.J., Wehner, M.F., O'Brien, T.A., Collins, W.D., 2019. A probabilistic gridded product for daily precipitation extremes over the United States. Clim. Dyn. 53 (5), 2517–2538. http://dx.doi.org/10.1007/s00382-019-04636-0.

Risser, M.D., Wehner, M.F., 2017. Attributable human-induced changes in the likelihood and magnitude of the observed extreme precipitation during Hurricane harvey. Geophys. Res. Lett. 44 (24), 12–457.

Risser, M.D., Wehner, M.F., O'Brien, J.P., Patricola, C.M., O'Brien, T.A., Collins, W.D., Paciorek, C.J., Huang, H., 2021. Quantifying the influence of natural climate variability on in situ measurements of seasonal total and extreme daily precipitation. Clim. Dyn. 56, 3205—-3230. http://dx.doi.org/10.1007/s00382-021-05638-7.

Rue, H., Riebler, A., Sørbye, S.H., Illian, J.B., Simpson, D.P., Lindgren, F.K., 2017. Bayesian computing with INLA: a review. Annu. Rev. Stat. Appl. 4 (1), 395–421.

Seneviratne, S., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S., Wehner, M., Zhou, B., 2021. Weather and climate extreme events. In: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J., Maycock, T., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B. (Eds.), Climate Change 2021: The Physical Science Basis. Contribution of Working Group I To the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1513–1766. http://dx.doi.org/10.1017/9781009157896.013, Retrieved from https://www.ipcc.ch/report/ar6/wg1/chapter/chapter-11/.

Sillmann, J., Croci-Maspoli, M., Kallache, M., Katz, R.W., 2011. Extreme cold winter temperatures in europe under the influence of north atlantic atmospheric blocking. J. Clim. 24 (22), 5899–5913. http://dx.doi.org/10.1175/2011JCLI4075.1, Retrieved from https://journals.ametsoc.org/view/journals/clim/24/22/2011jcli4075.1.xml.

Simpson, E.S., Opitz, T., Wadsworth, J.L., 2023. High-dimensional modeling of spatial and spatio-temporal conditional extremes using INLA and Gaussian Markov random fields. Extremes 26 (4), 669–713.

Sun, C., Zhu, L., Liu, Y., Wei, T., Guo, Z., 2022. CMIP6 model simulation of concurrent continental warming holes in eurasia and north america since 1990 and their relation to the indo-Pacific SST warming. Glob. Planet. Change 213, 103824. http://dx.doi.org/10.1016/j.gloplacha.2022.103824, Retrieved from http://dx.doi.org/10.1016/j.gloplacha.2022.103824.

Thompson, V., Kennedy-Asser, A.T., Vosper, E., Lo, Y.T.E., Huntingford, C., Andrews, O., Collins, M., Hegerl, G.C., Mitchell, D., 2022. The 2021 western North America heat wave among the most extreme events ever recorded globally. Sci. Adv. 8 (18), http://dx.doi.org/10.1126/sciadv.abm6860, Retrieved from http://dx.doi.org/10.1126/sciadv.abm6860.

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

Tobler, W.R., 1970. A computer movie simulating urban growth in the detroit region. Econ. Geogr. 46, 234. http://dx.doi.org/10.2307/143141, Retrieved from http://dx.doi.org/10.2307/143141.

Van Oldenborgh, G.J., Philip, S., Kew, S., Vautard, R., Boucher, O., Otto, F., Haustein, K., Soubeyroux, J.-M., Ribes, A., Robin, Y., et al., 2019. Human contribution to the record-breaking june 2019 heat wave in France. World Weather. Attrib..

Vautard, R., van Aalst, M., Boucher, O., Drouin, A., Haustein, K., Kreienkamp, F., van Oldenborgh, G.J., Otto, F.E.L., Ribes, A., Robin, Y., Schneider, M., Soubeyroux, J.-M., Stott, P., Seneviratne, S.I., Vogel, M.M., Wehner, M., 2020. Human contribution to the record-breaking june and july 2019 heatwaves in Western Europe. Environ. Res. Lett. 15 (9), 094077. http://dx.doi.org/10.1088/1748-9326/aba3d4, Retrieved from http://dx.doi.org/10.1088/1748-9326/aba3d4.

Vautard, R., Cattiaux, J., Happé, T., Singh, J., Bonnet, R., Cassou, C., Coumou, D., D'Andrea, F., Faranda, D., Fischer, E., Ribes, A., Sippel, S., Yiou, P., 2023. Heat extremes in Western Europe increasing faster than simulated due to atmospheric circulation trends. Nat. Commun. 14 (1), http://dx.doi.org/10.1038/s41467-023-42143-3, Retrieved from http://dx.doi.org/10.1038/s41467-023-42143-3.

Vehtari, A., Gelman, A., Gabry, J., 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat. Comput. 27, 1413–1432.

Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J. Clim. 23 (7), 1696–1718.

Wang, C., Zheng, J., Lin, W., Wang, Y., 2022. Unprecedented heatwave in western north america during late june of 2021: Roles of atmospheric circulation and global warming. Adv. Atmos. Sci. http://dx.doi.org/10.1007/s00376-022-2078-2, Retrieved from https://doi.org/10.1007/s00376-022-2078-2.

Wasserstein, R.L., Lazar, N.A., 2016. The ASA statement on p-values: Context, process, and purpose. Amer. Statist. 70 (2), 129–133. http://dx.doi.org/10.1080/00031305.2016.1154108, Retrieved from https://doi.org/10.1080/00031305.2016.1154108.

Watanabe, S., Opper, M., 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. J. Mach. Learn. Res. 11 (12).

Williams, I.N., Patricola, C.M., 2018. Diversity of ENSO events unified by convective threshold sea surface temperature: a nonlinear ENSO index. Geophys. Res. Lett. 45 (17), 9236–9244. http://dx.doi.org/10.1029/2018GL079203.

Zeder, J., Fischer, E.M., 2023. Quantifying the statistical dependence of mid-latitude heatwave intensity and likelihood on prevalent physical drivers and climate change. Adv. Stat. Clim. Meteorol. Ocean. 9 (2), 83–102. http://dx.doi.org/10.5194/ascmo-9-83-2023, Retrieved from http://dx.doi.org/10.5194/ascmo-9-83-2023.

Zeder, J., Sippel, S., Pasche, O.C., Engelke, S., Fischer, E.M., 2023. The effect of a short observational record on the statistics of temperature extremes. Geophys. Res. Lett. 50 (16), http://dx.doi.org/10.1029/2023gl104090, Retrieved from http://dx.doi.org/10.1029/2023GL104090.

Zhang, Y., Boos, W.R., 2023. An upper bound for extreme temperatures over midlatitude land. Proc. Natl. Acad. Sci. 120 (12), http://dx.doi.org/10.1073/pnas.2215278120, Retrieved from http://dx.doi.org/10.1073/pnas.2215278120.

Zhang, L., Ma, X., Wikle, C.K., Huser, R., 2023. Flexible and efficient spatial extremes emulation via variational autoencoders. arXiv preprint arXiv:2307.08079.

Zhang, L., Risser, M.D., Wehner, M.F., O'Brien, T.A., 2024. Leveraging extremal dependence to better characterize the 2021 Pacific northwest heatwave. J. Agric. Biol. Environ. Stat. http://dx.doi.org/10.1007/s13253-024-00636-8, Retrieved from http://dx.doi.org/10.1007/s13253-024-00636-8.

Zhang, L., Shaby, B.A., 2024. Reference priors for the generalized extreme value distribution. Statist. Sinica http://dx.doi.org/10.5705/ss.202021.0258, Retrieved from http://dx.doi.org/10.5705/ss.202021.0258.

Zhang, L., Shaby, B.A., Wadsworth, J.L., 2021. Hierarchical transformed scale mixtures for flexible modeling of spatial extremes on datasets with many locations. J. Amer. Statist. Assoc. 1–13.

Zhang, X., Wang, J., Zwiers, F.W., Groisman, P.Y., 2010. The influence of large-scale climate variability on winter maximum daily precipitation over North America. J. Clim. 23 (11), 2902–2915.