Describing the Relationship between a Weather Event and Climate Change: A New Statistical Approach

AURÉLIEN RIBES

Centre National de Recherches Météorologiques, Université de Toulouse, Météo France, CNRS, Toulouse, France

SOULIVANH THAO

Laboratoire des Sciences du Climat et de l'Environnement, UMR8212 CEA-CNRS-UVSQ, IPSL and U Paris Saclay, Gif-sur-Yvette, France

JULIEN CATTIAUX

Centre National de Recherches Météorologiques, Université de Toulouse, Météo France, CNRS, Toulouse, France

(Manuscript received 20 March 2019, in final form 20 January 2020)

ABSTRACT

Describing the relationship between a weather event and climate change—a science usually termed *event attribution*—involves quantifying the extent to which human influence has affected the frequency or the strength of an observed event. In this study we show how event attribution can be implemented through the application of nonstationary statistics to transient simulations, typically covering the 1850–2100 period. The use of existing CMIP-style simulations has many advantages, including their availability for a large range of coupled models and the fact that they are not conditional to a given oceanic state. We develop a technique for providing a multimodel synthesis, consistent with the uncertainty analysis of long-term changes. Last, we describe how model estimates can be combined with historical observations to provide a single diagnosis accounting for both sources of information. The potential of this new method is illustrated using the 2003 European heat wave and under a Gaussian assumption. Results suggest that (i) it is feasible to perform event attribution using transient simulations and nonstationary statistics, even for a single model; (ii) the use of multimodel synthesis in event attribution is highly desirable given the spread in single-model estimates; and (iii) merging models and observations substantially reduces uncertainties in human-induced changes. Investigating transient simulations also enables us to derive insightful diagnostics of how the targeted event will be affected by climate change in the future.

1. Introduction

Describing the relationship between a given weather or climate event and anthropogenic climate change is a growing area of activity in the field of climate science (National Academies of Sciences, Engineering, and Medicine 2016). Since the pioneering studies of Allen (2003) and Stott et al. (2004), the concept of event attribution has been applied to a wide variety of events, as synthesized in the annual special issues of the *Bulletin of the American Meteorological Society*

(BAMS) “Explaining Extreme Events from a Climate Perspective” (Peterson et al. 2012, and subsequent issues).¹

Multiple approaches have been introduced to address this question. Beyond issues related to the definition of the event of interest, the most commonly used approach is probabilistic and involves a comparison of the distributions of extreme events in the factual versus

Denotes content that is immediately available upon publication as open access.

¹ See <https://www.ametsoc.org/index.cfm/ams/publications/bulletin-of-the-american-meteorological-society-bams/explaining-extreme-events-from-a-climate-perspective/>.

This article is licensed under a Creative Commons Attribution 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

Corresponding author: Aurélien Ribes, aurelien.ribes@meteo.fr

counterfactual worlds (e.g., Stott et al. 2004; Pall et al. 2011). Particular attention is paid to changes in probability of the event associated with human influence. Various alternatives have been proposed in the literature; one of these involves focusing on the thermodynamic component of human influence (e.g., Trenberth et al. 2015; Cattiaux et al. 2010). However, this study will focus on the probabilistic approach and its statistical implementation, that is, how estimating changes in occurrence frequency and the corresponding uncertainty.

At least two methods are commonly used to derive such probabilities.

First, large ensembles of simulations are used to sample the factual and counterfactual statistical distributions (Pall et al. 2011; Massey et al. 2015; Christidis et al. 2013; Ciavarella et al. 2018; Wehner et al. 2018). Such ensembles are typically produced with atmospheric-only models forced by prescribed sea surface temperatures (SSTs); factual SSTs are usually taken from observations, while counterfactual SSTs are derived by subtracting an estimate of the anthropogenic influence. Such ensembles can be very large, typically from a few hundred to more than 10 000 simulations of one year or one season. One important advantage of using such large ensembles is that the signal-to-noise ratio is increased (sampling noise is reduced), and probabilities can be estimated very straightforwardly by counting exceedences, that is, using a minimal statistical inference—although more complex treatments have also been proposed (Paciorek et al. 2018). Several disadvantages should also be mentioned: the computational cost is relatively high (large number of simulations, which have to be redone on an annual basis at least), processes involving ocean–atmosphere coupling are missing (Dong et al. 2017), results are conditional on the particular pattern of SSTs considered (Risser et al. 2017), model bias or reliability issues can affect results (Bellprat and Doblas-Reyes 2016), and last, modeling uncertainty is usually not assessed comprehensively due to the lack of coordinated exercise [Wehner et al. (2018) is a noticeable exception].

Second, occurrence probabilities can be inferred from observations and observed trends, assuming that the trends are entirely related to human influence on climate (e.g., van Oldenborgh et al. 2015; Vautard et al. 2015). This approach eliminates concerns related to climate model biases and/or error in representing climate change. However, inference becomes dependent on the choice of a statistical model, and the signal-to-noise ratio is usually limited in observations (e.g., Li et al. 2019). Uncertainty in the observed trend estimate might lead to very wide uncertainty in the probability ratio or other diagnostics of the human influence. Further, this technique is highly dependent on the availability of a long, homogeneous historical record.

A few attempts have been made to consider these two approaches simultaneously (Uhe et al. 2016; Eden et al. 2016; Hauser et al. 2017), providing very helpful comparisons of methods for selected case studies. Several attribution studies from the World Weather Attribution² (WWA) also made some synthesis of both sources of information. However, to the best of our knowledge, no statistical method has been proposed to formally combine models and observations results in order to provide one single estimate of human influence.

In this paper, we tackle several of these issues. First, we propose to base event attribution on transient the Coupled Model Intercomparison Project (CMIP)-style simulations—typically a combination of historical and RCP scenarios. This is done through the use of non-stationary statistics (section 3). Second, we propose a statistical procedure to create a rigorous multimodel synthesis. This question has not been fully addressed in previous event attribution literature. We show that, if a large multimodel ensemble is available, the assumptions and techniques used to construct multimodel syntheses for large-scale mean variables can be extended to event attribution (section 4). Third, we introduce a statistical framework for merging information from models and observations. The proposed method is essentially Bayesian, in the sense that available observations are used to constrain the model range further (section 5).

In addition to the statistical inference itself, we promote the use of two additional diagnostics in describing the relationship between a particular event and climate change. First, the human influence is quantified both in terms of probability and intensity of the event. Although highlighting this duality is not new, using one point of view or the other may have contributed to past controversies (Otto et al. 2012), and both quantities can be derived from the same statistical analysis. Second, we describe how the characteristics of the event (frequency, intensity) evolve with time. This allows us to describe not only the human influence to date—the main diagnosis of event attribution studies—but also how a similar event will be affected by climate change in the future. This type of outcome is another benefit of using transient simulations and might be very helpful for communicating the relationship between an event and climate change in a comprehensive way.

Several facets of the proposed statistical method have been already introduced in the literature, for example, the investigation of future event probability (Sun et al. 2014; Christidis et al. 2015a; Kirchmeier-Young et al. 2017) or the use of some observational

² See <https://www.worldweatherattribution.org>.

constraints (e.g., Christidis et al. 2015b). The general structure of our method also shares several similarities with recent WWA studies [e.g., the use of a covariate and nonstationary statistics; see Van Oldenborgh et al. (2017), among other examples]. Using transient CMIP-style simulations for event attribution is not a new idea either (Lewis and Karoly 2013; Sun et al. 2014; Christidis et al. 2015b; King et al. 2015; Kirchmeier-Young et al. 2017, 2019). The main issue with such simulations is that the sample size is limited—usually to no more than 10 members. This is at least partly compensated by the fact that these simulations include a period of time (near the end of the twenty-first century) in which the human influence is much more pronounced than in the current climate—increasing the signal-to-noise ratio. Regardless, there are tremendous advantages in using such simulations: they are already available (reducing the additional computational cost for an event attribution analysis to almost zero), performed with fully coupled models (i.e., accounting for coupled processes, and also not conditional on a specific oceanic state), and available for many models (making it possible to account for a wider range of model uncertainties).

The main goal of this paper is to describe the proposed statistical method, and to provide a first illustration of its potential. The proposed algorithm is flexible and could be improved in several ways, without significantly affecting its general structure. Strengths and weaknesses are also discussed in the last section.

2. Framing and data

a. Event definition and indicators of the human influence

A relatively wide variety of events or classes of events has been considered in event attribution. In this study we focus on simple classes of events such as

$$E = \{Y > s\}, \quad (1)$$

where Y is a univariate random climate variable—typically temperature, rainfall, or wind speed, averaged over a given time window and spatial domain—and s is a pre-determined threshold. This class of events consists of all outcomes that are greater than or equal to the threshold s . The choice of s is, in some way, informed by the weather event that motivates the study. We assume that this event has happened at a time t_e in the factual (F) world.³

³ The factual world, or *world as it is*, is the world where all external forcings, including the anthropogenic forcings, have influenced climate.

The attribution analysis involves describing the characteristics of a *similar* event happening in the counterfactual (C) world.⁴ The meaning of *similar* depends on the point of view (frequency or intensity) and is discussed below. As we consider transient simulations where climate changes with time, describing how the characteristics of the event vary with time, for example, in the factual world, is also of interest.

Changes in occurrence frequency/probability can be assessed by comparing the probability of the event E happening in (F) versus (C), considering the same threshold s . Denoting $F_{F,t}$ and $F_{C,t}$ the cumulative distribution functions of Y at time t in the factual and counterfactual worlds, respectively, we define

$$\begin{aligned} p_F(t) &= \mathbb{P}_{F,t}(E) = 1 - F_{F,t}(s), \\ p_C(t) &= \mathbb{P}_{C,t}(E) = 1 - F_{C,t}(s). \end{aligned} \quad (2)$$

Human influence is then typically characterized through the probability ratio (PR) or the fraction of attributable probability⁵ (FAP; Stott et al. 2004):

$$\text{PR}(t) = \frac{p_F(t)}{p_C(t)}, \quad \text{FAP}(t) = \frac{p_F(t) - p_C(t)}{p_F(t)} = 1 - \frac{1}{\text{PR}(t)}. \quad (3)$$

As they are of critical importance, we will denote $p_F = p_F(t_e)$ and $p_C = p_C(t_e)$ the probabilities at time t_e .

Changes in intensity are assessed by comparing the magnitude of events with the same occurrence probability; this value is set to p_F , consistent with the observed event:

$$\begin{aligned} I_F(t) &= F_{F,t}^{-1}(1 - p_F), \quad I_C(t) = F_{C,t}^{-1}(1 - p_F), \quad \text{then} \\ \Delta I(t) &= I_F(t) - I_C(t). \end{aligned} \quad (4)$$

In other words, I_F and I_C are the quantiles of order p_F of $F_{F,t}$ and $F_{C,t}$, respectively. By definition, $I_F(t_e) = s$. ΔI tells how much more/less intense the event with exactly the same frequency would have been in the counterfactual world. Note that, for some climate variable such as precipitation, a relative difference (i.e., $\Delta I = I_F/I_C$) might be better appropriate than absolute difference.

Two important remarks can be added. First, conventional attribution studies only calculate PR, FAP, or ΔI at time t_e , that is, the exact date at which the event was

⁴ The counterfactual world, or *world that might have been*, is the world where anthropogenic influence is removed, while natural forcings still vary through time.

⁵ These indices were often referred to as risk ratio (RR) and fraction of attributable risk (FAR) in previous literature.

observed. Calculation of PR, FAP, or ΔI at a different date allows us to quantify the human influence, had the event occurred at that date. Second, describing how the characteristics of the event are changing through time, for example, in the factual world, is also helpful (see, e.g., Christidis et al. 2015a). This can be done using relative indices, for example,

$$\text{PR}_{\text{rel}}(t) = \frac{p_F(t)}{p_F(t_e)}, \quad \text{or} \quad \Delta I_{\text{rel}}(t) = I_F(t) - I_F(t_e). \quad (5)$$

All these diagnostics are calculated and illustrated subsequently.

b. Case study: 2003 European heat wave

To illustrate the method presented in this paper, we focus on the 2003 European heat wave (EHW03), an event that has long been scrutinized in event attribution studies (Stott et al. 2004; Schär et al. 2004; Christidis et al. 2015a). We define EHW03 (variable Y) as a 1-month event occurring in August 2003 near Paris, France. The spatial domain considered is a $5^\circ \times 5^\circ$ square surrounding Paris, that is, $[45^\circ\text{--}50^\circ\text{N}]$ and $[0^\circ\text{--}5^\circ\text{E}]$. The choice of this space–time window is debatable (see, e.g., Cattiaux and Ribes 2018). A monthly value was considered as a convenient choice in order to involve as many phase 5 of CMIP (CMIP5) models as possible and investigate their level of agreement. The threshold used is $s = +5.38^\circ\text{C}$ anomaly with respect to the 1961–90 mean, corresponding to actual value observed in 2003.

As further discussed below, our technique also requires the use of a covariate X , which is assumed to be representative of climate change magnitude over time. We consider the summer mean temperature over western Europe ($[35^\circ\text{--}70^\circ\text{N}]$ and $[10^\circ\text{W}\text{--}30^\circ\text{E}]$) in this respect.

c. Data

We use data from a collection of climate models from CMIP5 (all 24 models considered are listed in Fig. 4). For each model, we combine historical simulations (1850–2005) and RCP8.5 simulations (2006–2100). We use all available runs in cases where ensembles have been performed—using a different number of historical and RCP8.5 simulations is not problematic. Preindustrial control simulations are also used to quantify internal variability and derive confidence intervals.

Our method also requires using observed data. We use HadCRUT4 (Morice et al. 2012, <https://crudata.uea.ac.uk/cru/data/temperature/>) to provide historical summer mean temperatures over western Europe (1850–2016, $[10^\circ\text{W}, 30^\circ\text{E}] \times [35^\circ, 70^\circ\text{N}]$), and August mean

temperatures in the vicinity of Paris, France (1850–2016, $[0^\circ, 5^\circ\text{E}] \times [45^\circ, 50^\circ\text{N}]$).

3. Statistical analysis of transient simulations

In this section, we consider data from one single climate model and describe how changes in occurrence probability can be calculated from such data. By construction, transient simulations exhibit a nonstationary climate, so using nonstationary statistics is a key component of our approach. Therefore, we consider a covariate X that is assumed to be representative of climate change magnitude over time. The covariate will typically be a temperature, averaged either globally or over a large region, on a seasonal or annual basis. Several studies have already used the global mean temperature as such a covariate (van Oldenborgh et al. 2015; van der Wiel et al. 2017). Here we use summer mean European average temperature and refer to section 6 for further discussion on this choice. Once this covariate X has been selected, our procedure is as follows.

a. Anthropogenic and natural contributions to changes in X

As a first step, we need to estimate the forced responses in the covariate X and, in particular, the contributions of natural versus anthropogenic forcings to changes in X . This is typically the purpose of detection and attribution techniques. However, these techniques are not usually designed to provide smooth time series as a result. We therefore propose a hybrid approach using generalized additive models (GAM; Hastie 2017; Wood 2017).

We assume that

$$X_t = \underbrace{\mu_x + \beta e_t}_{\text{natural}} + \underbrace{f(t)}_{\text{anthropogenic}} + \varepsilon_t, \quad \text{and} \quad (6)$$

$$= X_{\text{nat}}(t) + X_{\text{ant}}(t) + \varepsilon_t, \quad (7)$$

where μ_x is a constant, e_t is an energy balance model (EBM; see Held et al. 2010) response to natural forcings only at the global scale, β is an unknown scaling factor, $f(t)$ is a smooth temporal function, and ε_t is a random term describing internal variability.

Within this framework, estimation of the response to natural forcing is quite consistent with usual practice in detection and attribution [D&A; more specifically, the ordinary least squares (OLS) formulation], as it involves the estimation of an unknown scaling factor β . The main innovation is the consideration of the response e_t derived from an EBM (similar to Huber and Knutti 2012), rather than a more complex model. In doing this, we take advantage of the forcing

time series and avoid involving additional noise (i.e., internal variability) from a climate model run. The impact of major volcanic eruptions, for instance, becomes clearly noticeable in that way. In practice, we consider the EBM parameter settings given in [Geoffroy et al. \(2013\)](#), for a subset of CMIP5 models. We calculate the EBM responses e_t using the corresponding settings (one calculation for each CMIP model). The multimodel average of all e_t is taken as a best estimate, while we randomly sample over the finite set of responses to quantify uncertainty. Finally, the rescaling of e_t is important if a regional temperature is considered, as EBMs are only suitable to describe global or hemispheric means.

The response to anthropogenic forcing f is assumed to be smooth over time and is estimated from the data x_t using smoothing splines. The regularity assumption can be regarded as sensible, as greenhouse gas and aerosols, that is, the two dominant drivers, vary quite slowly over time. Anthropogenically induced changes are computed with respect to a reference date t_{ref} , implying that $f(t_{\text{ref}}) = 0$; we consider $t_{\text{ref}} = 1850$, consistent with CMIP5 protocol, but another reference date could be used. Although the human influence is dominant on mean temperature changes over recent decades ([Bindoff et al. 2013](#)), the estimate of f is still influenced by low-frequency internal variability. It will be necessary to account for this component in the uncertainty analysis.

Estimation within model (6) can be made using standard GAM tools. Here we chose to estimate f using smoothing splines with 6 equivalent degrees of freedom—this number was tuned using cross validation.

Quantifying uncertainty in this decomposition is more difficult, since it is important to account for dependencies in ε_t . It is assumed that $\varepsilon_t \sim N(0, \Sigma)$, where Σ is known (derived from preindustrial control simulations, as usual in D&A) but not equal to identity. Uncertainties in X_{nat} and X_{ant} are assessed by using (i) perturbed values of e_t (using EBM parameters fitted to individual CMIP models), (ii) parametric uncertainty on β given $[f(\cdot), e_t]$, and (iii) parametric uncertainty on $f(\cdot)$ given (β, e_t) .

This decomposition procedure is illustrated in [Fig. 1](#) for one particular CMIP5 model (CNRM-CM5). Response to major volcanic eruptions can be easily identified in both the factual world (all forcings combined) and the counterfactual world (natural forcings only). The human influence emerges from noise near 1970 in this model. This is not necessarily contradictory with the fact that human influence is not attributable at that date in the instrumental record—10 model runs are

FIG. 1. ANT and NAT influences in covariate X . Changes in covariate X (raw data are shown as black points) are decomposed into changes related to (top) all forcings combined (red), (middle) anthropogenic forcings only (green), and (bottom) natural forcings only (blue). Shaded area corresponds to 5%–95% confidence regions. All data are from CNRM-CM5, including all runs available from this model (10 historical, 5 RCP8.5).

used, while only one observed realization is available, implying different signal-to-noise ratios.

This decomposition produces two major outputs: the estimated response to natural forcings only X_{nat} , corresponding to the expected value of X in the counterfactual world, and $X_{\text{all}} = X_{\text{ant}} + X_{\text{nat}}$, describing the state of X in the factual world.

b. Fitting a nonstationary distribution to Y

As a second step, a nonstationary distribution is fitted to the variable of interest Y . $X_{\text{all}}(t)$ is used as a covariate in this nonstationary fit. Two types of distributions are considered:

- Gaussian distribution:

$$Y_t \sim N\{\mu_0 + \mu_1 X_{\text{all}}(t), \sigma_0[1 + \sigma_1 X_{\text{all}}(t)]\}. \quad (8)$$

The parameters $\gamma = (\mu_0, \mu_1, \sigma_0, \sigma_1)$ can be estimated via maximum likelihood. However, no closed formula is available in this case, and an optimization algorithm is needed. We used the `nlminb` R routine, chosen from other possible options. Confidence regions on γ can be derived by bootstrapping $[X_{\text{all}}(t), Y_t]$ and simultaneously considering uncertainty on X_{all} , derived from the previous step. The bootstrap used assumes

year-to-year independence and therefore does not account for potential interannual correlation.

- Nonparametric distributions, assuming that the quantile of order α at time t , $q^\alpha(t)$, satisfies

$$q^\alpha(t) = \mu_0^\alpha + \mu_1^\alpha X_{\text{all}}(t). \quad (9)$$

In this case, the parameters $\gamma = (\mu_0^\alpha, \mu_1^\alpha)$ can be estimated, for a collection of α , using quantile regression (Koenker and Bassett 1978; Koenker and Hallock 2001). Given typical sample sizes (a few hundreds to thousands of data) and the computational cost of resampling, a fast algorithm is needed, and we used the Frisch–Newton approach after preprocessing (see Portnoy et al. 1997, implemented in *R* under “pfn”). Uncertainty on $(\mu_0^\alpha, \mu_1^\alpha)$ is assessed through a bootstrap procedure.

This list of distributions is obviously not exhaustive, and other families might be used. For instance, generalized extreme value (GEV) distributions could be of interest when the focus is on annual maxima (Coles et al. 2001). Still, a nonparametric distribution offers the highest flexibility, and might be nicely complemented by adjusting generalized Pareto distribution (GPD) to the tails, in order to improve estimation of rare values.⁶ In the remainder of this paper, we focus on nonstationary Gaussian distribution only for illustrating our method. So far, nonparametric distributions were only used to analyze transient simulations, and provided results consistent with Gaussian distributions.

The fit of a nonstationary Gaussian distribution is illustrated in Fig. 2. This figure suggests that $X_{\text{all}}(t)$ is an appropriate covariate for Y_t , as there is no evidence of a nonlinear relationship in the data. More generally, this type of diagnosis can be used to check if the choice of the covariate is appropriate. The fact that the three regression lines (corresponding to mean and quantiles) are almost parallel indicates that there are almost no changes in variance for this particular model and variable. We also find that, thanks to the temporal smoothing, internal variability does not contribute to the correlation between $X_{\text{all}}(t)$ and Y_t (not shown)—although internal variability does induce a significant correlation between X_t and Y_t in this particular case.

⁶Nonstationary GPD distribution could be used to model threshold exceedences. However, in many practical situations, it might be useful to obtain an estimate of the entire distribution, not only the tail. In particular, in the case of temperature, an event in the tail of the distribution in the counterfactual world can become quite common over the course of the twenty-first century, requiring estimation of the entire distribution in order to derive standard attribution statistics.

FIG. 2. Fit of a nonstationary Gaussian distribution. Illustration of the nonstationary fit for data from the CNRM-CM5 model, assuming Gaussian distribution. Black points: data $[X_{\text{all}}(t), Y_t]$, where $X_{\text{all}}(t)$ has been estimated following section 3a, and Y_t is the raw Y data, that is, August mean surface air temperatures from CNRM-CM5 for the $5^\circ \times 5^\circ$ grid square that includes Paris. Several simulations are considered, leading to several values of Y_t at each time t . Many points lie in the bottom-left corner, corresponding to the quasi-stationary climate of the period before 1980. Red lines: change in the mean of Y_t (solid) or 5%–95% quantiles (dashed lines) as estimated from the nonstationary fit.

In the following, γ will be split into (γ_0, γ_1) , where γ_0 are parameters describing the distribution of Y at a reference time (or in a stationary climate), while γ_1 are parameters describing the sensitivity of Y to changes in X . For instance, in the Gaussian case, $\gamma_0 = (\mu_0, \sigma_0)$ and $\gamma_1 = (\mu_1, \sigma_1)$.

c. Estimating changes in probability/intensity

Once a nonstationary distribution has been fitted on Y , the PDF of Y in the factual world is given by (8). The equivalent PDF in the counterfactual world is obtained by replacing X_{all} by X_{nat} . These PDFs can be translated into the corresponding CDFs $F_{F,t}$ and $F_{C,t}$. Then, the diagnostics proposed in section 2 can be derived easily, leading to PR and ΔI .

Changes in frequency and intensity, as estimated from one particular CMIP5 model, are illustrated in Fig. 3 for the EHW03 event as defined in section 2b. The event frequency moves from about 2.10^{-4} in 1850 to more than $1/2$ in 2100 in the factual world. These numbers are not inconsistent with Stott et al. (2004), as we consider a smaller space–time domain. Over the same period, the

FIG. 3. Changes in frequency and intensity in CNRM-CM5. Changes in the frequency and intensity of the event, as diagnosed from the analysis of transient historical and RCP8.5 scenarios from the CNRM-CM5 model. Changes in frequency are analyzed in terms of (top) occurrence probability in the factual world (p_F), (middle top) occurrence probability in the counterfactual world (p_C), (middle bottom) PR/FAP ($PR = p_F/p_C$), and (bottom) PR relative to year 2003 [$PR_{rel} = p_F(t)/p_F(t_e)$], that is, the year on which the event occurred (vertical bar). Changes in intensity are analyzed in terms of (top) intensity in the factual world (I_F), (middle top) intensity in the counterfactual world (I_C), (middle bottom) difference between these two [$\Delta I = I_F(t) - I_C(t)$], or (bottom) difference in the factual world with respect to year 2003 [$\Delta I_{rel} = I_F(t) - I_F(t_e)$]. Shaded areas correspond to 5%–95% confidence regions.

magnitude of the event (now defined by its occurrence probability) increases by about 6°C. These diagnostics suggest that recent changes are large in terms of probability ratios (PR near 10 in 2003) while remaining limited in terms of magnitude (near 1°C in 2003) in the CNRM-CM5 model. The influence of natural forcings is clearly discernible, and mainly driven by large volcanic eruptions. Consistent with Fig. 1, both frequency and intensity exhibit a discernible human influence as early as 1970 in this model. Human influence becomes huge during the twenty-first century, with PR higher than 10^3 in 2100. Overall, confidence intervals are relatively narrow but are consistent with the estimated changes in X (which exhibits limited uncertainty; Fig. 1) and the fact that there is a clear

relationship between X and Y (Fig. 2). The latter implies that any significant change in X translates into a significant change in Y .

4. Multimodel perspective and synthesis

a. Results from CMIP5 models

To give a broader picture, this procedure can then be applied to other CMIP models (Fig. 4). This reveals that model uncertainty is large—unlike estimation (or sampling) uncertainty, which remains very limited. Model best estimates of PR vary from 1.7 to more than 300 at the date of the event. Discrepancies among models are also very large in terms of ΔI , from 0.2° to 3°C in 2003.

FIG. 4. Attribution diagnostics in year 2003, for a collection of CMIP5 models.

FIG. 5. Multimodel statistics and synthesis. (left) Results for the 23 individual CMIP5 models considered are shown in terms of changes in (top) covariate X (degrees with respect to the reference year 2003), (middle) probability ratio (PR), and (bottom) ΔI . (center) The multimodel distribution estimated from this sample of models is illustrated through: the mean and 90% confidence region (i.e., confidence region of θ_0 , or resulting diagnostics, using the “models are statistically indistinguishable from the truth” paradigm), or (right) new realizations drawn from that distribution (which can be interpreted as virtual climate models).

Similar findings are made on the other parameters involved: p_C , p_F , I_C , and I_F —keeping in mind that model biases contribute substantially to discrepancies in I_C and I_F .

Unlike CNRM-CM5, some individual models exhibit a significant cooling trend (e.g., FGOALS-g2, ACCESS1.3, all versions of MIROC and CMCC) or warming trend [e.g., BCC_CSM1.1(m), INM-CM4.0, GISS-E2-R] in X_{all} prior to 1950 (Fig. 5a)—a period over which the anthropogenic forcings are limited. Most of this trend is

interpreted as resulting from human influence (i.e., falls into X_{ant}) according to the simple decomposition described in section 3a. Such trends typically result in PR becoming significantly different from 1, and ΔI significantly different from 0, soon after 1850 (Figs. 5d,g). At this stage it is unclear whether these trends (i) are related to low-frequency internal variability that is inappropriately taken into account, (ii) can be explained by a long-term regional drift (imbalance) in preindustrial control simulations, or (iii) highlight an early onset

of the anthropogenic influence, dominated by either aerosols or greenhouse gases (GHGs).

Though large, discrepancies among models can be easily understood. Models disagree on the magnitude of the changes in the covariate X (different global or local sensitivity to forcings), the variance of Y (which strongly influences the probability of exceeding a high threshold), and the strength of the relationship between Y and X . Each model might exhibit some bias in one of these characteristics. This highlights the need for a multimodel synthesis.

b. Building a multimodel synthesis

Techniques for building a multimodel synthesis have received much attention in both the literature and the IPCC assessments, due to their importance in providing climate change projections for the next century, including an assessment of uncertainty (Collins et al. 2013). Literature on the subject of how to use an ensemble of opportunity such as the CMIP ensemble, that is, where no particular design effort is made to cover the range of uncertainty (Tebaldi and Knutti 2007; Knutti et al. 2010a,b), is relatively abundant. However, these efforts have not been translated into event attribution thus far to the best of our knowledge, although multimodel event attribution studies have become more popular (e.g., Sun et al. 2014; Kirchmeier-Young et al. 2017, 2019; Stone et al. 2019). In this section we introduce one possible method for conducting such a synthesis in the context of the statistical framework described above. Here we propose a technique similar to that outlined in Ribes et al. (2017); we review the main concepts here but refer to that publication for a more detailed discussion.

Following section 3, the parameters describing the response of one single model are $\theta = [X_{\text{all}}(t), X_{\text{nat}}(t), \gamma]$ —all diagnostics can be derived from θ . Let θ_i be the true value of θ in model i ($i = 1, \dots, M$), let $\hat{\theta}_i$ be its estimator (as derived from section 3), and let θ^o be the true value of θ in the real world. The key idea behind the multimodel synthesis is to assume that $(\theta_i)_{i=1,\dots,M}$ are realizations of one multimodel distribution. Further, it is assumed that the truth θ^o is also a possible realization of this multimodel distribution—a paradigm known as *models are statistically indistinguishable from the truth* (Annan and Hargreaves 2010), or *model-truth exchangeability* (Rougier et al. 2013). These are strong assumptions, given that we do not know how the available sample of models was drawn from the space of plausible representations of the climate. There is no consensus on whether these assumptions lead to well-calibrated probabilistic forecasts (Collins et al. 2013), but some authors report empirical evidence that they are consistent with

past observations, in particular, for temperature (Annan and Hargreaves 2010; van Oldenborgh et al. 2013). In the following, this multimodel distribution is estimated using a Gaussian assumption. Multimodel statistics such as confidence regions can then be derived.

In more detail, we assume that (for $i = 1, \dots, m$)

$$\theta_i \sim N(\mu_m, \Sigma_m), \quad \text{and} \quad \hat{\theta}_i | \theta_i \sim N(\theta_i, \Sigma_{\theta,i}), \quad (10)$$

leading to

$$\hat{\theta}_i \sim N(\mu_m, \Sigma_m + \Sigma_{\theta,i}), \quad (11)$$

where θ_i is the value of θ for model i , $\hat{\theta}_i$ is its estimate, μ_m and Σ_m are the mean and variance of the multimodel population (i.e., Σ_m represents modeling uncertainty on θ), and $\Sigma_{\theta,i}$ describes the uncertainty related to internal variability in the estimation of θ_i . For each model, $\hat{\theta}_i$ and $\Sigma_{\theta,i}$ can be derived from the estimation procedure described in section 3. The next step is to estimate μ_m and Σ_m from the available sample of $\hat{\theta}_i$ —we refer to Ribes et al. (2017) for this technical step. Last, a prediction region for θ^o (i.e., a draw of a new value of θ) can be derived from (μ_m, Σ_m) .

Given a collection of CMIP models such as in Fig. 4, our procedure can be used to derive multimodel statistics and confidence regions (Fig. 5, and “MULTI” confidence ranges in Fig. 4). The fitted multimodel distribution can also be used to sample new realizations (using Monte Carlo simulations) corresponding to virtual climate models (Fig. 5), which is a way to check that the fitted distribution is consistent with the model sample.

Multimodel uncertainty is found to be much larger than the sampling uncertainty related to internal variability in one given model. This is not surprising for a monthlong temperature event such as the one investigated here and is consistent with many other studies (e.g., Hawkins and Sutton 2009)—but might not be the case with other variables or temporal scales. The multimodel confidence range (5%–95%) for PR is about [1.4, 260] in 2003, which better reflects the overall uncertainty than single-model estimates. It is worth noting that the reported multimodel confidence regions are not equal to the range of single-model results. Some models can be excluded from the final confidence region if they are outliers in terms of θ . And, in the presence of a very large sample of models, the bounds of the multimodel confidence region would converge to the corresponding quantiles of the model sample.

This uncertainty range in PR is larger than reported by Stott et al. (2004). Two important features of our method could contribute to this discrepancy, in addition to the different event definitions. First, the ensemble of models considered here is larger than in any other

attribution study, enabling a more comprehensive exploration of uncertainties—resulting in larger confidence intervals. In particular, differences in the variance of Y strongly contribute to the spread in p_F . Second, the attribution performed here is less constrained than other approaches. For instance, Stott et al. (2004) used historical observations (as opposed to model experiments only) to estimate the attributable warming, following standard D&A practice. In the most widespread event attribution procedure, the human-induced SST warming is also constrained by observations (Pall et al. 2011). This highlights the benefit of incorporating observed information in our procedure—a path explored in the next section. Our results also suggest that it is critical to consider a large ensemble of models for event attribution—as for assessing many other features of climate change.

5. Merging models and observations

In this section we introduce two techniques to combine observations and information provided by climate models, using the multimodel synthesis as a starting point. Among other possible approaches, we focus on using observations to constrain changes in X and estimate the distribution of Y . Other options are briefly discussed in section 6.

a. Observed changes in X

Detection and attribution studies have long illustrated that observations might be used to derive information on, for example, human-induced warming to date (Bindoff et al. 2013), particularly when the signal-to-noise ratio is high. In contrast with this, in Figs. 5a,b, models exhibit large differences in the simulated regional warming (covariate X) to date: 0.5° to 2.0°C in 2015. It is thus natural to investigate whether available observations enable us to narrow this range further.

In mathematical terms, real observations of covariable X , say X^o , can be written as

$$X_t^o = X_{\text{all}}^o(t) + \varepsilon_t^o, \quad (12)$$

where X_{all}^o is the real-world response to external forcings and ε^o is the contribution of internal variability, considered to be random. Using a Bayesian perspective, multimodel uncertainty on X_{all} (derived from the multimodel synthesis, e.g., Fig. 5b) can be considered as a prior distribution for X_{all}^o , say $\pi(X_{\text{all}}^o)$. If both this prior distribution and the distribution of ε^o are known, then the posterior distribution of $X_{\text{all}}^o|X^o$ can be derived, using a conventional Bayesian technique. The distribution of $\theta^o|X^o$ (θ^o contains X_{all}^o but is larger) can be derived similarly. This derivation is particularly easy under the assumptions made in section 4, as all distributions involved

FIG. 6. Observational constraint on covariate X . Observation of covariate X (here European summer mean temperature since 1850, black points), are compared to the multimodel distribution of externally forced changes in X , that is, $\pi(X_{\text{all}}^o)$ (light pink, 5%–95% confidence region, identical to Fig. 5b). Uncertainty in changes in X decreases after applying the observational constraint, that is, $X_{\text{all}}^o|X^o$ (dark pink, 5%–95% confidence region). Best estimates before (light brown) and after (brown) applying the observational constraint are almost indistinguishable in this case, as observations are consistent with the multimodel mean estimate. All values are temperature anomalies with respect to the 1961–90 period.

[i.e., $\pi(X_{\text{all}}^o)$ and the distribution of ε] are assumed to be Gaussian. As a consequence, $\theta^o|X^o$ follows a Gaussian distribution as well, but with mean and variance different from (μ_m, Σ_m) . Last, this constrained distribution can be used, instead of the unconstrained one, to derive all results, following the same procedure as in section 4.

Application of this procedure to summer mean temperature over Europe (i.e., our covariate X) suggests that some model responses to historical forcings are inconsistent with observations (Fig. 6). This phenomenon can be explained as follows. Intuitively, $X_{\text{all}}^o(t)$ is the expectation of X^o at time t , so observations should be distributed around $X_{\text{all}}^o(t)$. An estimate of X_{all}^o is not quite plausible if it lies far away from most of observed values x_t^o . In Fig. 6, the upper and lower bounds of the multimodel distribution $\pi(X_{\text{all}}^o)$ fall almost outside the set of observations over the beginning (before 1900) or the end (after 2000) of the observed period, suggesting some inconsistency.⁷ Using observational information

⁷ The period over which such an inconsistency is discernible depends on the choice of the reference period, here 1961–90.

therefore leads to a substantial reduction of the multi-model uncertainty in changes in X . This reduction is particularly clear over the historical period: the multi-model 5%–95% confidence range of total 1850–2015 warming is $[0.5^\circ, 2.0^\circ\text{C}]$ without the use of any observations [i.e., in $\pi(X_{\text{all}}^o)$], but shrinks to $[0.9^\circ, 1.4^\circ\text{C}]$ after applying the observational constraint (i.e., in $X_{\text{all}}^o|X^o$). The spread in future warming is also substantially reduced— $[3.8^\circ, 7.7^\circ\text{C}]$ and $[4.5^\circ, 6.9^\circ\text{C}]$ for the corresponding 5%–95% confidence ranges before and after applying the observational constraint, respectively.

The impact on event attribution diagnostics is also very clear, with a sharp reduction of uncertainties in PR or ΔI (Fig. 8). The lower bound of PR is particularly affected (3.1 after accounting for observations, as opposed to 1.4 without), because some model responses exhibit almost no human-induced warming in 2003, while observations suggest that such weak responses are unlikely. The effect on ΔI is also very clear: uncertainty reduces from $[+0.1^\circ, +2.3^\circ\text{C}]$ before accounting for observations to $[+0.5^\circ, +1.5^\circ\text{C}]$ after. Overall, these results suggest that taking observational information into account is very helpful, even if done only through the covariate X , that is, at a large spatiotemporal scale.

b. Observed distribution of Y

Another way of merging climate models' outputs with real-world observations is to estimate the distribution of Y , for example, at the time of the event. Climate models exhibit all sorts of biases (e.g., biases in the mean climate, the variability, the tails of the distribution, etc.) that can make the simulated p_F (and the entire probability distribution) erroneous (Bellprat and Doblas-Reyes 2016). Figure 4 shows that estimates of p_F vary widely, and are inconsistent among models; the multi-model synthesis exhibits large uncertainty: 5×10^{-7} to 3×10^{-2} . In such a case, the historical record can bring useful information to constrain the range of model estimate. Considering observations is also consistent with the practice of many national weather services in characterizing extreme events (Klein Tank et al. 2009).

Here, we illustrate how observations of Y , say Y^o , can be used to infer p_F and, more generally, the distribution of Y at time t_e . To do this, both changes in the covariate X , that is, X_{all} , and the nonstationary coefficients $\gamma_1 = (\mu_1, \sigma_1)$ are estimated from climate models (with some uncertainty) and subsequently treated as known in the statistical inference. Then, observations are used to estimate the stationary parameters $\gamma_0 = (\mu_0, \sigma_0)$. In this way, models are used to estimate some parameters (X_{all} and γ_1) while observations are used to estimate others (γ_0). As we make no use of the value of γ_0 previously

estimated in each model, the proposed treatment cannot be really considered as a *constraint*. This approach could rather be considered as a form of bias correction.

Under the Gaussian assumption, the parameters (μ_0, σ_0) of Eq. (8) have to be estimated. Given estimates of γ_1 and X_{all} , μ_0 can be naturally estimated by

$$\hat{\mu}_0 = \overline{y_t^o - \mu_1 X_{\text{all}}(t)}, \quad (13)$$

where the overbar denotes an average, y_t^o are the available observations of Y , and y_t^o , μ_1 and X_{all} are estimated from models. Then, σ_0 can be estimated by

$$\hat{\sigma}_0 = \text{sd} \left(\frac{y_t^o - \mu_1 X_{\text{all}}(t) - \hat{\mu}_0}{1 + \sigma_1 X_{\text{all}}(t)} \right), \quad (14)$$

where sd is the sample standard deviation, denoting again that every term in the right-hand side is known, and that $[y_t^o - \mu_1 X_{\text{all}}(t) - \hat{\mu}_0] \sim N\{0, \sigma_0[1 + \sigma_1 X_{\text{all}}(t)]\}$. Note that these estimators ($\hat{\mu}_0, \hat{\sigma}_0$) do not necessarily coincide with the maximum likelihood estimators but are, however, quite natural and attractive for computational reasons. The uncertainty of these parameters can be assessed by extending the bootstrap procedure to y_t^o (i.e., resampling observations y_t^o , as would be done in a stationary context), and considering simultaneously uncertainty in μ_1 , σ_1 , and X_{all} , as derived from the multimodel synthesis.

Our procedure is illustrated in Fig. 7. As the CanESM2 model simulates a larger change than CNRM-CM5, the 2003 event is relatively less abnormal according to that model, resulting in a larger estimate of p_F (changes in variance are small and do not substantially influence the results in this case). Model discrepancies in estimates of p_F are therefore largely related to the spread in the nonstationary term $\mu_1 X_{\text{all}}$.

Applying this procedure to all single models and to the multimodel synthesis leads to reduced uncertainty in the estimate of p_F (Fig. 8). Estimates of p_C are similarly improved. However, the impact on attribution diagnostics, that is, PR and ΔI , is quite limited. The proposed procedure refines the estimation of the Y distribution, but does not affect the estimation of human influence, and so coefficients measuring that influence are only indirectly impacted.

c. Applying the two constraints together

The two constraints presented above can also be applied simultaneously. If so, observations of X are used to constrain X_{all} first, then parameters γ_0 are estimated using observations of Y , given $(X_{\text{all}}, \gamma_1)$. Therefore, observed information is used in both X and γ_0 , in addition to model information. Note, however, that we

FIG. 7. Use of corrected observations to estimate p_F . (top) Observed record y_i^o (black line) is compared to the changes in Y [i.e., $\mu_1 X_{\text{all}}(t)$] as simulated by two climate models (CNRM-CM5 and CanESM2, color lines). Temperatures are anomalies with respect to the 1961–90 average. (middle),(bottom) Observed time series after correction for the mean change simulated by the models, that is, $y_i^o - \mu_1 X_i$. Correction is made such that the 2003 value is unchanged; p_F denotes the probability of exceeding the threshold (dotted line), as estimated from these corrected records. Uncertainty analysis and correction for changes in variance are not represented.

neglect potential dependencies between observations of X and Y . Results obtained in this way can be considered as the final results of our procedure described, as they combine all sources of information (Figs. 8 and 9 for the multimodel synthesis only).

Applying the two constraints simultaneously leads to a real narrowing of uncertainties in estimates of probabilities $p_F(t_e)$ and $p_C(t_e)$ (where γ_0 plays a critical role), and in the derived attribution diagnostics PR and ΔI (where X_{ant} is critical; Fig. 8). Uncertainty in PR shrinks from [1.4, 260] (multimodel synthesis, no use of observations) to [3.8, 90] (applying the two constraints).

Uncertainty in ΔI is also reduced, from [$+0.1^\circ$, $+2.3^\circ\text{C}$] to [$+0.5^\circ$, $+1.5^\circ\text{C}$] (i.e., roughly by a factor of 2). In all cases considered, applying the two constraints together reduces model spread further than using one single constraint or no observations at all.

Remarkably, time series of PR and ΔI can still be derived after applying these constraints (Fig. 9), suggesting a few comments. First, human influence on the likelihood of an event like EWH03 has been significant since the mid 1980s (Figs. 9a,c). Second, the odds of observing an event such as EHW03 (in the sense of the same magnitude) have strongly increased since 2003;

FIG. 8. Impact of (combined) constraints on (a) p_C , (b) p_F , (c) PR/FAP, and (d) ΔI . All results correspond to the multimodel synthesis (section 4). NO OBS: only model outputs (no observations) are used; results are identical to those presented in Fig. 4 for the MULTI synthesis. C0: γ_0 is estimated from observations Y^o (see section 5b). CX: Estimates of changes in X are constrained by observation x_i^o (see section 5a). CX + C0: the two constraints are applied simultaneously.

they were 3 to 9 times as large in 2018 as in 2003 [qualitatively consistent with Christidis et al. (2015a)]. Third, an event similar to EHW03 (in the sense of the same frequency) occurring in 2100 under an RCP8.5 scenario would imply a human contribution as large as $+7.7^\circ\text{C}$ [$+4.7$, $+11.1$] (Fig. 9c). Last, a very large fraction of this

human-induced warming is expected to take place after 2003: $+6.8^\circ\text{C}$ [$+4$, $+9.8$].

Overall, these results suggest that our approach, in addition to covering a wide range of uncertainties through the use of a large ensemble of models, can lead to relatively constrained attribution results. They also

FIG. 9. Attribution diagnostics as estimated after applying observational constraints CX and C0. Changes in (left) frequency (PR and PR_{rel}) and (right) intensity (ΔI and ΔI_{rel}) are used to quantify the human influence on an EHW03-like event. All results are shown as a function of time, had the event occurred at that time. The vertical bar indicates the year 2003. Results are for the multimodel synthesis only, after applying the two observational constraints CX and C0 simultaneously.

suggest that, in the particular case under consideration, the unconstrained parameters γ_1 do not vary widely among models.

6. Discussion

In this section we review several aspects of our proposed method that deserve particular attention.

a. Choice of the covariate X

Selecting an appropriate covariate X is a key step in our method. The choice of this covariate is at least partly subjective and can impact the final results. Using a global or regional temperature may be appropriate, as changes in many variables have been described to scale with temperature (e.g., Collins et al. 2013; Tebaldi and Arblaster 2014). Pattern scaling, however, works better if GHGs are the only source of external forcing. In practice, other forcings also contributed to recent changes—in particular, anthropogenic aerosols. As the strength of the aerosols forcing varies considerably over space, using a regional temperature as a covariate might better reflect the regional balance between various external forcings. In any case, relying on a covariate X is a strong assumption of our method, which must be properly acknowledged.

b. Choice of the Y distribution

In the present application, our method is only illustrated with Gaussian distributions. This is a clear limitation to the field of application of that method, as many climate variables do not follow a Gaussian distribution. Then, as the EHW03 event is very rare, a Gaussian assumption might not well represent the upper tail of the distribution of temperature, potentially leading to errors in the estimation of probabilities (particularly in the counterfactual world). Extending our procedure to other distributions, in particular, GEVs or nonparametric distributions (as suggested in section 3b) will open a broader scope of application.

c. Limitations in using nonstationary statistics

The use of nonstationary statistics is central in this approach, and some limitations must be pointed out. First, a sufficiently large SNR is needed in model data in order to allow fitting of the nonstationary model. The entire procedure can fail if nonstationary coefficients cannot be estimated properly [see, e.g., Li et al. (2019) for a discussion on this limitation]. In this respect, the temperature event considered in this study was an easy one. The method will have to be tested on other events/variables (e.g., precipitation, wind), to determine the extent of its field of application. Second, further statistical

developments might improve the fit of the statistical model. In the current analysis, Y data were limited to a specific space–time domain—we ignore any information available outside this domain. Using further spatial (e.g., a broader region than that of the event; Christidis et al. 2015b) or temporal (e.g., modeling the entire seasonal cycle; Rigal et al. 2019) information might be particularly attractive, but it would involve a sharp increase in the complexity of the statistical model and inference.

d. Climate model evaluation and reliability

Many previous attribution studies have devoted much attention to the evaluation of climate models, in particular, their ability to simulate extreme events consistent with observations, and whether the models include the appropriate physical processes and forcings. This model reliability issue has sometimes been tackled through implementing model selection (e.g., King et al. 2015). Recent studies have also suggested that model biases do impact event attribution results (Bellprat and Doblas-Reyes 2016; Bellprat et al. 2019).

In this study, we do not perform any model evaluation or selection, even though using CMIP models (of relatively low resolution) somewhat strengthens this reliability concern. Instead, we use bias-correction techniques (through the C0 observational constraint, and through defining the event as an anomaly with respect to a given reference period) to correct some of the model biases. Beyond that, more sophisticated bias-correction techniques could be applied prior to the statistical analysis—potentially contributing to reconcile event attribution with the use of coarse-resolution CMIP-style models. Such a treatment, however, is only acceptable if model-simulated changes are robust and little affected by present-day climate biases.

Broader application of our technique (e.g., to other spatiotemporal scales or other variables) might require case-by-case evaluation of the models' reliability, on the basis of available literature. In various cases, model limitations can prevent reliable attribution, for example, if models are just not able to simulate the type of event considered (e.g., tropical cyclones, weather events related to a specific topography), or if some forcings are missing (e.g., irrigation, changes in surface roughness). Two adaptations of our method could be proposed to address at least some of these issues. First, our method could be easily applied to transient experiments performed with high-resolution models, such as Coordinated Regional Climate Downscaling Experiment (CORDEX) models—following a general recommendation to use high-resolution models (National Academies of Sciences, Engineering, and Medicine 2016, and references therein). Note, however,

that limiting the analysis to a small number of models can have undesirable effects on the uncertainty analysis. Second, inconsistencies between models and observations might be identified by additional observational constraints, as discussed below.

e. Uncertainty quantification and modeling uncertainty

One key outcome of our analysis is that modeling uncertainty is critical in event attribution. Uncertainty ranges vary greatly in size if derived using one model only versus a multimodel ensemble, with ranges far too narrow in the former case. The technique proposed to build the multimodel synthesis could be improved in many ways, for example, by using a link function for some parameters in θ , a non-Gaussian dependence structure, or another paradigm than the model truth exchangeability. Last, the uncertainty ranges derived from the multimodel synthesis are relatively large, but it still might be useful to check that observations are consistent with these—no such check was implemented here.

f. Additional observational constraints

Section 5 explored two possible ways to use observations to constrain attribution results, in which observations of Y are only used to estimate the *stationary* parameters γ_0 . A natural extension would be to also constrain γ_1 , that is, the magnitude of changes in Y . This could be done using a Bayesian approach, in which the multimodel uncertainty on γ_1 is used as a prior distribution.

g. Consistency with other approaches

Assessing the consistency of our results with previous studies that also focused on the EWH03 event is not easy, primarily because variations in the event definition can contribute to discrepancies in the results (Cattiaux and Ribes 2018). Roughly speaking, our results in terms of PR lie somewhere between Stott et al. (2004) and Christidis et al. (2015a). Our estimate of ΔI seems consistent with the figures in Christidis et al. (2015a). Providing closer comparisons between our approach and other event attribution methods will be of primary interest in the future. From another perspective, our approach provides an adaptation of projections uncertainty plumes to a specific event, sharing some common features with the treatment used in IPCC AR5 (e.g., use of CMIP models, treatment of uncertainties).

7. Conclusions

This study describes and illustrates a new statistical method for event attribution that can be decomposed

into three steps. First, event attribution diagnostics are derived from transient CMIP-style experiments using nonstationary statistics and an appropriate covariate. Single-model results derived from this step typically exhibit large discrepancies. Second, a multimodel synthesis is performed, assuming model–truth exchangeability. Evidence suggests that this synthesis might represent uncertainty better than single-model analyses. Third, multimodel information is combined with historical observations in order to account for all sources of information available. This blending typically reduces uncertainty in the final attribution diagnostics, while providing a more comprehensive description of the event and human influence on it.

This study illustrates that it is possible to perform event attribution using available transient simulations. This is an important result because the use of such experiments offers several advantages. In particular, it offers the possibility of characterizing the human influence on a singular event in a way that is consistent with long-term projections, that is, using the same data and a similar quantification of uncertainty. The calculation of uncertainty plumes covering both the past and future also provides a new perspective on the human influence on a singular event. And, obviously, reusing available simulations could save much computation time and effort.

Overall, this method could facilitate communication about the human influence on a particular event, as the diagnostics it provides are, by construction, consistent with other long-term indicators of climate change. The method is also promising in that it allows a rapid analysis of events, as all input data are already available. Testing this approach on a broad range of event types and scales will be necessary before any systematic application.

Acknowledgments. We acknowledge Geert Jan van Oldenborgh, Francis Zwiers, and one anonymous reviewer for their constructive and insightful comments on the previous versions of this paper. Part of this work was supported by the French Ministère de la Transition Énergétique et Solidaire (through *Extremoscope* and *Convention pour les Services Climatiques* grants), the European Research Area for Climate Services (through the Eupheme project, Grant 690462), and the French National Research Agency (ANR, through the REMEMBER project, Contract ANR-12-SENV-001). The authors are supported by Météo France (AR), CEA (ST), and CNRS (JC). The R scripts used to make the analysis described in this paper will be made available on the personal worldwide web page of author Aurélien Ribes.

Data availability statement: All data used in this study are publicly available. 1) HadCRUT4 observations are available at <https://crudata.uea.ac.uk/cru/data/temperature/>. 2) Outputs of CMIP5 simulations are available at <https://esgf-node.llnl.gov/projects/cmip5/>. Parameters of the EBMs were taken from Geoffroy et al. (2013). Time series of natural forcings used to compute EBMs response were taken from IPCC AR5, Annex II (IPCC 2013).

REFERENCES

- Allen, M., 2003: Liability for climate change. *Nature*, **421**, 891–892, <https://doi.org/10.1038/421891a>.
- Annan, J. D., and J. C. Hargreaves, 2010: Reliability of the CMIP3 ensemble. *Geophys. Res. Lett.*, **37**, L02703, <https://doi.org/10.1029/2009GL041994>.
- Bellprat, O., and F. Doblas-Reyes, 2016: Attribution of extreme weather and climate events overestimated by unreliable climate simulations. *Geophys. Res. Lett.*, **43**, 2158–2164, <https://doi.org/10.1002/2015GL067189>.
- , V. Guemas, F. Doblas-Reyes, and M. G. Donat, 2019: Towards reliable extreme weather and climate event attribution. *Nat. Commun.*, **10**, 1732, <https://doi.org/10.1038/s41467-019-09729-2>.
- Bindoff, N. L., and Coauthors, 2013: Detection and attribution of climate change: From global to regional. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 867–952.
- Cattiaux, J., and A. Ribes, 2018: Defining single extreme weather events in a climate perspective. *Bull. Amer. Meteor. Soc.*, **99**, 1557–1568, <https://doi.org/10.1175/BAMS-D-17-0281.1>.
- , R. Vautard, C. Cassou, P. Yiou, V. Masson-Delmotte, and F. Codron, 2010: Winter 2010 in Europe: A cold extreme in a warming climate. *Geophys. Res. Lett.*, **37**, L20704, <https://doi.org/10.1029/2010GL044613>.
- Christidis, N., P. A. Stott, A. A. Scaife, A. Arribas, G. S. Jones, D. Copey, J. R. Knight, and W. J. Tennant, 2013: A new HadGEM3-A-based system for attribution of weather- and climate-related extreme events. *J. Climate*, **26**, 2756–2783, <https://doi.org/10.1175/JCLI-D-12-00169.1>.
- , G. S. Jones, and P. A. Stott, 2015a: Dramatically increasing chance of extremely hot summers since the 2003 European heatwave. *Nat. Climate Change*, **5**, 46–50, <https://doi.org/10.1038/nclimate2468>.
- , P. A. Stott, and F. W. Zwiers, 2015b: Fast-track attribution assessments based on pre-computed estimates of changes in the odds of warm extremes. *Climate Dyn.*, **45**, 1547–1564, <https://doi.org/10.1007/s00382-014-2408-x>.
- Ciavarella, A., and Coauthors, 2018: Upgrade of the HadGEM3-A based attribution system to high resolution and a new validation framework for probabilistic event attribution. *Wea. Climate Extremes*, **20**, 9–32, <https://doi.org/10.1016/J.WACE.2018.03.003>.
- Coles, S., J. Bawa, L. Trenner, and P. Dorazio, 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics, Vol. 208, Springer, 209 pp.
- Collins, M., and Coauthors, 2013: Long-term climate change: Projections, commitments and irreversibility. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1029–1136.
- Dong, B., R. T. Sutton, L. Shaffrey, and N. P. Klingaman, 2017: Attribution of forced decadal climate change in coupled and uncoupled ocean–atmosphere model experiments. *J. Climate*, **30**, 6203–6223, <https://doi.org/10.1175/JCLI-D-16-0578.1>.
- Eden, J. M., K. Wolter, F. E. Otto, and G. J. van Oldenborgh, 2016: Multi-method attribution analysis of extreme precipitation in Boulder, Colorado. *Environ. Res. Lett.*, **11**, 124009, <https://doi.org/10.1088/1748-9326/11/12/124009>.
- Geoffroy, O., D. Saint-Martin, D. J. Olivié, A. Voldoire, G. Bellon, and S. Tytca, 2013: Transient climate response in a two-layer energy-balance model. Part I: Analytical solution and parameter calibration using CMIP5 AOGCM experiments. *J. Climate*, **26**, 1841–1857, <https://doi.org/10.1175/JCLI-D-12-00195.1>.
- Hastie, T. J., Ed., 2017: Generalized additive models. *Statistical models in S*, Routledge, 249–307.
- Hauser, M., and Coauthors, 2017: Methods and model dependency of extreme event attribution: The 2015 European drought. *Earth's Future*, **5**, 1034–1043, <https://doi.org/10.1002/2017EF000612>.
- Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1108, <https://doi.org/10.1175/2009BAMS2607.1>.
- Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis, 2010: Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *J. Climate*, **23**, 2418–2427, <https://doi.org/10.1175/2009JCLI3466.1>.
- Huber, M., and R. Knutti, 2012: Anthropogenic and natural warming inferred from changes in Earth's energy balance. *Nat. Geosci.*, **5**, 31–36, <https://doi.org/10.1038/ngeo1327>.
- IPCC, 2013: Annex II: Climate system scenario tables. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker, Eds., Cambridge University Press, 1395–1446.
- King, A. D., G. J. van Oldenborgh, D. J. Karoly, S. C. Lewis, and H. Cullen, 2015: Attribution of the record high central England temperature of 2014 to anthropogenic influences. *Environ. Res. Lett.*, **10**, 054002, <https://doi.org/10.1088/1748-9326/10/5/054002>.
- Kirchmeier-Young, M. C., F. W. Zwiers, and N. P. Gillett, 2017: Attribution of extreme events in Arctic sea ice extent. *J. Climate*, **30**, 553–571, <https://doi.org/10.1175/JCLI-D-16-0412.1>.
- , N. P. Gillett, F. W. Zwiers, A. J. Cannon, and F. S. Anslow, 2019: Attribution of the influence of human-induced climate change on an extreme fire season. *Earth's Future*, **7**, 2–10, <https://doi.org/10.1029/2018EF001050>.
- Klein Tank, A. M., F. W. Zwiers, and X. Zhang, 2009: Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation. Climate Data and Monitoring WCDMP-72, WMO/TD-1500, 56 pp, http://www.wmo.int/pages/prog/wcp/wcdmp/wcdmp_series/documents/WCDMP_72_TD_1500_en_1.pdf.
- Knutti, R., G. Abramowitz, M. Collins, V. Eyring, P. J. Gleckler, B. Hewitson, and L. Mearns, 2010a: Good practice guidance paper on assessing and combining multi model climate projections. IPCC Expert Meeting on Multi Model Evaluation, T. F. Stocker et al., Eds., IPCC Rep. 0165–0009, 15 pp., https://wg1.ipcc.ch/guidancepaper/IPCC_EM_MME_GoodPracticeGuidancePaper.pdf.
- , R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl, 2010b: Challenges in combining projections from multiple climate

- models. *J. Climate*, **23**, 2739–2758, <https://doi.org/10.1175/2009JCLI3361.1>.
- Koenker, R., and G. Bassett Jr., 1978: Regression quantiles. *Econometrica*, **46**, 33–50, <https://doi.org/10.2307/1913643>.
- , and K. F. Hallock, 2001: Quantile regression. *J. Econ. Perspect.*, **15**, 143–156, <https://doi.org/10.1257/jep.15.4.143>.
- Lewis, S. C., and D. J. Karoly, 2013: Anthropogenic contributions to Australia's record summer temperatures of 2013. *Geophys. Res. Lett.*, **40**, 3705–3709, <https://doi.org/10.1002/grl.50673>.
- Li, C., F. Zwiers, X. Zhang, and G. Li, 2019: How much information is required to well constrain local estimates of future precipitation extremes? *Earth's Future*, **7**, 11–24, <https://doi.org/10.1029/2018EF001001>.
- Massey, N., and Coauthors, 2015: Weather@home—Development and validation of a very large ensemble modelling system for probabilistic event attribution. *Quart. J. Roy. Meteor. Soc.*, **141**, 1528–1545, <https://doi.org/10.1002/QJ.2455>.
- Morice, C., J. Kennedy, N. Rayner, and P. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res.*, **117**, D08101, <https://doi.org/10.1029/2011JD017187>.
- National Academies of Sciences, Engineering, and Medicine, 2016: *Attribution of Extreme Weather Events in the Context of Climate Change*. National Academies Press, 186 pp.
- Otto, F. E., N. Massey, G. Van Oldenborgh, R. Jones, and M. Allen, 2012: Reconciling two approaches to attribution of the 2010 Russian heat wave. *Geophys. Res. Lett.*, **39**, L04702, <https://doi.org/10.1029/2011GL050422>.
- Paciorek, C. J., D. A. Stone, and M. F. Wehner, 2018: Quantifying statistical uncertainty in the attribution of human influence on severe weather. *Wea. Climate Extremes*, **20**, 69–80, <https://doi.org/10.1016/j.wace.2018.01.002>.
- Pall, P., T. Aina, D. A. Stone, P. A. Stott, T. Nozawa, A. G. Hilberts, D. Lohmann, and M. R. Allen, 2011: Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature*, **470**, 382–385, <https://doi.org/10.1038/nature09762>.
- Peterson, T. C., P. A. Stott, and S. Herring, 2012: Explaining Extreme Events of 2011 from a Climate Perspective. *Bull. Amer. Meteor. Soc.*, **93**, 1041–1067, <https://doi.org/10.1175/BAMS-D-12-00021.1>.
- Portnoy, S., R. Koenker, R. A. Thisted, M. R. Osborne, S. Portnoy, and R. Koenker, 1997: The Gaussian hare and the Laplacian tortoise: Computability of squared-error versus absolute-error estimators. *Stat. Sci.*, **12**, 279–300, <https://doi.org/10.1214/ss/1030037960>.
- Ribes, A., F. W. Zwiers, J.-M. Azaïs, and P. Naveau, 2017: A new statistical approach to climate change detection and attribution. *Climate Dyn.*, **48**, 367–386, <https://doi.org/10.1007/s00382-016-3079-6>.
- Rigal, A., J.-M. Azaïs, and A. Ribes, 2019: Estimating daily climatological normals in a changing climate. *Climate Dyn.*, **53**, 275–286, <https://doi.org/10.1007/s00382-018-4584-6>.
- Risser, M. D., D. A. Stone, C. J. Paciorek, M. F. Wehner, and O. Angéilil, 2017: Quantifying the effect of interannual ocean variability on the attribution of extreme climate events to human influence. *Climate Dyn.*, **49**, 3051–3073, <https://doi.org/10.1007/s00382-016-3492-x>.
- Rougier, J., M. Goldstein, and L. House, 2013: Second-order exchangeability analysis for multimodel ensembles. *J. Amer. Stat. Assoc.*, **108**, 852–863, <https://doi.org/10.1080/01621459.2013.802963>.
- Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller, 2004: The role of increasing temperature variability in European summer heatwaves. *Nature*, **427**, 332–336, <https://doi.org/10.1038/nature02300>.
- Stone, D. A., and Coauthors, 2019: Experiment design of the International CLIVAR C20C+ Detection and Attribution project. *Wea. Climate Extremes*, **24**, 100206, <https://doi.org/10.1016/J.WACE.2019.100206>.
- Stott, P., D. Stone, and M. Allen, 2004: Human contribution to the European heatwave of 2003. *Nature*, **432**, 610–614, <https://doi.org/10.1038/nature03089>.
- Sun, Y., X. Zhang, F. W. Zwiers, L. Song, H. Wan, T. Hu, H. Yin, and G. Ren, 2014: Rapid increase in the risk of extreme summer heat in eastern China. *Nat. Climate Change*, **4**, 1082–1085, <https://doi.org/10.1038/nclimate2410>.
- Tebaldi, C., and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. Roy. Soc. London*, **365A**, 2053–2075, <https://doi.org/10.1098/rsta.2007.2076>.
- , and J. Arblaster, 2014: Pattern scaling: Its strengths and limitations, and an update on the latest model simulations. *Climatic Change*, **122**, 459–471, <https://doi.org/10.1007/s10584-013-1032-9>.
- Trenberth, K. E., J. T. Fasullo, and T. G. Shepherd, 2015: Attribution of climate extreme events. *Nat. Climate Change*, **5**, 725–730, <https://doi.org/10.1038/nclimate2657>.
- Uhe, P., F. Otto, K. Haustein, G. van Oldenborgh, A. King, D. Wallom, M. Allen, and H. Cullen, 2016: Comparison of methods: Attributing the 2014 record European temperatures to human influences. *Geophys. Res. Lett.*, **43**, 8685–8693, <https://doi.org/10.1002/2016GL069568>.
- van der Wiel, K., and Coauthors, 2017: Rapid attribution of the August 2016 flood-inducing extreme precipitation in south Louisiana to climate change. *Hydrol. Earth Syst. Sci.*, **21**, 897–921, <https://doi.org/10.5194/hess-21-897-2017>.
- van Oldenborgh, G. J., F. J. Doblas Reyes, S. S. Drijfhout, and E. Hawkins, 2013: Reliability of regional climate model trends. *Environ. Res. Lett.*, **8**, 014055, <https://doi.org/10.1088/1748-9326/8/1/014055>.
- , F. E. Otto, K. Haustein, and H. Cullen, 2015: Climate change increases the probability of heavy rains like those of storm Desmond in the UK—An event attribution study in near-real time. *Hydrol. Earth Syst. Sci. Discuss.*, **12**, 13 197–13 216, <https://doi.org/10.5194/hessd-12-13197-2015>.
- , and Coauthors, 2017: Attribution of extreme rainfall from Hurricane Harvey, August 2017. *Environ. Res. Lett.*, **12**, 124009, <https://doi.org/10.1088/1748-9326/AA9EF2>.
- Vautard, R., and Coauthors, 2015: Extreme fall 2014 precipitation in the Cévennes mountains [in “Explaining Extreme Events of 2014 from a Climate Perspective”]. *Bull. Amer. Meteor. Soc.*, **96** (12), S56–S60, <https://doi.org/10.1175/BAMS-D-15-00088.1>.
- Wehner, M., D. Stone, H. Shioyama, P. Wolski, A. Ciavarella, N. Christidis, and H. Krishnan, 2018: Early 21st century anthropogenic changes in extremely hot days as simulated by the C20C+ detection and attribution multi-model ensemble. *Wea. Climate Extremes*, **20**, 1–8, <https://doi.org/10.1016/j.wace.2018.03.001>.
- Wood, S. N., 2017: *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 410 pp.