ANNUAL REVIEWS

Annual Review of Statistics and Its Application

Statistical Methods for Extreme Event Attribution in Climate Science

Philippe Naveau,[1] Alexis Hannart,[2] and Aurélien Ribes[3]

[1]Laboratoire de Sciences du Climat et de l'Environnement, IPSL-CNRS, 91191 Gif-sur-Yvette, France; email: philippe.naveau@ipsl.lsce.fr

[2]Ouranos, Montréal, Quebec H3A 1B9, Canada

[3]Centre National de Recherches Météorologiques, 31057 Toulouse, France

Annu. Rev. Stat. Appl. 2020. 7:89–110

First published as a Review in Advance on January 3, 2020

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

https://doi.org/10.1146/annurev-statistics-031219-041314

Keywords

causality theory, detection, attribution, extreme value theory

Abstract

Changes in the Earth's climate have been increasingly observed. Assessing the likelihood that each of these changes has been caused by human influence is important for decision making on mitigation and adaptation policy. Because of their large societal and economic impacts, extreme events have garnered much media attention—have they become more frequent and more intense, and if so, why? To answer such questions, extreme event attribution (EEA) tries to estimate extreme event likelihoods under different scenarios. Over the past decade, statistical methods and experimental designs based on numerical models have been developed, tested, and applied. In this article, we review the basic probability schemes, inference techniques, and statistical hypotheses used in EEA. To implement EEA analysis, the climate community relies on the use of large ensembles of climate model runs. We discuss, from a statistical perspective, how extreme value theory could help to deal with the different modeling uncertainties. In terms of interpretation, we stress that causal counterfactual theory offers an elegant framework that clarifies the design of event attributions. Finally, we pinpoint some remaining statistical challenges, including the choice of the appropriate spatio-temporal scales to enhance attribution power, the modeling of concomitant extreme events in a multivariate context, and the coupling of multi-ensemble and observational uncertainties.

1. INTRODUCTION

Extreme event attribution (EEA) (see, e.g., Stott et al. 2016) aims to evaluate how the likelihood of a specific extreme climate event has been affected by anthropogenic forcings, such as the increase of greenhouse gas concentrations associated with fossil fuel burning. The main idea in EEA is to compare the probability of extreme climate events under a factual scenario of conditions that are close to those observed around the time of the event (e.g., greenhouse gas concentrations and ocean temperatures) with the probability under a counterfactual scenario in which anthropogenic emissions had never occurred (see, e.g., Angélil et al. 2017). Given the random variable of interest, say, daily maxima of temperatures over a given region and a duration, the goal in EEA is to compute the probability that such a random variable is higher than a given large threshold. One trait of EEA is that these small probabilities must be inferred from in silico numerical simulations, under factual and counterfactual scenarios. Another feature of EEA is the choice of the spatiotemporal domain defining the event. Often, an EEA study is driven by a strong societal need (e.g., a specific flood or heat wave) and/or a rare climatological phenomenon (e.g., a record snowstorm). Defining precisely the duration and spatial spread of such events is key, and this leads to various climatological and statistical strategies.

In recent years, a series of review/assessment articles (e.g., Natl. Acad. Sci. Eng. Med. 2016, Shepherd 2016, Stott et al. 2016, Angélil et al. 2017, Chen et al. 2018) have described and summarized the main climatological results and challenges of EEA, but these articles have not focused on the methodological side of the statistical modeling—they have concentrated on climatological advances, pitfalls, and challenges related to climate impact modeling. Our work attempts to complement these assessments by reviewing the different statistical approaches used in EEA and identifying possible bridges between the climatological and statistical communities. In particular, we hope to provide statisticians with the main tools in terms of notation, hypotheses and statistical models. In this context of describing the EEA statistical blueprint, we do not cover the so-called story-line approach, based on decomposing possible causal pathways to explain a specific realization (Trenberth et al. 2015).

Section 2 provides the basic setup. In particular, we stress that the comparison of such probabilities can be mathematically interpreted within the realm of causality theory. In Section 3, we review the types of data at hand and the different sources of uncertainties specific to the climate system. Section 4 covers the choice of the event of interest, whose spatial and temporal scales are key.

Given the data and the event of interest, inference for small probabilities is needed, for which different techniques have been proposed. Climate data providers have favored nonparametric approaches based on large ensembles of specific runs, while applied statisticians have preferred parametric models that leverage extreme value theory (EVT) (see, e.g., Davison & Huser 2015). In Section 5, we explain these approaches and identify their pros and cons. We also emphasize a few recent attempts to integrate either multi-errors (e.g., within a Bayesian hierarchical framework), or spatial dependence or multivariate EVT. In Section 6, we identify and discuss current challenges.

2. BASIC SETUPS

A major difficulty in EEA is the nonstationarity of the climate system, global warming being its most obvious feature. To deal with this lack of temporal stationarity, mainly due to anthropogenic forcings, climatologists have introduced two stationary worlds: the so-called factual and counterfactual worlds (Stott et al. 2004). The factual world contains the effect of human influence on the climate, while the counterfactual world does not. It is assumed that independent random draws from numerical experiments can be obtained for both worlds, and that the probability density

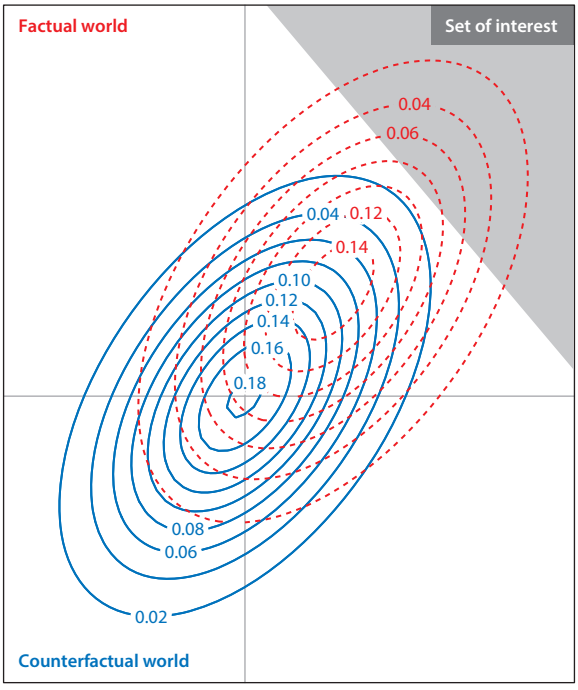

Figure 1

Classical scheme used in EEA: A counterfactual world (*solid contours* of the bivariate pdf) represents a world without any anthropogenic influence, and a factual world (*dotted contours* of the bivariate pdf) corresponds to our actual climate. For example, the *x*-axis and *y*-axis could represent mean decadal summer temperatures at two different locations. The main EEA question is to compare the probability of being in the set of interest (*shaded area*) with respect to the two worlds. In this figure, the event of interest could be the average temperature difference between the two locations being large. Abbreviations: EEA, extreme events attribution; pdf, probability density function.

function (pdf) of the variable of interest, e.g., mean seasonal temperatures, can be estimated. To help visualize this thought experiment, **Figure 1** displays bivariate pdf contours in the two different worlds, with blue contours for the counterfactual world and red contours for the factual one. One can imagine that the axes corresponds to two different locations, and the variable to some mean temperatures. It is clear that the mean has shifted from the (blue) counterfactual world toward a warmer (red) factual world. In this made-up example, a change in variability is also present, but less obvious. Concerning the class of events of interest, the EEA community typically focuses on sets defined by the upper right corner[1] of **Figure 1**. To move from the generic setup described by **Figure 1** to operational case studies, many choices have to be made.

Stott et al. (2004) published a seminal paper assessing how much human activities may have increased the risk of the occurrence of heat waves, like the 2003 heat wave observed over Europe. They did not look at a bivariate problem like the one in **Figure 1**, but focused on the mean June–August temperatures over a European region. Basically, a complex spatio-temporal temperature

[1]In this expository graph, we have chosen to display the bivariate framework. Although there is recent interest in compound events and multivariate modeling, most EEA studies have focused on univariate sets of interest, i.e., one variable being above a large threshold. Still, with **Figure 1**, we want to emphasize that the choice of the set is not universal in a multivariate context. Instead of $\{X_1 + X_2 > u\}$, one could choose $\{\min(X_1, X_2) > u\}$ or $\{\max(X_1, X_2) > u\}$, leading to different statistical modeling strategies.

field was reduced to its average in time and space, a single scalar random variable, X. In this case, the event of interest was the set $\{X > u\}$, where $u = 1.6$ Kelvin was chosen to mimic the 2003 mean European summer anomaly temperatures. The definition of X (its type, dimension in space and time, and so on) is paramount. This choice depends on the application at hand, and also on the data available. Before dealing with this delicate choice of X in a multivariate framework, it is important to introduce some notation and definitions to discuss the univariate setup. In this case, one key objective in most EEA analysis is the computation of small tail probabilities

$$p = \mathrm{P}(X > u),$$

where u represents a large threshold and X corresponds to a single summary of some atmospheric random field. Instead of setting u, one can impose the value of p and find u. In hydrology (e.g., Katz et al. 2002), this is equivalent to providing a return period, say T, and finding the return level u_T, defined by $\mathrm{P}(X > u_T) = 1/T$. For example, Luu et al. (2018) estimated 1-in-1,000-year events of daily autumn rainfall in the South of France, i.e., the random variable was daily precipitation quantities and the threshold u corresponded to a 1-in-1,000-year return period. At this stage, inference of such a small p or, equivalently, large u is classical in terms of statistical analysis. This estimation problem of large return periods, given a sample of independent and identically distributed (i.i.d.) copies of X, has been tackled by statisticians working on extreme value analysis (see, e.g., Davison & Smith 1990, Coles 2001, Beirlant et al. 2004, Davison & Huser 2015). Cooley et al. (2019) summarize the main elements of EVT applied to environmental sciences.

To contrast the potential differences between the factual and counterfactual worlds, it is natural to distinguish

$$p_0 = \mathrm{P}(X > u), \quad p_1 = \mathrm{P}(Z > u), \tag{1.}$$

where Z follows the factual pdf F and X represents the same random variable but in the counterfactual world, which consequently may have a different pdf, say G. To simplify notation and follow common practice in EEA, we drop u in p_0 and p_1, but these quantities obviously depend on u. Many authors (see the bibliography of Stott et al. 2016) have looked at two types of probability ratio. The fraction of attributable risk (FAR) and the risk ratio (RR) are defined as

$$\mathrm{FAR}(u) = 1 - \frac{p_0}{p_1}, \quad \mathrm{RR}(u) = \frac{p_1}{p_0}.$$

The FAR(u) has been interpreted as the contribution of the anthropogenic causal factor. The solid line in **Figure 2** shows FAR(u) in two typical setups. The left panel mimics mean temperature anomaly pdfs by assuming that X is standard Gaussian, $X \sim \mathrm{N}(0, 1)$, while Z follows a Gaussian distribution with mean 1 (i.e., one degree warmer) and standard deviation 1.5 (higher climate variability). In the right panel, by choosing a Pareto tail, $(1 + \xi x/\sigma)^{-1/\xi}$, with $\sigma = 1$ for X and $\sigma = 1.5$ for Z, we investigate daily precipitation behavior with a typical shape parameter $\xi = 0.2$. For both the Gaussian and Pareto cases, FAR(u) increases as u increases. The main difference is that, in the Gaussian setup, a rare event in the factual world (small p_1) would be nearly impossible in the counterfactual world (almost-zero p_0). This explains why FAR(u) goes to unity, as u increases, in the left panel of **Figure 2**. In contrast, FAR(u) tends to $1 - \sigma^{1/\xi} = 0.87$ in the Pareto case. Unlike the Gaussian case, in the Pareto case, a very rare event in the factual world remains possible in the counterfactual world. This simple example implies that the distributional assumptions made for average temperatures (rather light tails) could lead to different EEA conclusions for heavy-tailed atmospheric variables, like heavy rainfall, and will also affect inferences, especially in terms of confidence intervals.

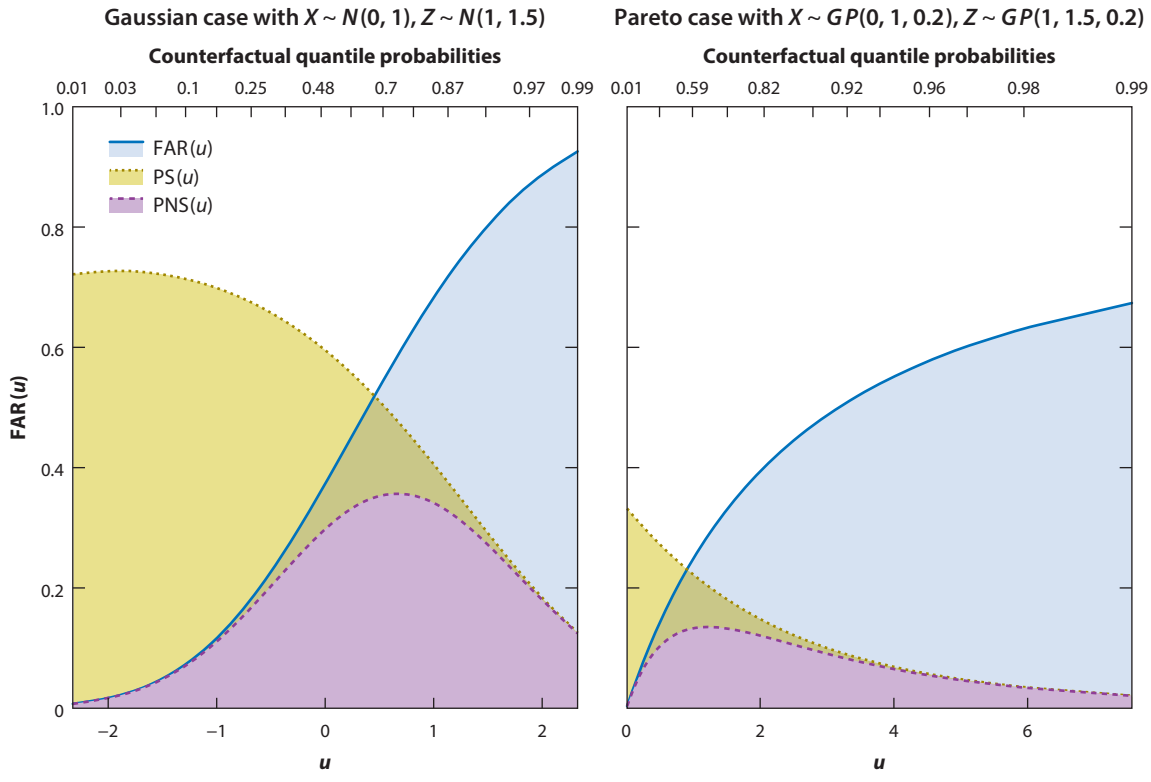

Figure 2

FAR(u) (*solid blue line*), probability of sufficient causation (*dotted dark yellow line*) and probability of sufficient and necessary causation (*dashed purple line*) in function of the threshold u (Equation 2). The left and right panels correspond to the Gaussian and Pareto cases, respectively. Abbreviations: FAR, fraction of attributable risk; GP, generalized Pareto; PNS, probability of necessary and sufficient causation; PS, probability of sufficient causation.

In classical EEA studies, factual simulations from a given numerical model have to be understood as samples from today's world, i.e., of the current climate, as opposed to transient simulations mimicking climate drifts over one or two centuries.[2] This implies that p_1 should have a temporal subscript indicating the current year of the event analysis. For notational simplicity, we do not use such a time indexing.[3] The EEA strategy recommended in the National Academies report (Natl. Acad. Sci. Eng. Med. 2016) is to compare the factual (today) and the counterfactual worlds, so the comparison is only valid for one fixed climatological instant.

In terms of interpretation, one may wonder if it is better to use FAR(u), or RR(u), or to simply look at p_1. Despite its apparent simplicity, the latter may be the most complicated, because the value of p_1, say, of observing a specific heat wave, will certainly be different in 2100 than today. From an impact point of view, it is unclear if long-term decisions (e.g., building dams or planting specific tree species) should be made with respect to today's climate or with respect to the climate of 2100, which will be different—but then, why 2100 specifically? Interpreting return levels and

[2]Global climate model runs, in contrast to EEA runs, aim to reproduce transient trajectories over long time periods and at the global scale. This explains why very few global runs are produced.

[3]The counterfactual world providing p_0 is quasi-stationary for the temporal scale of interest of EEA studies, given that natural forcings (such as changes in solar radiation or major volcanic eruptions) have limited impact at the decadal scale.

periods in nonstationary situations remains complicated (Rootzén & Katz 2013, Gilleland et al. 2017), and although it is a crucial point in other types of attribution analysis (Kim et al. 2016, Ribes et al. 2016), the statistical modeling of nonstationarities is rarely treated in EEA (a recent exception is Ribes et al. 2019).

As any study of EEA is intertwined with the question of attribution, and consequently rooted in the assessment of causality, it would be of interest to make a bridge between p_0 and p_1 and some type of causality. Hannart et al. (2016) were able to make a mathematical link between FAR(u) and the causal counterfactual probability theory of Pearl (2009). More precisely, increasing and external anthropogenic forcing can be understood as a monotone and exogenous intervention on the climate system. Running climate models with and without this intervention allows us to distinguish between the probabilities of sufficient causation (PS) and necessary causation (PN). In a general setup, these are difficult to compute, but the monotonicity and external nature of anthropogenic forcing simplify them to (Hannart et al. 2016):

$$\begin{aligned}&\text{probability of necessary causation: } \mathrm{PN}(u) = \max(\mathrm{FAR}(u), 0),\\ &\text{probability of sufficient causation: } \mathrm{PS}(u) = \max\left(1 - \frac{1-p_1}{1-p_0}, 0\right), \text{and}\\ &\text{probability of necessary and sufficient causation: } \mathrm{PNS}(u) = \max\left(p_1 - p_0, 0\right).\end{aligned} \quad 2.$$

From these definitions, it is clear that whenever $p_1 \geq p_0$, FAR(u) can be interpreted as the PN. For the special cases shown in **Figure 2**, PS(u) and PN(u) play opposite roles with respect to u. According to the dotted lines, sufficient causation (PS) is maximized for small values of u, but it goes to zero for large values of u, i.e., for extreme events. The reverse holds for necessary causation (PN, here represented by FAR). As one may expect, PNS(u) balances these two effects. The highest value for PNS(u) appears to be smaller for the Pareto case than the Gaussian case, which can be explained by recalling that any Gaussian random variable is very unlikely to have values outside three standard deviations from its mean, so it almost behaves like a random variable with finite support. In contrast, the support of a Pareto random variable cannot be considered finite. Large realizations in the factual world could also have happened in the counterfactual world. This effect diminishes whenever the Pareto shape parameter differs strongly between the factual and counterfactual worlds.

Before treating estimation, we need to review the different types of data available for EEA and their associated uncertainties.

3. CLIMATE DATA TYPES

Computing the counterfactual p_0 would require a world without anthropogenic forcing, but as human influence has been accelerating since the Industrial Revolution and few in situ observations are available before the Industrial Revolution, data from the counterfactual world are not directly accessible from observational records. For this reason, counterfactual samples in many EEA studies are produced in silico from numerical experiments. For example, the Met Office Hadley Centre for Climate Prediction and Research provided the simulations used by Stott et al. (2004). This leads to the question of the reliability of such numerical simulations and observational data.

3.1. Model Errors and Uncertainty Sources

In terms of statistical modeling, the Earth's climate, although an extremely complex dynamical system, is viewed as a spatio-temporal random process with respect to the variable of interest. This

stochastic representation can be applied to the three pillars of EEA: the factual and counterfactual worlds, say Z and X, and the true observed climate system, say Y. These three data sets are tainted by errors, and the following sources of uncertainty can be listed:

- natural internal variability
- model uncertainty from approximating the true climate system with numerical experiments
- observational uncertainties due to instrumental errors, homogenization problems, and mismatches between data sources, e.g., merging satellite data with observations by data assimilation
- sampling uncertainty in space and time
- statistical modeling error by assuming a specific statistical model, e.g., assuming a generalized extreme-value distribution for independent block maxima

The last three items of this list are commonplace for statisticians, so, we focus on the first two.

Even without any forcing variations, the true climate system is considered a chaotic dynamical system, which exhibits its own internal variability. This source of uncertainty, over decadal or longer timescales, can be interpreted as the inherent noise of a stationary system. In practice, the observed climate is always influenced by some forcings and is never at equilibrium, so a stationary version is never observable.

Numerical climate models are imperfect due to numerical approximations and parameterizations of a number of processes. Two numerical climate models from two different teams will produce different climatological pdfs under the same forcings. Modeling errors associated with their numerical and physical imperfections should be taken into account by introducing some climate modeling error. Thus, the Coupled Model Intercomparison Project (CMIP) of the World Climate Research Programme contains numerous types of experiments (including control runs with no change in forcings, idealized perturbations, historical simulations, etc.), different initial condition members (sampling variability) and different numerical models (model uncertainty). In EEA, assessment studies like that of Angélil et al. (2017) compare multiple methods and multiple studies (meta-analysis) (see their table 1). Paciorek et al. (2018) also list the different sources of uncertainty.

3.2. Data Types

There are three typical data sets in EEA: the observational database and the numerically simulated factual and counterfactual worlds. In many EEA numerical experiments, boundary conditions over a specific region are needed to simulate large ensemble runs. These given conditions (e.g., sea surface temperatures) are chosen with respect to the observed extreme event under study. By providing boundary conditions, the observational database indirectly drives the factual and counterfactual worlds. For example, Angélil et al. (2017) analyzed two ensembles of 390 members spanning the period January 2010–December 2013, using the American Community Atmosphere Model. Each realization was driven by the same external boundary conditions, but with different initial weather states to spread possible weather trajectories. This type of conditioning based on the extreme event [see also weather@home (**https://climateprediction.net**); Otto et al. 2018] constrains the simulation space and reduces confidence intervals, but it renders the result conditional on a given situation. In addition, the setup of the counterfactual runs is not trivial: Unobserved ocean temperatures and other conditions have to be set with respect to unknown preindustrial information. Finally, p_0 and p_1 become conditional probabilities:

$$p_{C_0} = \mathrm{P}(X > u|C_0), \quad p_{C_1} = \mathrm{P}(Z > u|C_1), \qquad 3.$$

where C_0 and C_1 are conditions based on the climate system state related to the event of interest.

In summary, numerical experiments based on boundary conditions involve a layer of arbitrary choices (e.g., how to find C_1 and C_0). This setup does not entirely answer the questions of stakeholders who are not interested by conditional return periods, but by unconditional ones. Some risks could even become less frequent under some atmospheric circulation patterns while becoming more frequent in general. EEA conditional simulations are application specific, are computationally expensive, and need high expertise (see, e.g., Otto et al. 2018). On the positive side, they can produce very large sample sizes, a key feature when dealing with extremes, especially for nonparametric approaches, and their spatial resolution can be high, a crucial point for some impact studies.

Another data repository is the aforementioned CMIP experiment. This database contains global simulations that have the advantages of being non-event-specific and unconditional. In addition, many other models from different research institutes are available, so numerical model uncertainties can be explored. Their main drawbacks are their small sample sizes and their spatial resolution, which can be too coarse for many applications. Another complication is the transient nature of these simulations, which implies that factual runs in CMIP contain some trend that must be taken into account in the statistical analysis. This issue was absent in classical EEA conditional simulations.

Surprisingly, at least for statisticians, extreme value analysis in attribution studies has been separated into two independent topics: event attribution and long-term trend attribution. Although the topic of interest, comparing high return periods from two different climatological setups, is identical, the event attribution and trend attribution research communities have been mostly working separately. For example, early work by Kharin & Zwiers (2000), who inferred relative changes in return levels and periods at the continental scale, was mostly viewed as outside of the realm of EEA. Among other differences, long-term trend studies attribute whether observations also exhibit a significant trend, while EEA provides frequency estimates typically derived from numerical models, often disregarding the historical record. Another difference is that trend attribution studies (see, e.g., Kharin & Zwiers 2005, Kharin et al. 2007) primarily consider the global scale, while EEA studies are triggered by a local event.

3.3. Variables of Interest and Their Distributions

The special report *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (Field et al. 2012) reviewed the climate literature on observed global- and regional-scale changes in climate extremes (see also the *Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker et al. 2013). From these reports, it is clear that various physical variables can be of interest for EEA. Quoting from Chen et al. (2018, p. 37),

> Typical extreme indices include the number or fraction of days with maximum temperature (Tmax), minimum temperature (Tmin), or precipitation, below the first, fifth, or tenth percentile, or above the 90th, 95th, or 99th percentile, as generally defined for given timeframes (e.g., days, months, seasons, annual) with respect to a reference period (e.g., 1961–1990). Other definitions relate to, for example, the number of days above specific absolute temperature or precipitation thresholds, or more complex definitions relate to the length or persistence of climate extremes.

Recent studies have also focused on surface wind (Vautard et al. 2019), extent of sea ice cover (Kirchmeier-Young et al. 2017), and fires (Kirchmeier-Young et al. 2019). For temperatures averaged in time and space, tracking changes in means and variances of Gaussian variables can be enough, but some important climate variables are not normally distributed. As they are always positive, skewed, and heavy tailed, the distributional features of precipitation differ from those

of temperatures. 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 (see, e.g., Kharin et al. 2007). Unlike for temperatures, both signal (warming climate) and noise (variability) vary with scale for precipitation. In this context, a natural way to move away from the normal distribution when looking at extremes is to apply EVT (see, e.g., Davison & Smith 1990, Coles 2001, Beirlant et al. 2004) (discussed further in Section 5.2).

4. EVENT DEFINITIONS

Most EEA studies are performed after witnessing an extreme occurrence. Typically, a very unusual event like the 2003 European heat wave or the United Kingdom flooding in autumn 2000 attracts a lot of media attention, causes major damage, and may even be responsible for deaths. Such high-impact events also trigger a series of research articles. For example, the *Bulletin of the American Meteorological Society* produces a yearly special report to assess how human-caused climate change may have affected the strength and likelihood of individual extreme events that have already occurred. The seventh edition of this report included analyses of marine heat waves in the Tasman Sea off of Australia in 2017 and 2018 and the summer 2017 heat wave in southern Europe (see, e.g., Kew et al. 2019).

There is no need to observe an extreme realization to compute the probability of extreme occurrences over a selected spatiotemporal domain. The choice of threshold, defining the extreme quantile under study, can also be made independently of any specific realization. For example, Kim et al. (2016) attributed extreme temperature changes using updated observations and multi-model CMIP data sets for an extended period of 1951–2010. A regression-based generalized extreme value (GEV) model was applied to block maxima for each grid point and at the global level.

In most EEA, the definition of the event itself in terms of duration and spatial footprint is physically based rather than statistically justified. Mathematically speaking, this step can be viewed as transforming a real-valued spatiotemporal random field, $Y(t,s)$, where the index t represents time and s the spatial component, into a scalar statistic, say

$$\overline{Y}(D,R) = \sum_{t\in D}\sum_{s\in R} Y(t,s), \qquad 4.$$

where D corresponds to a duration, usually from a few days to a few months, and R a region of a few hundred square kilometers. For example, Stott et al. (2004) chose to analyze a three-month event (June–August) over a region covering Europe and the Mediterranean Sea. This choice had an impact on the causal analysis. Cattiaux & Ribes (2018) scanned a large spatiotemporal domain and found that maximizing rarity within the factual world, i.e., minimizing p_1 in time and space, will highlight the regions most prone to extreme events (see their figures 2 and 3). As quoted by Cattiaux & Ribes (2018, p. 1562), "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." In other words, journalists and climatologists like to know the spatiotemporal scale that makes the event not rare, but exceptional. To apply their method, Cattiaux & Ribes estimated p_1 and p_0 from observations in the following way. The climate change component was estimated by smoothing spline and removed from temperatures recorded from 1950 to 2015. For a given D and R, they expressed Equation 4 as

$$\overline{Y}_i(D,R) = h(i) + A_i(D,R), \quad i \in \{1950,\ldots,2015\},$$

where $b(i)$ is obtained by averaging smoothing splines over R obtained from CMIP data and $A_i(D,R)$ corresponds to Gaussian noise for temperatures. The factual and counterfactual samples were then built, respectively, as

$$Z_i(D,R) = \overline{Y}_i(D,R) - [b(t) - b(1950)], \quad X_i(D,R) = \overline{Y}_i(D,R) - [b(t) - b(2015)],$$

by assuming that 1950 represents the counterfactual world and 2003 (the date of the event under scrutiny) the factual world. The optimization problem was to find

$$(\hat{D}, \hat{R}) = \text{argmin } \mathrm{P}(Z(D,R) > u(D,R)),$$

where the threshold $u(D,R)$ corresponds to the observed temperatures of the 2003 summer for the spatiotemporal region (D,R). Cattiaux & Ribes (2018) emphasized that their main objective was to offer a new blueprint for defining the event, and their approach could be improved, especially on the statistical modeling side. For instance, simple assumptions were used to estimate the occurrence probability p_1: The forced response simulated by models was assumed to be correct, and temperatures followed a Gaussian distribution; uncertainty on the computed p_1 was not taken into account, and the modeling framework did not cover multivariate events.

Another approach, motivated by the desire to evidence a causal link, was proposed by Hannart & Naveau (2018). By taking advantage of counterfactual theory, these authors proposed to maximize PNS defined by Equation 2, with $p_0 = \mathrm{P}(M\mathbf{X} > u)$ and $p_1 = \mathrm{P}(M\mathbf{Z} > u)$, and

$$(\hat{M}, \hat{u}) = \text{argmin } \mathrm{PNS}(u),$$

where u is a scalar, and $\mathbf{Z}$ and $\mathbf{X}$ represent multivariate vectors of size d concatenating the spatiotemporal data from the factual and counterfactual worlds, respectively. Here, the vector M of size d indicates that the aim is to find the best linear combination with respect to PNS. The event has to be understood as $\{M\mathbf{X} > u\}$, and not as $\{X(D,R) > u(D,R)\}$ like in Cattiaux & Ribes (2018). The linear projection, $M\mathbf{X}$, reduces the multivariate vector $\mathbf{X}$ into a univariate random variable, while ensuring that the projected data contain the most information in terms of causality. In the hierarchical Gaussian framework, Hannart & Naveau (2018) were able to integrate climate internal variability, model uncertainty, observational error, and sampling within a single model. They applied this approach to data from the HadCRUT4 observational data set for $\mathbf{Y}$, and runs of the IPSL-CM5A-LR model for the factual and counterfactual worlds, $\mathbf{Z}$ and $\mathbf{X}$, showed that anthropogenic forcings have virtually certainly caused the observed evolution of temperature at the global scale. More precisely, the probability that the observed evolution of temperature is not caused by anthropogenic forcings is then 1:10,000 instead of 1:20.

4.1. Events Based on Weather Types

The variability of the large-scale atmospheric circulation plays an essential role in EEA. Conditioning on the atmospheric circulation can help us to interpret the differences between the factual and counterfactual worlds. For example, Yiou et al. (2017) explained p_1 and p_0 as a function of a dynamical component and a thermodynamical one. They assumed that extreme values of a climate variable are generally reached for given patterns of atmospheric circulation. More precisely, given an observed circulation C, the authors introduced the notion of neighborhood circulation trajectories, defined as the distance from the event either to a known weather regime that is computed independently of the event itself, or to the observed trajectory of circulation. So, neighborhoods around C were defined in the counterfactual and factual worlds. Using Bayes' formula, the relative

ratio p_0/p_1 was decomposed as

$$\frac{\mathrm{P}(X > u)}{\mathrm{P}(Z > u)} = \rho_{\text{thermodynamical}} \times \rho_{\text{dynamical}},$$

with

$$\rho_{\text{thermodynamical}} = \frac{\mathrm{P}(X > u|C_1)}{\mathrm{P}(Z > u|C_0)}, \quad \rho_{\text{dynamical}} = \frac{\mathrm{P}(C_1)}{\mathrm{P}(C_0)} \times \frac{\mathrm{P}(C_0|Z > u)}{\mathrm{P}(C_1|X > u)}.$$

This approach was applied to a record precipitation event that hit the southern United Kingdom in January 2014. This is similar to the additive decomposition of Shepherd (2016), who also introduced a nondynamical term. Climatologically, such a decomposition helps practitioners to understand the event at hand and to focus the interpretation on specific circulation patterns. Statistically, some criticisms made in Section 3.2 can be repeated. In particular, some of the main caveats in EEA remain the uncertainty in the counterfactual world in terms of conditioning, how to define C_0, and the error associated with wrongly defining C_0.

4.2. Record Events

In event definitions, the basic ingredient has so far been the probability of exceedances, i.e., the probability that a well-chosen univariate random variable exceeds a large threshold. Still, many EEA studies have been motivated not by a large observed value, but by a record (see, e.g., King 2017). Breaking a record simply means that the current observation exceeds all past measurements (see, e.g., Resnick 1987). In this context, climatologists are often asked by media outlets if the frequency of record breaking has increased. Mathematically, the year r, say, in the counterfactual world, can be defined as a record year if

$$p_{0,r} = \mathrm{P}(X_r > \max(X_1, \ldots, X_{r-1})), \quad r = 1, 2, \ldots.$$

Naveau et al. (2018) noted that, if the counterfactual world time series can be assumed exchangeable, a reasonable hypothesis in the counterfactual world, then $p_{0,r}$ is exactly known, is equal to $1/r$, and does not need to be estimated. Hence, one can compare this probability with

$$p_{1,r} = \mathrm{P}(Z_r > \max(X_1, \ldots, X_{r-1})) \qquad 5.$$

to assess the chance that the factual observation Z_r is a record in the counterfactual world. The authors applied their approach to two data sets of annual maxima of daily temperature maxima: observations recorded in Paris and output from the Météo-France climate model. In the same way that no specific realizations need to be observed to compute changes in the frequencies of being above a threshold, one does not need to observe a record at year r to compute $p_{1,r}$ (in fact, it is likely that the record in the observed sample will not occur at year r). In general, EEA studies do not compare a specific realization occurrence, say, $X = x$, in the factual and counterfactual world but rather evaluate probabilities of an event like $X > u$ or like $Z_r > \max(X_1, \ldots, X_{r-1})$ for records.

5. STATISTICAL METHODS

In this section, we review and discuss common statistical models used in EEA. To see how these different techniques can be applied in practice, we refer to Yiou et al. (2019), who compared different approaches to the analysis of the Northern European summer heat wave of 2018.

5.1. Nonparametric Approaches

To our knowledge, there have been very few studies in EEA concerning the convergence of any type of nonparametric estimators of p_0 and p_1, especially with respect to the different uncertainties listed in Section 3.1. A popular current approach is modeling exceedance numbers as a simple binomial count, with a success corresponding to being above the threshold u, and computing the ratio of these counts in the factual and counterfactual worlds. Climatologists are clearly aware that large sample sizes are needed to estimate small probabilities in this case. For example, the weather@home experiment (Otto et al. 2018) produces thousands of samples (or more) of counterfactual and factual realizations, although these experiments are conditional. As the sample sizes are large, simply counting exceedances above a high threshold is valid, but it is computationally expensive if u is large. It is easy to study the properties of the natural estimator of the FAR[4]:

$$\widehat{\mathrm{FAR}}_n(u) = 1 - \left(\sum_{i=1}^{n} \mathbb{I}(X_i > u)\right) / \left(\sum_{i=1}^{n} \mathbb{I}(Z_i > u) + \frac{1}{2}\right),$$

where $\mathbb{I}(A)$ represents the indicator function, and $(X_1, \ldots, X_n)^T$ and $(Z_1, \ldots, Z_n)^T$ are our independent counterfactual and factual samples of size n. The fraction $1/2$ is there to avoid dividing by zero. For large n, $\widehat{\mathrm{FAR}}_n(u)$ is asymptotically unbiased and its distribution can be approximated by a Gaussian law (for a fixed u),

$$\sqrt{n}\left(\widehat{\mathrm{FAR}}_n(u) - \mathrm{FAR}(u)\right) \sim \mathrm{N}\left(0, \sigma^2(u)\right), \quad \sigma^2(u) = \frac{p_0^2(u)}{p_1^2(u)}\left[\frac{1 - p_0(u)}{p_0(u)} + \frac{1 - p_1(u)}{p_1(u)}\right]. \qquad 6.$$

This provides an asymptotic confidence interval at level α, $\widehat{\mathrm{FAR}}_n(u) \pm z_{1-\alpha/2}\frac{\hat{\sigma}(u)}{\sqrt{n}}$, where $z_{1-\alpha/2}$ represents the standard normal quantile at $1 - \alpha/2$ and $\hat{\sigma}(u)$ is the version of $\sigma(u)$ where $p_1(u)$ and $p_0(u)$ have been replaced by their empirical estimators. The drawback of nonparametric estimators like $\widehat{\mathrm{FAR}}_n(u)$ can be seen by considering the standard deviation $\sigma(u)$. Whenever $\lim np_0(u_n) = 0$ and $\lim p_0(u_n)/p_1(u_n) \in (0, 1]$ for large n, $\sigma(u_n)$ will explode, and the confidence interval will be useless. Consequently, the sample size n should always be much greater than the return period of interest. This is particularly relevant whenever the tail behavior of the random variable of interest, say, extreme rainfall, is heavy tailed, and becomes paramount if the time and spatial scales are fine. Practitioners are acquainted with such problems, and exhaustive numerical sensitivity analyses have been done in this context. To increase the sample size, one simple solution is to raise computer capacities and produce more simulations with appropriate boundary conditions. This brute-force option changes the estimation goal, i.e., inferring Equation 3 instead of Equation 1. To get confidence intervals for Equation 1 from estimators of Equation 3, one needs to model the error in estimating $\mathrm{P}(C_0)$ and $\mathrm{P}(C_1)$ and include this uncertainty in $\sigma(u)$.

Paciorek et al. (2018) compared ten statistical methods (Wilson's method, Koopman's asymptotic score test inversion, Wang-Shan exact test inversion, normal-theory with delta method, likelihood-ratio test inversion, bootstrap normal, percentile bootstrap, basic bootstrap, bootstrap-t, and adjusted percentile bootstrap) to build confidence intervals for the relative ratio p_1/p_0. When using nonparametric binomial counting, they recommended, based on numerical simulations, either the Koopman or Wang-Shan methods. They also concluded (Paciorek et al. 2018, p. 69) that "existing statistical methods not yet in use for event attribution have several advantages over the widely used bootstrap, including better statistical performance in repeated samples and robustness to small estimated probabilities."

[4]The same reasoning could be made for $\mathrm{RR}(u)$.

If one wants to move away from conditional analysis based on the numerical simulation of very large samples at a high computational cost, a possible avenue is to impose parametric forms on distributions. As p_0 and p_1 correspond to rare events, especially in the counterfactual world, EVT represents an attractive alternative to go beyond the range of the observations.

5.2. Univariate Extreme Value Theory

Davison & Huser (2015) gave an exhaustive review of the main modeling and estimation strategies to describe univariate and multivariate extremes. In the univariate case (e.g., Embrechts et al. 1997, Coles 2001), the probability that some random variable exceeding a well-chosen cutoff level v is larger than x can be approximated by a generalized Pareto (GP) tail defined as

$$\mathrm{P}(X > x | X > v) = \overline{H}_\xi \left(\frac{x - v}{\sigma} \right),$$

where $x > v, \sigma > 0$, and

$$\overline{H}_\xi(x) = \begin{cases} (1 + \xi x)_+^{-1/\xi}, & \xi \neq 0, \\ \exp(-x), & \xi = 0, \end{cases} \quad 7.$$

with $a_+ = \max(a, 0)$. The shape parameter ξ provides key information about the tail behavior. A zero shape parameter corresponds to an exponential tail, while $\xi > 0$ characterizes heavy tails and $\xi < 0$ implies a bounded support. For example, most daily rainfall (e.g., Katz et al. 2002) appears to be slightly heavy tailed. The GP tail has the important property of being invariant to thresholding. If X follows a GP distribution (GPD) with shape parameter ξ, then $X - v | X > v$ will also follow a GPD with the same shape parameter (the scale parameter σ will vary linearly with the threshold v). The GP is the only continuous distribution with this feature, which is why exceedances above a high threshold u are likely to follow a GPD (for theoretical aspects, see Beirlant et al. 2004, de Haan & Ferreira 2006). For similar reasons, the only distribution stable for rescaled maxima is the GEV:

$$\mathrm{GEV}(x; \mu, \sigma, \xi) = \exp\left(-\overline{H}_\xi \left(\frac{x - \mu}{\sigma} \right) \right), \quad 8.$$

where the new scalar μ corresponds to a location parameter. If the variable of interest corresponds to block maxima, say, annual maxima of daily rainfall maxima, a GEV fit can be tested. For exceedances, the GP should be preferred.

The GP and GEV distributions can be viewed as building blocks. For example, in the analysis of record frequency changes in annual maxima, Naveau et al. (2018) were able, by leveraging the GEV distribution, to simplify the FAR expression for records (under an easily testable hypothesis) into

$$\mathrm{far}(r) = 1 - \frac{p_{0,r}}{p_{1,r}} = (1 - \theta)\left(1 - \frac{1}{r}\right), \quad \theta = \frac{1}{\mathrm{E}(G(Z))} - 1,$$

where the single parameter θ, although a function of the GEV factual and counterfactual parameters, can be estimated without fitting GEVs.

There are many ways to estimate the GPD and GEV parameters. For example, likelihood-based approaches (e.g., Davison & Smith 1990), methods-of-moments (e.g., Hosking & Wallis 1987, Diebolt et al. 2008), or Bayesian inference (e.g., Coles & Tawn 1996) can be implemented. As noted by Davison & Huser (2015, p. 222), "the incorporation of external information

and borrowing strength across related time series may require a Bayesian approach; a Markov chain Monte Carlo approximation to the posterior density of the GEV parameters based on independent maxima is implemented in the R package `evdbayes`. Such methods are widely used in more complex problems (see, e.g., Cooley et al. 2007, Sang & Gelfand 2009, Shaby & Reich 2012, Reich et al. 2013)."

EVT techniques that could help to model the different sources of uncertainties in event attributions have not yet been fully implemented in EEA. Most EEA studies, via various numerical experiments, are based on simple binomial counting approaches, sometimes complemented by a GP fit with covariates (see, e.g., van der Wiel et al. 2017, van Oldenborgh et al. 2017). But even recent papers comparing statistical methods can bring confusing messages about EVT. For example, Paciorek et al. (2018, p. 71) wrote, "While EVA is often used for estimating probabilities of extreme events,. . .simple nonparametric estimators are often a good choice....We focus more on the binomial approach because of its greater generality." As the main goal of EVT is to extrapolate beyond the range of the data sample, a binomial approach will always put zero as the inferred value of p for extremes beyond the largest values. As already pointed out, the variance in the convergence in law for the nonparametric estimator of FAR(u) described by Equation 6 clearly indicates that a nonparametric approach will not work for very large u.

Three possible reasons can be invoked to explain why the EEA community has not relied on EVT in a systematic way and not yet developed complex EVT models. The first one is that a so-called extreme in EEA may not be considered an extreme event by the EVT community. The community focuses on very high quantiles—0.99, 0.999, or even 0.9999—e.g., 10,000-year return levels for nuclear safety design. In such a context, it is impossible to use a nonparametric approach to estimate very high quantiles, and one needs to apply the EVT basic principle, i.e., $\mathrm{P}(X > u) = \mathrm{P}(X > u | X > v) \times \mathrm{P}(X > v)$ for any large v, and any $u > v$ can be approximated by

$$\overline{H}_{\xi}\left(\frac{u-v}{\sigma}\right) \times \mathrm{P}(X > v).$$

Given estimates of σ and ξ obtained for the sampled data above the cutoff v, the above equivalence leads to the estimator

$$\hat{p}_0 = \overline{H}_{\hat{\xi}}\left(\frac{u-v}{\hat{\sigma}}\right) \times \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i > v).$$

Here, a nonparametric estimator, $\sum_{i=1}^{n} \mathbb{I}(X_i > v)$, has been used to estimate $\mathrm{P}(X > v)$ for a large but still moderate cutoff v, i.e., there are some observations above v. So, classical EVT techniques complement counting methods when interpolation above very large observations becomes impossible. Advocating a choice between the binomial and GP distributions (see, e.g., Paciorek et al. 2018) misses the inherent flexibility of the EVT approach that bridges both.

The second reason for the preference of basic binomial approaches over EVT in most EEA studies may be the delicate question of the cutoff choice v. If the threshold u is not too large and is comparable to v, then an EVT analysis may not bring strong added value. A large v ensures that the GP approximation is likely to be valid, but very few data points may be available for its fit. A smaller v provides more exceedances, but the GP approximation may be incorrect. Many papers discuss this (see, e.g., Bader et al. 2018). In particular, de Haan et al. (2015) offered a general theoretical framework to explore relative RRs in function of the convergence rate toward the GPD. The main objective of de Haan et al. (2015) was to detect trends, but the proposed methodology could be adapted to the EEA context. Another possibility is to bypass threshold choice completely and

replace it by a smooth transition. This approach has been applied to hourly and daily rainfall time series (see, e.g., Tencaliec et al. 2019, Naveau et al. 2016).

A third reason is the sample size. Depending on the data at hand, there is a tradeoff between the short length of instrumental time series (or small ensembles of existing global climate models) and large samples of conditional simulations. As previously mentioned, there has recently been a strong push to continuously generate numerical experiments over a small domain with prefixed boundary conditions. Although computationally expensive and tainted by error modeling uncertainties due to the choice of the domain size and of the boundary conditioning, these simulations can produce very large samples (see, e.g., weather@home in Otto et al. 2018). Provided that these data correctly mimic the factual and counterfactual worlds, moderately extreme conditional quantiles can be estimated by binomial count techniques.

In the past two years, some authors have tried to explore a different path based on nonstationary statistical methods. In spirit, the main idea is to replace large ensemble conditional simulations by transient data sets. For example, the existing CMIP database is readily available for a wide range of global coupled models with different forcing scenarios. These global models do not depend on a prefixed oceanic state. In addition, historical measurements can be used. Ribes et al. (2019) showed how event attribution can be implemented through the application of nonstationary statistics to transient simulations, typically covering the period from 1850 to 2100. They developed techniques for handling a multi-model synthesis but relied on a Gaussian assumption. Approaches based on transient global simulations go back to Kharin & Zwiers (2000, 2005) and Kharin et al. (2007), who even leveraged the GEV framework within a hierarchical model.

5.3. Hierarchical Modeling

To our knowledge, no comprehensive statistical models dealing with all identified sources of uncertainties have yet been developed within the EEA and EVT communities. Still, a promising avenue could be the hierarchical modeling of the different sources of uncertainties. In a multivariate Gaussian setup adapted to spatially averaged temperatures, Katzfuss et al. (2017) used a regression-based attribution Bayesian hierarchical approach to model the uncertainty in the true climate signal under different forcing scenarios and the uncertainty associated with estimating the climate variability covariance matrix. This Bayesian errors-in-variable model coupled with PCA truncations based on Bayesian model averaging enabled the authors to incorporate a large spectrum of uncertainties into inference on the regression coefficients (see, e.g., Hannart et al. 2014 for attribution errors-in-variable models). This approach does not address the question of estimating PN, PS, and PNS (see Equation 2). One step in this direction is Hannart & Naveau (2018), who, under the same type of linear model with a built-in hierarchy of uncertainties, were able to improve the quantification of causal probabilities. One key idea was to project the signal of interest into a linear subset in order to maximize the causal evidence. This strategy appears to bring a significant gain when the model is assumed to be normally distributed, but may be challenged when the variable of interest departs from normality.

5.4. Multivariate Extreme Value Theory

Another important step will be to integrate the latest multivariate EVT advances in the EEA context, because most studies in atmospheric sciences rely on independence. This latter is rarely tenable in either space or time. In addition, it will underestimate systemic risks and/or miss compound events. More precisely, let $\mathbf{Y} = (Y_1, \ldots, Y_d)^T$ denote a multivariate random vector with cumulative distribution function F and marginal cumulative distribution functions $F_1, \ldots, F_d$. In

the bivariate case, one could wonder how extreme events are correlated. The classical tail dependence coefficient χ measures the probability of $F_1(Y_1)$ being large given that $F_2(Y_2)$ is large as the quantile level q increases:

$$\chi = \lim_{q \uparrow 1} P\left[F_1(Y_1) > q \mid F_2(Y_2) > q\right]. \qquad 9.$$

To study such summaries, univariate theory based on an asymptotic argument (the block maxima length going to infinity) has to generalize to a multivariate framework (Coles 2001, Beirlant et al. 2004, de Haan & Ferreira 2006). To do so, let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{id})^T$, $i \in \{1, \ldots, n\}$, be n i.i.d. copies of $\mathbf{Y}$. Let $\mathbf{M}_n := (M_{n,1}, \ldots, M_{nd})$ with $M_{nj} := \max(Y_{1j}, \ldots, Y_{nj})$ for $j = 1, \ldots, d$. The cumulative distribution function F is said to be in the max-domain of attraction of an extreme-value distribution G if there exist sequences of normalizing constants $\mathbf{a}_n = (a_{n1}, \ldots, a_{nd})^T > \mathbf{0}$ and $\mathbf{b}_n = (b_{n1}, \ldots, b_{nd})^T$ such that

$$\mathrm{P}\left[\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \leq \mathbf{x}\right] = F^n(\mathbf{a}_n \mathbf{x} + \mathbf{b}_n) \to G(\mathbf{x}), \qquad n \to \infty. \qquad 10.$$

The convergence in Equation 10 implies that the margins $F_1, \ldots, F_d$ of F are in the max-domain of attraction of the univariate GEV extreme-value distributions $G_1, \ldots, G_d$. Let $\mathbf{l} = (l_1, \ldots, l_d)$ denote the vector of lower endpoints of G. If Equation 10 holds, we have

$$\max\left\{\frac{\mathbf{Y} - \mathbf{b}_n}{\mathbf{a}_n}, \mathbf{l}\right\} \mid \mathbf{Y} \not\leq \mathbf{b}_n \xrightarrow{d} \mathbf{V}, \qquad n \to \infty, \qquad 11.$$

where $\mathbf{V} = (V_1, \ldots, V_d)^T$ is said to follow a multivariate generalized Pareto distribution (MGPD) (Rootzén & Tajvidi 2006, Rootzén et al. 2018b). For $j \in \{1, \ldots, d\}$, the conditional random variables $V_j \mid V_j > 0$ are univariate GPDs with parameters $\sigma_j > 0$ and ξ_j. The properties of the multivariate GPD have been studied in detail by Tajvidi (1996), Rootzén & Tajvidi (2006), Falk & Guillou (2008), Ferreira & de Haan (2014), and Rootzén et al. (2018a, 2018b), while statistical modeling is quite recent (Huser et al. 2016, Kiriliouk et al. 2019). In terms of simulation and inference, a key property is that every standardized MGP vector $\mathbf{V}^*$ with $\sigma_j = 1$ and $\xi_j = 0$ can be represented as

$$\mathbf{V}^* = E + \mathbf{U} - \max_{1 \leq j \leq d} U_j, \qquad 12.$$

where E is a unit exponential random variable and $\mathbf{U}$ is a d-dimensional random vector independent of E. To illustrate the variety of extremes generated by such a model, **Figure 3** displays the case where $\mathbf{U}$ corresponds to a bivariate Gaussian random vector such that $U_1 - U_2 \sim \mathcal{N}(\beta_1 - \beta_2, \tau^2)$ for some $\tau > 0$. Given the same Gaussian marginal behavior with $\xi = (0, 0)$, $\sigma = (1, 1)$, $\beta = (0, 0)$, the three panels of **Figure 3** show the impact of τ on the dependence strength. In particular, the tail dependence coefficient (Equation 9) can be expressed as

$$\chi = 2\left(1 - \frac{1}{1 + 2e^{\tau^2/2}\Phi(-\tau)}\right). \qquad 13.$$

Φ is the standard Gaussian cumulative distribution function. From the three panels, one sees that, as χ increases, the dependence among the bivariate samples increases along the diagonal for extremes. More complex extremal dependence structures (asymmetric, etc.) can be obtained by modifying with the Gaussian marginal structure.

To explore how quantities like those in Expression 2 change in a multivariate context, one possibility is to project the multivariate MGPD vector into a one-dimensional object. If $\mathbf{V}$ follows

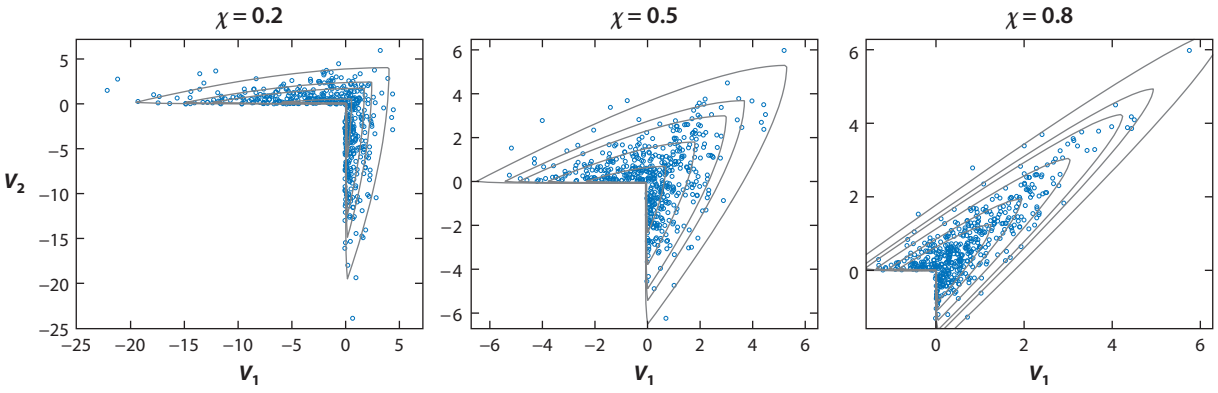

Figure 3

Scatter plots and density contours for the bivariate Gaussian multivariate generalized Pareto distribution model (Equation 12) with $\xi = (0, 0)$, $\sigma = (1, 1)$, $\beta = (0, 0)$. From left to right, the tail dependence coefficient (given by Equation 13) equals $\chi = 0.2$, $\chi = 0.5$, and $\chi = 0.8$. A large χ indicates a strong extremal dependence.

a multivariate GPD with marginal parameters σ and $\xi = (\xi, \ldots, \xi)^T$, then any linear projection with positive coefficients $\mathbf{w}$ satisfies $\mathbf{w}^T\mathbf{V} \mid \mathbf{w}^T\mathbf{V} > 0 \sim \text{GPD}(\mathbf{w}^T\sigma, \xi)$. This property was fundamental to compute and study probabilities of necessary and sufficient causation (Kiriliouk & Naveau 2019).

In terms of statistical inference and modeling, the recent work of de Fondeville and colleagues (see, e.g., de Fondeville & Davison 2018, Engelke et al. 2019, and R. de Fondeville & A.C. Davison, manuscript in preparation) could also help to build multivariate Pareto processes in time, space, or both.

A drawback of MGPD modeling is that, in some important cases like the Gaussian bivariate situation, the coefficient χ is equal to zero, so the dependence is then hidden in the rate of convergence to zero. To handle this so-called asymptotic independence, Ledford & Tawn (1996, 1997) measured the speed of the decay toward independence at high levels via the coefficient of tail dependence, $\eta \in (0, 1]$ under the model

$$P\left[Y_1 > y \mid Y_2 > y\right] \sim L(y)y^{1-1/\eta},$$

where $L(y)$ is slowly varying, i.e., $L(vy)/L(y) \to 1$ for any $v > 0$ and large y. If $\eta = 1$, then the variables are asymptotically dependent and asymptotically independent otherwise. As stated by Davison & Huser (2015, p. 217), "Hence, if an asymptotic dependence model is wrongly assumed to be valid, probabilities of extremely rare events will be overestimated, and conversely, potentially leading to serious misestimation of risk....In practice, however, the coefficient of tail dependence is difficult to estimate, as it relates to the joint behavior of the data at infinity. Thus, careful assessment of the plausibility of asymptotic independence is required." Still, the class of asymptotic independent models offers additional tools that have yet to be used in a EEA context.

6. CONCLUSIONS

From a statistical point of view, the field of EEA represents rich territory. It combines various sources of uncertainties and data sets. To enhance causality and to reduce the noise/signal ratio, the need for appropriate definitions of the extreme event of interest is essential. The main difficulty is that the classical setup used by EVA statisticians in environmental studies, one data set

and one statistical model, can be too narrow. A research push to frame a generic blueprint (maybe a Bayesian hierarchical model or a meta-analysis) that integrates different data sources (transient and conditional simulations, numerical and observational data sets, etc.) appears necessary. The use of numerical models and reanalysis products in the climate community, with higher accuracy in terms of the spatial and temporal resolution, will likely increase in the coming years. Still, as the object of interest is the analysis of rare and even very rare events, the computational costs will stay high and the extreme value community should invent statistical concepts and tools to deal with the estimation of very small p_0 and p_1. In particular, if p_1 is very small, then the corresponding event may be nearly impossible in the counterfactual world, and one has to model extremely rare multivariate events in the counterfactual setting. Another aspect of causality is the setup of climate numerical experiments with respect to multiple factors. As global climate model runs are expensive in terms of resources and computational time, few numerical experiments can be done in CMIP. From a counterfactual theory point of view, we advocate all-but-one simulations. In terms of estimation, only relying on binomial counts may limit the study of very rare events (going beyond the largest values). Recent multivariate extreme value advances could help to tackle this issue in a spatio-temporal context.

The following list highlights some challenges and points that need more statistical development:

- Combining all known sources of uncertainties for extreme climate event analysis is an area for further development. For example, Bayesian model averaging for multivariate extremes (see, e.g., Sabourin et al. 2013) could be a first step. Extending some of the ideas of Gaussian hierarchical modeling of Katzfuss et al. (2017) and Hannart & Naveau (2018) to a multivariate EVA framework could be another direction. Hammerling et al. (2019) laid solid foundations to explore this in the attribution context.
- Developing statistical strategies to automatically define the extreme event, i.e., relevant spatio-temporal scales of interest with a given nonstationary random field, remains a key and fundamental problem in EEA.
- Finding a needle in a haystack, i.e., finding some particular event with low probability of occurrence (but with important impacts) may be particularly challenging in a nonstationary climate system.
- Summarizing the uncertainty of p_0 and p_1 with one number. This is particularly important for impact studies and communication with stakeholders.

Future issues can complement this list, and some statistical techniques could be leveraged to address them.

- Compound events (such as simultaneous precipitation deficit and high temperatures), concomitant and concurrent extremes (see, e.g., Zscheischler et al. 2018) may have a strong impact in hazard and risk assessment. This combination of causes could lead to underestimation of risk because drivers of extreme events often interact in space and time. There is a large need for research development in this area.
- Return times of compound events can be tricky to interpret and can lead to misleading risk assessments. Serinaldi (2016) listed important points to reduce such misinterpretation.
- In statistical environmental studies, the modeling of subasymptotic multivariate extremes described in Section 5.4 has been coupled with the idea of conditional extremal models, i.e., modeling of the multivariate vector given that one component is large (see, e.g., Tawn et al. 2018, Huser & Wadsworth 2019, Shooter et al. 2019). This regression-based approach

for extremes, to our knowledge, has not been used in EEA. This could offer new ways to improve causality assessment.

- Instead of sampling all trajectories with a numerical climate model, there has been recent interest in coupling large deviation algorithms within the climate numerical code itself. For example, Ragone et al. (2018) developed rare event algorithms to compute probabilities of events that could not be observed in direct numerical simulations. This strategy reduced the computational cost of expensive numerical models for the study of heat waves in climate models.
- Model misspecification and multiple testing in space (see, e.g., Risser et al. 2019) have rarely been addressed by the EEA community and further development of these tools is needed.
- Another line of approach, not addressed in this review, is the physical explanation of the event itself, i.e., evidencing the causal mechanistic chain of causation (Trenberth et al. 2015). How to integrate such an approach within statistical reasoning remains an open question.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We would like to thank Soulivanh Thao and Julien Cattiaux for their helpful comments and discussions. Part of this work was supported by the French national program FRAISE-LEFE/INSU, Melody-PRC-ANR, Eupheme, FUSIMET (PEPS I3A), DAMOCLES-COST-ACTION on compound events. Finally, the insightful remarks from the reviewer have greatly improved the scope and the clarity of this work.

LITERATURE CITED

Angélil O, Stone D, Wehner M. 2017. An independent assessment of anthropogenic attribution statements for recent extreme temperature and rainfall events. *J. Clim.* 30:5–16

Bader B, Yan J, Zhang X. 2018. Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *Ann. Appl. Stat.* 12:310–29

Beirlant J, Goegebeur Y, Segers J, Teugels J. 2004. *Statistics of Extremes: Theory and Applications*. New York: Wiley

Cattiaux J, Ribes A. 2018. Defining single extreme weather events in a climate perspective. *Bull. Am. Meteorol. Soc.* 99:1557–68

Chen Y, Moufouma-Okia W, Masson-Delmotte V, Zhai P, Pirani A. 2018. Recent progress and emerging topics on weather and climate extremes since the fifth assessment report of the Intergovernmental Panel on Climate Change. *Annu. Rev. Environ. Resour.* 43:35–59

Coles SG. 2001. *An Introduction to Statistical Modeling of Extreme Values*. New York: Springer

Coles SG, Tawn JA. 1996. A Bayesian analysis of extreme rainfall data. *Appl. Stat.* 45:463–78

Cooley D, Hunter BD, Smith RL. 2019. Univariate and multivariate extremes for the environmental sciences. In *Handbook of Environmental and Ecological Statistics*, ed. AE Gelfand, M Fuentes, JA Hoeting, RL Smith. Boca Raton, FL: Chapman & Hall/CRC

Cooley D, Nychka D, Naveau P. 2007. Bayesian spatial modeling of extreme precipitation return levels. *J. Am. Stat. Assoc.* 102:824–40

Davison AC, Huser RG. 2015. Statistics of extremes. *Annu. Rev. Stat. Appl.* 2:203–35

Davison AC, Smith RL. 1990. Models for exceedances over high thresholds (with comments). *J. R. Stat. Soc. B* 52:393–442

de Fondeville R, Davison AC. 2018. High-dimensional peaks-over-threshold inference. *Biometrika* 105:575–92
de Haan L, Ferreira AF. 2006. *Extreme Value Theory: An Introduction*. New York: Springer
de Haan L, Tank A, Neves C. 2015. On tail trend detection: modeling relative risk. *Extremes* 18:141–78
Diebolt J, Guillou A, Naveau P, Ribereau P. 2008. Improving probability-weighted moment methods for the generalized extreme value distribution. *REVSTAT Stat. J.* 6:33–50
Embrechts P, Klüppelberg C, Mikosch T. 1997. *Modelling Extremal Events for Insurance and Finance*. New York: Springer
Engelke S, de Fondeville R, Oesting M. 2019. Extremal behaviour of aggregated data with an application to downscaling. *Biometrika* 106:127–44
Falk M, Guillou A. 2008. Peaks-over-threshold stability of multivariate generalized Pareto distributions. *J. Multivar. Anal.* 99:715–34
Ferreira AF, de Haan L. 2014. The generalized Pareto process; with a view towards application and simulation. *Bernoulli* 20:1717–37
Field C, Barros V, Stocker T, Qin D, Dokken D, et al. 2012. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge Univ. Press
Gilleland E, Katz RW, Naveau P. 2017. Quantifying the risk of extreme events under climate change. *CHANCE* 30:30–36
Hammerling D, Katzfuss M, Smith RL. 2019. Climate change detection and attribution. In *Handbook of Environmental and Ecological Statistics*, ed. AE Gelfand, M Fuentes, JA Hoeting, RL Smith. Boca Raton, FL: Chapman & Hall/CRC
Hannart A, Naveau P. 2018. Probabilities of causation of climate changes. *J. Clim.* 31:5507–24
Hannart A, Pearl J, Otto FEL, Naveau P, Ghil M. 2016. Counterfactual causality theory for the attribution of weather and climate-related events. *Bull. Am. Meteorol. Soc.* 97:99–110
Hannart A, Ribes A, Naveau P. 2014. Optimal fingerprinting under multiple sources of uncertainty. *Geophys. Res. Lett.* 41:1261–68
Hosking JRM, Wallis JR. 1987. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* 29:339–49
Huser RG, Davison AC, Genton MG. 2016. Likelihood estimators for multivariate extremes. *Extremes* 19:79–103
Huser RG, Wadsworth JL. 2019. Modeling spatial processes with unknown extremal dependence class. *J. Am. Stat. Assoc.* 114:434–44
Katz RW, Parlange MB, Naveau P. 2002. Statistics of extremes in hydrology. *Adv. Water Resour.* 25:1287–304
Katzfuss M, Hammerling D, Smith RL. 2017. A Bayesian hierarchical model for climate change detection and attribution. *Geophys. Res. Lett.* 44:5720–28
Kew SF, Philip SY, van Oldenborgh GJ, van der Schrier G, Otto FEL, Vautard R. 2019. The exceptional summer heat wave in Southern Europe 2017. *Bull. Am. Meteorol. Soc.* 100:S49–53
Kharin VV, Zwiers FW. 2000. Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. *J. Clim.* 13:3760–88
Kharin VV, Zwiers FW. 2005. Estimating extremes in transient climate change simulations. *J. Clim.* 18:1156–73
Kharin VV, Zwiers FW, Zhang X, Hegerl GC. 2007. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Clim.* 20:1419–44
Kim YH, Min SK, Zhang X, Zwiers F, Alexander LV, et al. 2016. Attribution of extreme temperature changes during 1951–2010. *Clim. Dyn.* 46:1769–82
King AD. 2017. Attributing changing rates of temperature record breaking to anthropogenic influences. *Earth's Future* 5:1156–68
Kirchmeier-Young MC, Gillett NP, Zwiers FW, Cannon A, Anslow F. 2017. Attribution of extreme events in arctic sea ice extent. *J. Clim.* 30:553–71
Kirchmeier-Young MC, Zwiers FW, Gillett N. 2019. Attribution of the influence of human-induced climate change on an extreme fire season. *Earth's Future* 7:2–10

Kiriliouk A, Naveau P. 2019. Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory. arXiv:1908.03107 [stat.AP]
Kiriliouk A, Rootzén H, Segers J, Wadsworth JL. 2019. Peaks over thresholds modelling with multivariate generalized Pareto distributions. *Technometrics* 61:123–35
Ledford A, Tawn JA. 1996. Statistics for near independence in multivariate extreme values. *Biometrika* 83:169–87
Ledford A, Tawn JA. 1997. Modelling dependence within joint tail regions. *J. R. Stat. Soc. B* 59:475–99
Luu LN, Vautard R, Yiou P, van Oldenborgh GJ, Lenderink G. 2018. Attribution of extreme rainfall events in the South of France using EURO-CORDEX simulations. *Geophys. Res. Lett.* 45:6242–50
Naveau P, Huser RG, Ribereau P, Hannart A. 2016. Modelling jointly low, moderate and heavy rainfall intensities without a threshold selection. *Water Resour. Res.* 52:2753–69
Naveau P, Ribes A, Zwiers FW, Hannart A, Tuel A, Yiou P. 2018. Revising return periods for record events in a climate event attribution context. *J. Clim.* 31:3411–22
Natl. Acad. Sci. Eng. Med. 2016. *Attribution of Extreme Weather Events in the Context of Climate Change*. Washington, DC: Natl. Acad.
Otto F, Philip S, Kew S, Li S, King A, Cullen H. 2018. Attributing high-impact extreme events across timescales—a case study of four different types of events. *Clim. Change* 149:399–412
Paciorek CJ, Stone DA, Wehner MF. 2018. Quantifying statistical uncertainty in the attribution of human influence on severe weather. *Weather Clim. Extrem.* 20:69–80
Pearl J. 2009. *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge Univ. Press. 2nd ed.
Ragone F, Wouters J, Bouchet F. 2018. Computation of extreme heat waves in climate models using a large deviation algorithm. *PNAS* 115:24–29
Reich BJ, Shaby BA, Cooley D. 2013. A hierarchical model for serially dependent extremes: a study of heat waves in the western US. *J. Agric. Biol. Environ. Stat.* 19:119–35
Resnick SI. 1987. *Extreme Values, Regular Variation, and Point Processes*. New York: Springer
Ribes A, Thao S, Cattiaux J. 2019. *Describing the relationship between a weather event and climate change: a new statistical approach*. Work. Pap. hal-02122780. **https://hal.archives-ouvertes.fr/hal-02122780/document**
Ribes A, Zwiers FW, Azaïs JM, Naveau P. 2016. A new statistical approach to climate change detection and attribution. *Clim. Dyn.* 48:29–56
Risser MD, Paciorek CJ, Stone DA. 2019. Spatially dependent multiple testing under model misspecification, with application to detection of anthropogenic influence on extreme climate events. *J. Am. Stat. Assoc.* 114:61–78
Rootzén H, Katz RW. 2013. Design life level: quantifying risk in a changing climate. *Water Resour. Res.* 49:5964–72
Rootzén H, Segers J, Wadsworth JL. 2018a. Multivariate generalized Pareto distributions: parametrizations, representations and properties. *J. Multivar. Anal.* 165:117–31
Rootzén H, Segers J, Wadsworth JL. 2018b. Multivariate peaks over thresholds models. *Extremes* 21:115–45
Rootzén H, Tajvidi N. 2006. Multivariate generalized Pareto distributions. *Bernoulli* 12:917–30
Sabourin A, Naveau P, Fougères AL. 2013. Bayesian model averaging for multivariate extremes. *Extremes* 16:325–50
Sang H, Gelfand A. 2009. Hierarchical modeling for extreme values observed over space and time. *Environ. Ecol. Stat.* 16:407–26
Serinaldi F. 2016. Can we tell more than we can know? The limits of bivariate drought analyses in the United States. *Stoch. Environ. Res. Risk Assess.* 30:1691–704
Shaby BA, Reich BJ. 2012. Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland. *Environmetrics* 23:638–48
Shepherd TG. 2016. A common framework for approaches to extreme event attribution. *Curr. Clim. Change Rep.* 2:28–38
Shooter R, Ross E, Tawn JA, Jonathan P. 2019. On spatial conditional extremes for ocean storm severity. *Environmetrics* 30:e2562

Stocker T, Qin D, Plattner G, Tignor M, Allen S, et al. 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge Univ. Press

Stott PA, Christidis N, Otto FEL, Sun Y, Vanderlinden JP, et al. 2016. Attribution of extreme weather and climate-related events. *WIREs Clim. Change* 7:23–41

Stott PA, Stone DA, Allen MR. 2004. Human contribution to the European heatwave of 2003. *Nature* 432:610–13

Tajvidi N. 1996. *Characterisation and some statistical aspects of univariate and multivariate generalized Pareto distributions*. PhD Thesis, Dep. Math., Chalmers Univ., Göteborg

Tawn JA, Shooter R, Towe R, Lamb R. 2018. Modelling spatial extreme events with environmental applications. *Spat. Stat.* 28:39–58

Tencaliec P, Favre A, Naveau P, Prieur C, Nicolet G. 2019. Flexible semiparametric generalized Pareto modeling of the entire range of rainfall amount. *Environmetrics*. In press. **https://doi.org/10.1002/env.2582**

Trenberth KE, Fasullo JT, Shepherd TG. 2015. Attribution of climate extreme events. *Nat. Clim. Change* 5:725–30

van der Wiel K, Kapnick SB, van Oldenborgh GJ, Whan K, Philip S, et al. 2017. Rapid attribution of the August 2016 flood-inducing extreme precipitation in south Louisiana to climate change. *Hydrol. Earth Syst. Sci.* 21:897–921

van Oldenborgh GJ, van der Wiel K, Sebastian A, Singh R, Arrighi J, et al. 2017. Attribution of extreme rainfall from Hurricane Harvey, August 2017. *Environ. Res. Lett.* 12:124009

Vautard R, van Oldenborgh GJ, Otto FEL, Yiou P, de Vries H, et al. 2019. Human influence on European winter wind storms such as those of January 2018. *Earth Syst. Dyn.* 10:271–86

Yiou P, Cattiaux J, Faranda D, Kadygrov N, Jézéquel A, et al. 2019. Analyses of the Northern European summer heatwave of 2018. *Bull. Am. Meteorol. Soc.* In press

Yiou P, Jézéquel A, Naveau P, Otto FEL, Vautard R, Vrac M. 2017. A statistical framework for conditional extreme event attribution. *Adv. Stat. Climatol. Meteorol. Oceanogr.* 3:17–31

Zscheischler J, Westra S, van den Hurk BJJM, Seneviratne SI, Ward PJ, et al. 2018. Future climate risk from compound events. *Nat. Clim. Change* 8:469–77