



HAL
open science

Statistical Methods for Extreme Event Attribution in Climate Science

Philippe Naveau, Alexis Hannart, Aurélien Ribes

► **To cite this version:**

Philippe Naveau, Alexis Hannart, Aurélien Ribes. Statistical Methods for Extreme Event Attribution in Climate Science. Annual Reviews of Statistics and its Application, 2020, 7, pp.89-110. 10.1146/annurev-statistics-031219-. hal-02408516

HAL Id: hal-02408516

<https://hal.science/hal-02408516>

Submitted on 3 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Statistical methods for extreme event attribution in climate science

Philippe Naveau¹ | Alexis Hannart² | Aurélien Ribes³

¹Laboratoire de Sciences du Climat et de l'Environnement, IPSL-CNRS, Gif-sur-Yvette, 91191, France

²Ouranos, Montréal, Canada

³Meteo France, Toulouse, France

Correspondence

Philippe Naveau, Laboratoire de Sciences du Climat et de l'Environnement, IPSL-CNRS, Gif-sur-Yvette, 91191, France
Email: philippe.naveau@ipsl.lscce.fr

Funding information

Changes in the Earth's climate have been increasingly observed. Assessing how likely each of these changes have been caused by human influence is important for decision making on mitigation and adaptation policy. Due to their high societal and economical impacts, media attention has been particularly focused on whether extreme events have become more frequent and and more intense, if so, why? To answer such questions, extreme event attribution (EEA) has tried to estimate extreme event likelihoods under different scenarios. Statistical methods and experimental designs based on numerical models have been developed, tested and applied during the last decade. In this paper, we review the basic probability schemes, inference techniques and statistical hypothesis used in EEA. To implement EEA analysis, the climate community relies on the use of large ensemble of climate models runs. From a statistical perspective, we discuss 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. These include 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.

KEYWORDS

Causality theory, detection and attribution, extreme value theory

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., Angéilil 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 specificity 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, a EEA study is driven by a strong societal need (e.g., a specific flooding or heatwave) 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, there has been a series of review/assessment articles (e.g., Chen et al., 2018; Angéilil et al., 2017; National Academies of Sciences, Engineering and Medicine, 2016; Stott et al., 2016; Shepherd, 2016) that describe and summarize the main climatological results and challenges about EEA, but these articles did not focus on the methodological side of the statistical modeling, and have preferred to concentrate on climatological advances, pitfalls and identified challenges in terms of climate impact modelling. 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, one goal of this article is to provide to any statistician, not well versed in geosciences but interested by this topic, the main tools in terms of notation, hypotheses and statistical models. In this context of describing the EEA statistical blueprint, the so-called story-line approach based on decomposing possible causal pathways to explain a specific realization (Trenberth et al., 2015) will not be covered.

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 recall the type 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 non-parametric 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 and Huser, 2015). In Section 5, we explain these approaches and identify their pros and cons. We also emphasize a few recent attempts to either integrate multi-errors (e.g., within a Bayesian hierarchical framework), or spatial dependence or multivariate extreme value theory. In Section 6, a list of current challenges is identified and discussed.

2 | BASIC SETUPS

A major difficulty in EEA is the non-stationarity 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 pdf of the variable of interest, e.g., mean seasonal temperatures, can be estimated. To visualise 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 corner¹ of Figure 1. To move from the generic setup

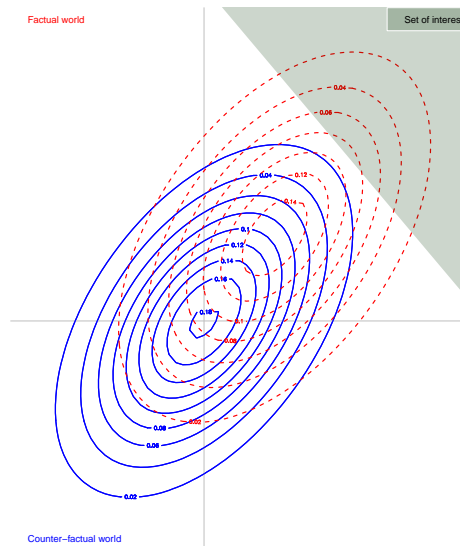


FIGURE 1 Classical scheme used in EEA: A counterfactual world (solid contours of the bivariate pdf) represents a world without any anthropogenic influence and a factual world (dotted contours of the bivariate pdf) corresponds to our actual climate. For example, the x-axis and y-axis could represent mean decadal summer temperatures at two different locations. The main EEA question is to compare the probability of being in the set of interest (greenish area) with respect to the two worlds. In this figure, the event of interest could be the average temperature between the two locations being large.

¹In this expository graph, we have chosen to display the bivariate framework. Although there is a 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.

described by Figure 1 to operational case studies, many choices have to be made.

Stott et al. (2004) published a seminal paper assessing by how much human activities may have increased the risk of the occurrence of heatwaves, like the 2003 heatwave observed over Europe. They did not look at a bivariate problem like in Figure 1, but focused on the mean June-August temperatures over a European region. Basically, a complex spatio-temporal temperature field was reduced to its average in time and space, a single scalar random variable, X . In this case, the so-called 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

$$p = P(X > u),$$

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 (see, 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 $P(X > u_T) = 1/T$. For example, Luu et al. (2018) estimated 1-in-1,000-year events of daily Fall 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 a large u , is classical in terms of statistical analysis. Given a sample of i.i.d. copies of X , this estimation problem of large return periods has been tackled by statisticians working on extreme value analysis (see, e.g., Davison and Smith, 1990; Coles, 2001; Beirlant et al., 2004; Davison and Huser, 2015). Cooley et al. (2019) summarizes 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

$$p_0 = P(X > u), \quad p_1 = P(Z > u), \quad (1)$$

where Z follows the factual pdf F and X represents the same random variable but in the counterfactual world and 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 so-called fraction of attributable risk (FAR) and the risk ratio (RR) are defined as

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

The $\text{FAR}(u)$ has been interpreted as the contribution of the anthropogenic causal factor. The solid line in Figure 2 shows $\text{FAR}(u)$ in two typical setups. The left panel mimics mean temperature anomaly pdfs by assuming that X is standard Gaussian, $X \sim N(0, 1)$ 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, $\text{FAR}(u)$ increases as u increases. The main difference is that, in the Gaussian setup, a rare event in the factual world (p_1 small) would be nearly impossible in the counterfactual world (p_0 almost zero). This explains why $\text{FAR}(u)$ goes to unity, as u increases, in the left panel of Figure 2. In contrast, $\text{FAR}(u)$ 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

Gaussian case with $X \sim N(0, 1), Z \sim N(1, 1.5)$

Pareto case with $X \sim GP(0, 1, .2), Z \sim GP(0, 1.5, .2)$

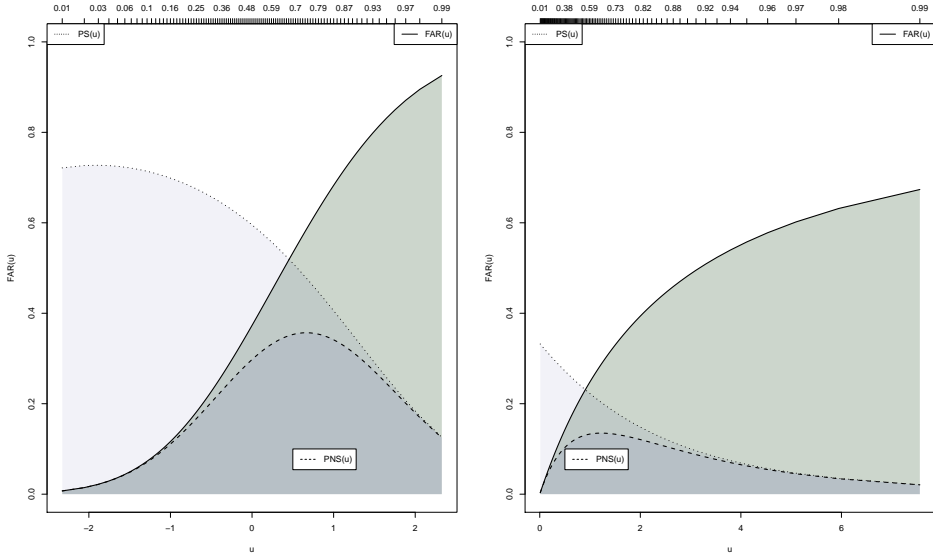


FIGURE 2 $FAR(u)$ (solid line), probability of sufficient causation (dotted lines) and probability of sufficient and necessary causation (dashed lines) in function of the threshold u , see Equation (2). The left and right panels correspond to the Gaussian and Pareto cases, respectively.

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.

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 drifts² over one or two centuries. 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 indexing³. The EEA strategy recommended in the National Academies of Sciences, Engineering and Medicine (2016) report 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 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 heatwave, will be certainly 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 the respect to the climate of 2100 that will be different. But then, why 2100? Interpreting return levels and periods in non-stationary situations remains complicated (Rootzén and 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 non-stationarities is

²Global climate model runs are different from EEA runs in the sense that the former aim at reproducing transient trajectories over long time periods and at the global scale. This explains why very few global runs are produced, while counterfactual simulations often consist of thousands of runs.

³The 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.

rarely treated in EEA (Ribes et al., 2019, is a recent exception).

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 of 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)

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

From these definitions, it is clear that, whenever $p_1 \geq p_0$, $FAR(u)$ can be interpreted as the probability of necessary causation. For the special cases described in Figure 2, $PS(u)$ and $PN(u)$ play opposite roles with respect to u . According to the dotted lines, sufficient causation 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, see the solid lines. As one may expect, $PNS(u)$, the probability of necessary and sufficient causation, balances these two effects. One can also notice that the highest value for $PNS(u)$ appears to be smaller for the Pareto case than the Gaussian case. This 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

To compute the counterfactual p_0 , one would need a world without anthropogenic forcing, but, as human influence has been accelerating since the industrial revolution⁴, 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 a numerical experiment. For example, the Hadley Centre provided the simulations used in 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,

⁴and few in situ observations are available before the industrial revolution.

say Y . These three datasets 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 as a chaotic dynamical system, which exhibits its own internal variability. This source of uncertainty, over decadal or longer time scales, 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 parametrizations 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 to their numerical and physical imperfections should be taken into account by introducing some climate modeling error. This explains why the Coupled Model Intercomparison Project (CMIP) contains not one but 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 Angéilil et al. (2017) compared multiple methods and multiple studies (meta-analysis); see their Table 1. Paciorek et al. (2018) also listed the different sources of uncertainty.

3.2 | Data types

There are three typical datasets in EEA: the observational database, and the numerically simulated factual and counterfactual worlds. In numerous 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 boundaries conditions, the observational database indirectly drives the factual and counterfactual worlds. For example, Angéilil 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 climateprediction.net (Weather@home) in Otto et al., 2018) constrains the simulation space and reduces confidence intervals, but 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

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

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 involves a layer of arbitrary choices (e.g., how

to find C_1 and C_0). This setup does not entirely answer to the question of some 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, 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 non-parametric 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, a lot of 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 object of interest, comparing high return periods from two different climatological setups, was identical, the event attribution and trend attribution research communities have been mostly working separately. For example, early work by Kharin and Zwiers (2000) who inferred relative changes in return levels and periods at the continental scale, were 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 and 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 on Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation, (Field et al., 2012) has 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), *typical extreme indices include the number or fraction of days with maximum temperature (T_{max}), minimum temperature (T_{min}), 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 and Smith, 1990; Coles, 2001; Beirlant et al., 2004), see 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 heatwave or the United Kingdom flooding in Autumn2000. 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 American Meteorological Society in its bulletin 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 heatwaves in the Tasman Sea off of Australia in 2017 and 2018 and the summer 2017 heatwave 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 datasets 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 rather physically based 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

$$\bar{Y}(D, R) = \sum_{t \in D} \sum_{s \in R} Y(t, s), \quad (4)$$

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 3-month event (June-August) over a region covering Europe and the Mediterranean sea. This choice has an impact of the causal analysis, see Figure 2 and Figure 3 in Cattiaux and Ribes (2018), who proposed to scan a large spatiotemporal domain with respect to an objective criterion. They argued 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. As quoted in Cattiaux and Ribes (2018) *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, they 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 over 1950-2015. For a given D and R , they expressed Equation (4) as

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

where $h(i)$ 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

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

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

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

where the threshold $u(D, R)$ corresponds to the observed temperatures of the 2003 summer for the spatiotemporal region (D, R) . Cattiaux and 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, temperatures follow 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 fact that one wishes to evidence a causal link, was proposed by Hannart and Naveau (2018). By taking advantage of counterfactual theory, these authors proposed to maximize the probability of necessary and sufficient causation (PNS) defined by Equation (2) with $p_0 = P(M\mathbf{X} > u)$ and $p_1 = P(M\mathbf{Z} > u)$, and

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

where u is a scalar and \mathbf{Z} represents a multivariate vector of size d , concatenating the spatiotemporal data from the factual world, and \mathbf{X} from the counterfactual one. 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 and 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 contains the most information in terms of causality. In the hierarchical Gaussian framework, Hannart and Naveau (2018) were able to integrate climate internal variability, model uncertainty, observational error and sampling within a single model. This approach was applied to data from the HadCRUT4 observational dataset for \mathbf{Y} and runs of the IPSL-CM5A-LR model for factual and counterfactual world, \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 one in ten thousand (1:10 000) instead of one in twenty (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 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 either the distance to a known weather regime that is computed independently of the event itself, or from the distance 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

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

with

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

This approach was applied to a record precipitation event that hit southern United Kingdom in January 2014. This is similar to the "additive" decomposition of Shepherd (2016), who also introduced a non-dynamical 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, one of the main caveats in EEA remains the uncertainty in the counterfactual world in terms of conditioning, how to define C_0 and what is the error associated with wrongly defining C_0 .

4.2 | Record events

In terms of 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

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

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, equal to $1/r$ and does not need to be estimated. Hence, one can compare this probability with

$$p_{1,r} = P(Z_r > \max(X_1, \dots, X_{r-1})), \quad (5)$$

to assess the chance that the factual observation Z_r is a record in the counterfactual world. These authors applied their approach to two datasets of annual maxima of daily temperature maxima: observations recorded in Paris and output from the Meteo-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 won't 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 $Z_r > \max(X_1, \dots, X_{r-1})$ 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 heatwave of 2018.

5.1 | Non-parametric approaches

To our knowledge, there have been very few studies in EEA concerning the convergence of any type of non-parametric estimators of p_0 and p_1 , especially with respect to the different uncertainties listed in Section 3.1. A popular current approach is to model exceedance numbers as a simple binomial count, with a success corresponding 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 (see climateprediction.net and Otto et al., 2018) produces thousands or more samples 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 FAR⁵

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

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

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

This provides an asymptotic confidence interval at level α , $\widehat{\text{FAR}}_n(u) \pm z_{1-\alpha/2} \frac{\hat{\sigma}(u)}{\sqrt{n}}$, 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 non-parametric estimators like $\widehat{\text{FAR}}_n(u)$ can be seen by considering the standard deviation $\sigma(u)$. Whenever $\lim n p_0(u_n) = 0$ 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 be always 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 analysis 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 (3) instead of (1). To get confidence intervals for (1) from estimators of (3), one need to model the error in estimating $P(C_0)$ and $P(C_1)$, and to 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 BCa bootstrap) to build confidence intervals for the relative ratio p_1/p_0 . When using non-parametric binomial counting, they recommended, based on numerical simulations, either the Koopman or Wang-Shan methods. They also concluded 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*.

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 and Huser (2015) gave an exhaustive review of the main modeling and estimation strategies to describe

⁵The same reasoning could be made for $\text{RR}(u)$.

univariate and multivariate extremes. In the univariate case (e.g., Coles, 2001; Embrechts et al., 1997), 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

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

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) appear to be slightly heavy-tailed. The GP tail has the important property of being invariant to thresholding. If X follows a 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). As the GP is the only continuous distribution with this feature, this explains why exceedances above a high threshold u are likely to follow a GPD (see, Beirlant et al., 2004; de Haan and Ferreira, 2006, for theoretical aspects). For similar reasons, the only distribution stable for rescaled maxima is the so-called generalized extreme value distribution (GEV)

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

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

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

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 and Smith, 1990), methods-of-moments (e.g., Hosking and Wallis, 1987; Diebolt et al., 2008) or Bayesian inference (e.g., Coles and Tawn, 1996) can be implemented. As noticed by Davison and Huser (2015), *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 and Gelfand, 2009; Shaby and Reich, 2012; Reich et al., 2013).*

EVT techniques that could help to model the different source 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 to 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) 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 non-parametric estimator of $FAR(u)$ described by (6) clearly indicates that a non-parametric 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 as an extreme event by the EVT community. The latter focuses on very high quantiles, 0.99, 0.999 or even 0.9999, e.g., 10000 year return levels for nuclear safety design. In such context, it is impossible to use a non-parametric approach to estimate very high quantiles and one needs to apply the EVT basic principle, i.e., $P(X > u) = P(X > u|X > v) \times P(X > v)$ for any large v and any $u > v$ can be approximated by

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

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

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

Here, a non-parametric estimator, $\sum_{i=1}^n \mathbb{I}(X_i > v)$, has been used to estimate $P(X > v)$ 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 comparable to v , then a 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 risk ratios in function of the convergence rate towards 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 times 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., climateprediction.net (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 last two years, some authors have tried to explore a different path based on non-stationary statistical

methods. In spirit, the main idea is replace large ensemble conditional simulations by transient datasets. For example, the existing CMIP-style 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 non-stationary statistics to transient simulations, typically covering the 1850-2100 period. 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 and 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 that 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 probabilities, see Equation (2). One step in this direction is Hannart and 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 was assumed to be normally distributed, but may be challenged when the variable of interest departs from normality.

5.4 | Multivariate EVT

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 either in space or time. In addition, it will underestimate systemic risks and/or miss compound events. More precisely, let $\mathbf{Y} = (Y_1, \dots, Y_d)^T$ denote a multivariate random vector with cumulative distribution function F and marginal cumulative distribution functions F_1, \dots, 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,

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

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 and Ferreira, 2006). To do so, let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{id})^T$, $i \in \{1, \dots, n\}$, be n independent and identically distributed (iid) copies of \mathbf{Y} . Let $\mathbf{M}_n := (M_{n1}, \dots, M_{nd})$ with $M_{nj} := \max(Y_{1j}, \dots, Y_{nj})$ for $j = 1, \dots, d$. The cdf F is said to be in the max-domain of attraction of an extreme-value distribution G if there exist sequences of normalizing constants $\mathbf{a}_n = (a_{n1}, \dots, a_{nd})^T > 0$

and $\mathbf{b}_n = (b_{n1}, \dots, b_{nd})^T$ such that

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

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

$$\max\left\{\frac{\mathbf{Y} - \mathbf{b}_n}{\mathbf{a}_n}, \mathbf{l}\right\} \mid \mathbf{Y} \not\leq \mathbf{b}_n \xrightarrow{d} \mathbf{V}, \quad n \rightarrow \infty, \quad (11)$$

where $\mathbf{V} = (V_1, \dots, V_d)^T$ is said to follow a multivariate generalized Pareto distribution (Rootzén and Tajvidi, 2006; Rootzén et al., 2018b). For $j \in \{1, \dots, d\}$, 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 and Tajvidi (2006), Falk and Guillou (2008), Ferreira and de Haan (2014), and Rootzén et al. (2018b,a), while statistical modelling is quite recent (Huser et al., 2016; Kirilouk 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

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

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 (9) can be expressed as

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

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 (2) change in a multivariate context, one possibility is to project the multivariate MGPD vector into a one-dimensional object. If \mathbf{V} follows a multivariate GPD with marginal parameters σ and $\xi = (\xi_1, \dots, \xi_d)^T$, 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 (Kirilouk and Naveau, 2019).

In terms of statistical inference and modeling, the recent work of de Fondeville and his colleagues (see, e.g., de Fondeville and Davison, 2019, 2018; Engelke et al., 2019) 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 and Tawn (1996, 1997) measured the speed of the decay toward independence at

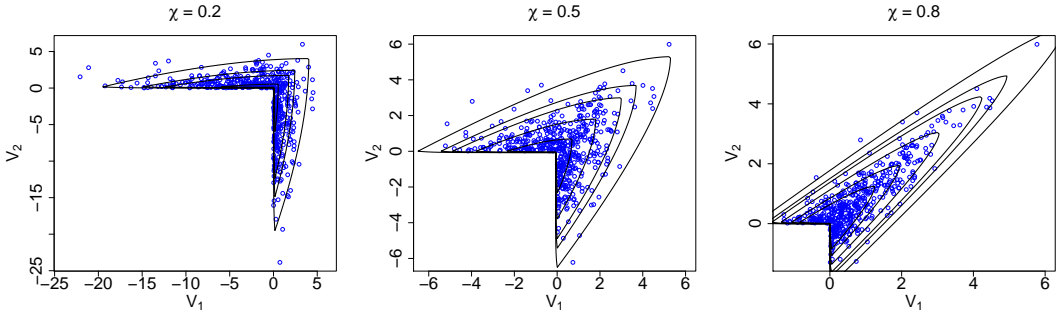


FIGURE 3 Scatterplots and density contours for the bivariate Gaussian MGPD model (12) with $\xi = (0, 0)$, $\sigma = (1, 1)$, $\beta = (0, 0)$. From left to right, the tail dependence coefficient (13) equals to $\chi = 0.2$, $\chi = 0.5$ and $\chi = 0.8$.

high levels via the coefficient of tail dependence, $\eta \in (0, 1]$ under the model

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

where $L(y)$ is slowly varying, i.e., $L(vy)/L(y) \rightarrow 1$ for any $v > 0$ and y large. If $\eta = 1$, then the variables are asymptotically dependent and asymptotically independent otherwise. As quoted in Davison and Huser (2015) *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 datasets. 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 dataset 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 datasets, 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 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. For example, Bayesian model averaging for multivariate extremes (see, e.g., Sabourin et al., 2013) could be a first step. Extending some ideas of Gaussian hierarchical modeling of Katzfuss et al. (2017) and Hannart and 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 non-stationary 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 sensitive in a non-stationary 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 than one component is large (see, e.g., Tawn et al., 2018; Huser and 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 heatwaves 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 open.

ACKNOWLEDGMENTS

We would like to 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.

REFERENCES

- Angéilil, O., Stone, D. and Wehner, M. (2017) An independent assessment of anthropogenic attribution statements for recent extreme temperature and rainfall events. *Journal of Climate*, **30**, 5–16.
- Bader, B., Yan, J. and Zhang, X. (2018) Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *The Annals of Applied Statistics*, **12**, 310–329.
- Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004) *Statistics of Extremes: Theory and Applications*. Wiley : Hoboken, ISBN:978-0-471-97647-9, 522 pages.
- Cattiaux, J. and Ribes, A. (2018) Defining single extreme weather events in a climate perspective. *Bulletin of the American Meteorological Society*, **99**, 1557–1568.
- Chen, Y., Moufouma-Okia, W., Masson-Delmotte, V., Zhai, P. and 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. *Annual Review of Environment and Resources*, **43**, 35–59. URL: <https://doi.org/10.1146/annurev-environ-102017-030052>.
- Coles, S. G. (2001) *An Introduction to Statistical Modeling of Extreme Values*. New York: Springer.
- Coles, S. G. and Tawn, J. A. (1996) A bayesian analysis of extreme rainfall data. *Applied Statistics*, **45**, 463–478.
- Cooley, D., Hunter, B. D. and Smith, R. L. (2019) *Handbook of Environmental and Ecological Statistics*, chap. Univariate and Multivariate Extremes for the Environmental Sciences. Chapman & Hall/CRC Handbooks of Modern Statistical Methods.
- Cooley, D., Nychka, D. and Naveau, P. (2007) Bayesian spatial modeling of extreme precipitation return levels. *Journal of The American Statistical Association*, **102**, 824–840.
- Davison, A. C. and Huser, R. G. (2015) Statistics of extremes. *Annual Review of Statistics and Its Application*, **2**, 203–235. URL: <https://doi.org/10.1146/annurev-statistics-010814-020133>.
- Davison, A. C. and Smith, R. L. (1990) Models for exceedances over high thresholds (with comments). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **52**, 393–442.
- Diebolt, J., Guillou, A., Naveau, P. and Ribereau, P. (2008) Improving probability-weighted moment methods for the generalized extreme value distribution. *REVSTAT - Statistical Journal*, **6**, 33–50.
- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*, vol. 33 of *Applications of Mathematics*. Springer-Verlag, Berlin.
- Engelke, S., de Fondeville, R. and Oesting, M. (2019) Extremal behaviour of aggregated data with an application to downscaling. *Biometrika*, **106**, 127–144.
- Falk, M. and Guillou, A. (2008) Peaks-over-threshold stability of multivariate generalized Pareto distributions. *Journal of Multivariate Analysis*, **99**, 715–734.
- Ferreira, A. F. and de Haan, L. (2014) The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, **20**, 1717–1737.

- Field, C., Barros, V., Stocker, T., Qin, D., Dokken, D. and al. (2012) *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate*. Cambridge, UK: Cambridge Univ. Press.
- de Fondeville, R. and Davison, A. C. (2018) High-dimensional peaks-over-threshold inference. *Biometrika*, **105**, 575–592.
- (2019) Functional peaks-over-threshold analysis and generalized r-Pareto processes. *Submitted*.
- Gilleland, E., Katz, R. W. and Naveau, P. (2017) Quantifying the risk of extreme events under climate change. *CHANCE*, **30**, 30–36.
- de Haan, L. and Ferreira, A. F. (2006) *Extreme Value Theory: an Introduction*. Springer-Verlag New York.
- de Haan, L., Tank, A. and Neves, C. (2015) On tail trend detection: modeling relative risk. *Extremes*, **18**, 141–178.
- Hammerling, D., Katzfuss, M. and Smith, R. L. (2019) *Handbook of Environmental and Ecological Statistics*, chap. Climate Change Detection and Attribution. Chapman & Hall/CRC Handbooks of Modern Statistical Methods.
- Hannart, A. and Naveau, P. (2018) Probabilities of causation of climate changes. *Journal of Climate*, **31**, 5507–5524.
- Hannart, A., Pearl, J., Otto, F. E. L., Naveau, P. and Ghil, M. (2016) Counterfactual causality theory for the attribution of weather and climate-related events. *Bulletin of the American Meteorological Society*, **97**, 99–110.
- Hannart, A., Ribes, A. and Naveau, P. (2014) Optimal fingerprinting under multiple sources of uncertainty. *Geophysical Research Letters*, **41**, 1261–1268.
- Hosking, J. R. M. and Wallis, J. R. (1987) Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, **29**, 339–49.
- Huser, R. G., Davison, A. C. and Genton, M. G. (2016) Likelihood estimators for multivariate extremes. *Extremes*, **19**, 79–103.
- Huser, R. G. and Wadsworth, J. L. (2019) Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, **114**, 434–444.
- Katz, R. W., Parlange, M. B. and Naveau, P. (2002) Statistics of extremes in hydrology. *Advances in Water Resources*, **25**, 1287–1304.
- Katzfuss, M., Hammerling, D. and Smith, R. L. (2017) A Bayesian hierarchical model for climate change detection and attribution. *Geophysical Research Letters*, **44**, 5720–5728.
- Kew, S. F., Philip, S. Y., Jan van Oldenborgh, G., van der Schrier, G., Otto, F. E. L. and Vautard, R. (2019) The exceptional summer heat wave in Southern Europe 2017. *Bulletin of the American Meteorological Society*, **100**, S49–S53. URL: <https://doi.org/10.1175/BAMS-D-18-0109.1>.
- Kharin, V. V. and Zwiers, F. W. (2000) Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. *Journal of Climate*, **13**, 3760–3788.
- (2005) Estimating extremes in transient climate change simulations. *Journal of Climate*, **18**, 1156–1173.
- Kharin, V. V., Zwiers, F. W., Zhang, X. and Hegerl, G. C. (2007) Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *Journal of Climate*, **20**, 1419–1444. URL: <https://doi.org/10.1175/JCLI4066.1>.
- Kim, Y. H., Min, S. K., Zhang, X., Zwiers, F., Alexander, L. V., Donat, M. G. and Tung, Y.-S. (2016) Attribution of extreme temperature changes during 1951–2010. *Climate Dynamics*, **46**, 1769–1782.
- King, A. D. (2017) Attributing changing rates of temperature record breaking to anthropogenic influences. *Earth's Future*, **5**, 1156–1168.

- Kirchmeier-Young, M. C., Gillett, N. P., Zwiers, F. W., Cannon, A. and Anslow, F. (2017) Attribution of extreme events in arctic sea ice extent. *Journal of Climate*, **30**, 553–571.
- Kirchmeier-Young, M. C., Zwiers, F. W. and 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. and Naveau, P. (2019) Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory. *Submitted*, <https://arxiv.org/abs/1908.03107>.
- Kiriliouk, A., Rootzén, H., Segers, J. and Wadsworth, J. L. (2019) Peaks over thresholds modelling with multivariate generalized Pareto distributions. *Technometrics*, **61**, 123–135.
- Ledford, A. and Tawn, J. A. (1996) Statistics for near independence in multivariate extreme values. *Biometrika*, **83**, 169–187.
- (1997) Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **59**, 475–499.
- Luu, L. N., Vautard, R., Yiou, P., van Oldenborgh, G. J. and Lenderink, G. (2018) Attribution of extreme rainfall events in the south of France using Euro-CORDEX simulations. *Geophysical Research Letters*, **45**, 6242–6250.
- National Academies of Sciences, Engineering and Medicine (2016) *Attribution of Extreme Weather Events in the Context of Climate Change*. Washington, DC: The National Academies Press. URL: <https://www.nap.edu/catalog/21852/attribution-of-extreme-weather-events-in-the-context-of-climate-change>.
- Naveau, P., Huser, R. G., Ribereau, P. and Hannart, A. (2016) Modelling jointly low, moderate and heavy rainfall intensities without a threshold selection. *Water Resources Research*, **52**, 2753–2769.
- Naveau, P., Ribes, A., Zwiers, F. W., Hannart, A., Tuel, A. and Yiou, P. (2018) Revising return periods for record events in a climate event attribution context. *Journal of Climate*, **31**, 3411–3422.
- van Oldenborgh, G. J., van der Wiel, K., Sebastian, A., Singh, R., Arrighi, J., Otto, F., Haustein, K., Li, S., Vecchi, G. and Cullen, H. (2017) Attribution of extreme rainfall from Hurricane Harvey, August 2017. *Environmental Research Letters*, **12**, 124009. URL: <https://doi.org/10.1088/1748-9326/12/2/Faa9ef2>.
- Otto, F., Philip, S., Kew, S., Li, S., King, A. and Cullen, H. (2018) Attributing high-impact extreme events across timescales—a case study of four different types of events. *Climatic Change*, **149**, 399–412.
- Paciorek, C. J., Stone, D. A. and Wehner, M. F. (2018) Quantifying statistical uncertainty in the attribution of human influence on severe weather. *Weather and Climate Extremes*, **20**, 69–80. URL: <http://www.sciencedirect.com/science/article/pii/S2212094717300841>.
- Pearl, J. (2009) *Causality: Models, Reasoning, and Inference*. Cambridge University Press, second edn.
- Ragone, F., Wouters, J. and Bouchet, F. (2018) Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences*, **115**, 24–29. URL: <https://www.pnas.org/content/115/1/24>.
- Reich, B. J., Shaby, B. A. and Cooley, D. (2013) A hierarchical model for serially dependent extremes: a study of heat waves in the western US. *Journal of Agricultural, Biological and Environmental Statistics*, **19**, 119–135.
- Resnick, S. I. (1987) *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag New York Berlin Heidelberg.
- Ribes, A., Thao, S. and Cattiaux, J. (2019) Describing the relationship between a weather event and climate change: a new statistical approach. *Submitted*.
- Ribes, A., Zwiers, F. W., Azais, J. M. and Naveau, P. (2016) A new statistical approach to climate change detection and attribution. *Climate Dynamics*, 29–56.

- Risser, M. D., Paciorek, C. J. and Stone, D. A. (2019) Spatially dependent multiple testing under model misspecification, with application to detection of anthropogenic influence on extreme climate events. *Journal of the American Statistical Association*, **114**, 61–78. URL: <https://doi.org/10.1080/01621459.2018.1451335>.
- Rootzén, H. and Katz, R. W. (2013) Design life level: Quantifying risk in a changing climate. *Water Resources Research*, **49**, 5964–5972.
- Rootzén, H., Segers, J. and Wadsworth, J. L. (2018a) Multivariate generalized Pareto distributions: parametrizations, representations and properties. *Journal of Multivariate Analysis*, **165**, 117–131.
- (2018b) Multivariate peaks over thresholds models. *Extremes*, **21**, 115–145.
- Rootzén, H. and Tajvidi, N. (2006) Multivariate generalized Pareto distributions. *Bernoulli*, **12**, 917–930.
- Sabourin, A., Naveau, P. and Fougères, A. L. (2013) Bayesian model averaging for multivariate extremes. *Extremes*, **16**, 325–350.
- Sang, H. and Gelfand, A. (2009) Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics*, **16**, 407–426.
- Serinaldi, F. (2016) Can we tell more than we can know? the limits of bivariate drought analyses in the United States. *Stochastic Environmental Research and Risk Assessment*, **30**, 1691–1704.
- Shaby, B. A. and Reich, B. J. (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, T. G. (2016) A common framework for approaches to extreme event attribution. *Current Climate Change Reports*, **2**, 28–38.
- Shooter, R., Ross, E., Tawn, J. A. and Jonathan, P. (2019) On spatial conditional extremes for ocean storm severity. *Environmetrics*, 1–18.
- Stocker, T., Qin, D., Plattner, G., Tignor, M., Allen, S. and 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, P. A., Christidis, N., Otto, F. E. L., Sun, Y., Vanderlinden, J. P., van Oldenborgh, G., Vautard, R., von Storch, H., Walton, P., Yiou, P. and Zwiers, F. W. (2016) Attribution of extreme weather and climate-related events. *WIREs Climate Change*, **7**, 23–41.
- Stott, P. A., Stone, D. A. and Allen, M. R. (2004) Human contribution to the European heatwave of 2003. *Nature*, **432**, 610–613.
- Tajvidi, N. (1996) *Characterisation and Some Statistical Aspects of Univariate and Multivariate Generalized Pareto Distributions*. Ph.D. thesis, Department of Mathematics, Chalmers, Göteborg.
- Tawn, J. A., Shooter, R., Towe, R. and Lamb, R. (2018) Modelling spatial extreme events with environmental applications. *Spatial statistics*, **28**, 39–58.
- Tencaliec, P., Favre, A., Naveau, P., Prieur, C. and Nicolet, G. (2019) Flexible semiparametric generalized Pareto modeling of the entire range of rainfall amount. *Environmetrics*, <https://doi.org/10.1002/env.2582>, 1–20.
- Trenberth, K. E., Fasullo, J. T. and Shepherd, T. G. (2015) Attribution of climate extreme events. *Nature Climate Change*, **5**, 725–730.
- Vautard, R., van Oldenborgh, G. J., Otto, F. E. L., Yiou, P., de Vries, H., van Meijgaard, E., Stepek, A., Soubeyroux, J. M., Philip, S., Kew, S. F., Costella, C., Singh, R. and Tebaldi, C. (2019) Human influence on european winter wind storms such as those of january 2018. *Earth System Dynamics*, **10**, 271–286.

- van der Wiel, K., Kapnick, S. B., van Oldenborgh, G. J., Whan, K., Philip, S., Vecchi, G. A., Singh, R. K., Arrighi, J. and Cullen, H. (2017) Rapid attribution of the August 2016 flood-inducing extreme precipitation in south Louisiana to climate change. *Hydrology and Earth System Sciences*, **21**, 897–921. URL: <https://www.hydro1-earth-syst-sci.net/21/897/2017/>.
- Yiou, P., Cattiaux, J., Faranda, D., Kadygrov, N., Jézéquel, A., Naveau, P., Ribes, A., Robin, Y., Thao, S., van Oldenborgh, G. J. and Vrac, M. (2019) Analyses of the northern european summer heatwave of 2018. *Bulletin of the American Meteorological Society*, in review.
- Yiou, P., Jezequel, A., Naveau, P., Otto, F. E. L., Vautard, R. and Vrac, M. (2017) A statistical framework for conditional extreme event attribution. *Advances in Statistical Climatology, Meteorology and Oceanography*, **3**, 17–31.
- Zscheischler, J., Westra, S., van den Hurk, B. J. J. M., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T. and Zhang, X. (2018) Future climate risk from compound events. *Nature Climate Change*, **8**, 469–477. URL: <https://doi.org/10.1038/s41558-018-0156-3>.