



HAL
open science

Describing the relationship between a weather event and climate change: a new statistical approach.

Aurélien Ribes, Soulivanh Thao, Julien Cattiaux

► To cite this version:

Aurélien Ribes, Soulivanh Thao, Julien Cattiaux. Describing the relationship between a weather event and climate change: a new statistical approach.. 2019. hal-02122780

HAL Id: hal-02122780

<https://hal.science/hal-02122780>

Preprint submitted on 7 May 2019

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.

1 **Describing the relationship between a weather event and climate change: a**
2 **new statistical approach**

3 Aurélien Ribes*

4 *CNRM, Météo France - CNRS, Toulouse, France*

5 Soulivanh Thao

6 *LSCE, IPSL, Gif-sur-Yvette, France*

7 Julien Cattiaux

8 *CNRM, Météo France - CNRS, Toulouse, France*

9 *Corresponding author address: Aurélien Ribes, CNRM, Météo France - CNRS, 42 Avenue Gas-
10 pard Coriolis, Toulouse, France

11 E-mail: aurelien.ribes@meteo.fr

ABSTRACT

12 Describing the relationship between a weather event and climate change –
13 a science usually termed *event attribution* – involves quantifying the extent
14 to which human influence has affected the frequency or the strength of an
15 observed event. In this study we show how event attribution can be imple-
16 mented through the application of non-stationary statistics to transient simu-
17 lations, typically covering the 1850-2100 period. The use of existing CMIP-
18 style simulations has many advantages, including their availability for a large
19 range of coupled models, and the fact that they are not conditional to a given
20 oceanic state. We develop a technique for providing a multi-model synthe-
21 sis, consistent with the uncertainty analysis of long-term changes. Lastly, we
22 describe how model estimates can be combined with historical observations
23 to provide a single diagnosis accounting for both sources of information. The
24 potential of this new method is illustrated using the 2003 European Heat Wave
25 and under a Gaussian assumption. Results suggest that (i) it is feasible to per-
26 form event attribution using transient simulations and non-stationary statis-
27 tics, even for a single model, (ii) the use of multi-model synthesis in event
28 attribution is highly desirable given the spread in single model estimates, and
29 (iii) merging models and observations substantially reduces uncertainties in
30 human-induced changes. Investigating transient simulations also enables us
31 to derive insightful diagnoses of how the targeted event will be affected by
32 climate change in the future.

33 **1. Introduction**

34 Describing the relationship between a given weather or climate event and anthropogenic climate
35 change is a growing area of activity in the field of climate science (National Academies of Sciences
36 and Medicine 2016). Since the pioneering studies of (Allen 2003; Stott et al. 2004), the concept of
37 event attribution has been applied to a wide variety of events, as synthesized in the annual special
38 issues of BAMS “Explaining extreme events in a climate perspective” (Peterson et al. 2012, and
39 subsequent issues¹).

40 Multiple approaches have been introduced to address this question. Beyond issues related to the
41 definition of the event of interest, the most commonly used approach is probabilistic, and involves
42 a comparison of the distributions of extreme events in the factual vs counterfactual worlds (Stott
43 et al. 2004; Pall et al. 2011, e.g.). Particular attention is paid to changes in probability of the event
44 associated with human influence. Various alternatives have been proposed in the literature; one
45 of these involves focusing on the thermodynamic component of human influence (Trenberth et al.
46 2015; Cattiaux et al. 2010, e.g.). However, this study will focus on the probabilistic approach
47 and its statistical implementation, i.e. how estimating changes in occurrence frequency and the
48 corresponding uncertainty.

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

50 First, large ensembles of simulations are used to sample the factual and counter-factual statisti-
51 cal distributions (Pall et al. 2011; Massey et al. 2015; Christidis et al. 2013; Ciavarella et al. 2018;
52 Wehner et al. 2018). Such ensembles are typically produced with atmospheric-only models forced
53 by prescribed sea surface temperatures (SSTs); factual SSTs are usually taken from observations,
54 while counterfactual SSTs are derived by subtracting an estimate of the anthropogenic influence.
55 Such ensembles can be very large, typically from a few hundred to more than 10.000 simulations

¹see <https://www.ametsoc.org/index.cfm/ams/publications/bulletin-of-the-american-meteorological-society-bams/explaining->

56 of one year or one season. One important advantage of using such large ensembles is that the sig-
57 nal to noise ratio is increased (sampling noise is reduced), and probabilities can be estimated very
58 straightforwardly by counting exceedances, i.e. using a minimal statistical inference – although
59 more complex treatments have also been proposed (Paciorek et al. 2018). Several disadvantages
60 should also be mentioned: the computational cost is relatively high (large number of simulations,
61 which have to be redone on an annual basis at least), processes involving ocean-atmosphere cou-
62 pling are missing (Dong et al. 2017), results are conditional on the particular pattern of SSTs
63 considered (Risser et al. 2017), model bias or reliability issues can affect results (Bellprat and
64 Doblas-Reyes 2016), and lastly, modelling uncertainty is usually not assessed comprehensively
65 due to the lack of coordinated exercise (Wehner et al. 2018, is a noticeable exception).

66 Second, occurrence probabilities can be inferred from observations and observed trends, assum-
67 ing that the trends are entirely related to human influence on climate (van Oldenborgh et al. 2015;
68 Vautard et al. 2015, e.g.). This approach eliminates all concerns related to model bias and/or error
69 in representing climate change. However, one strong limitation is that the signal to noise ratio
70 is usually limited in observations – climate change to date might be relatively small compared to
71 internal variability. In many cases, observations do not provide evidence for any significant trend,
72 while models do suggest sensitivity to anthropogenic forcings. Even if a significant trend is found,
73 uncertainty in the trend estimate might lead to very wide uncertainty in the risk ratio or other di-
74 agnoses of the human influence. Further, this technique is highly dependent on the availability of
75 a long, homogeneous historical record – and such data are not always available.

76 A few attempts have been made to consider these two approaches simultaneously (Uhe et al.
77 2016; Eden et al. 2016; Hauser et al. 2017). These studies provide very helpful comparisons
78 of methods for selected case studies. However, to the best of our knowledge, there has been

79 no real attempt to combine the two available sources of information together (i.e. models and
80 observations) in order to provide one single estimate of human influence.

81 In this paper, we tackle several of these issues. First, we propose to base event attribution on
82 transient CMIP-style simulations – typically a combination of historical and RCP scenarios. This
83 is done through the use of non-stationary statistics (Section 3). Second, we propose a statistical
84 procedure to create a rigorous multi-model synthesis. This question has not been fully addressed in
85 previous event attribution literature, primarily because no large multi-model ensemble was avail-
86 able. We show that, if such an ensemble were available, the assumptions and techniques used
87 to construct multi-model syntheses for large scale mean variables could be extended to event at-
88 tribution (Section 4). Third, we introduce a statistical framework for merging information from
89 models and observations. The proposed method is essentially Bayesian, in the sense that available
90 observations are used to constrain the model range further (Section 5).

91 Using transient CMIP-style simulations for event attribution is not a new idea (Lewis and Karoly
92 2013; King et al. 2015). The main issue with such simulations is that the sample size is limited
93 – usually to no more than 10 members. This is at least partly compensated by the fact that these
94 simulations include a period of time (near the end of the 21st century) in which the human influ-
95 ence is much more pronounced than in the current climate, resulting in a higher signal to noise
96 ratio. Another potential concern is related to the capacity of CMIP models to simulate extreme
97 events adequately – a point that we carefully discuss below. Regardless, there are tremendous
98 advantages in using such simulations: they are already available (dragging the computational cost
99 down to almost zero), performed with fully-coupled models (i.e. accounting for coupled processes,
100 and also not conditional on a specific oceanic state), and available for many models (allowing the
101 possibility to account for modelling uncertainty in a comprehensive way, consistent with IPCC
102 practice).

103 In addition to the statistical inference itself, we promote the use of two additional diagnoses in
104 describing the relationship between a particular event and climate change. First, the human influ-
105 ence is quantified both in terms of probability and intensity of the event. Although highlighting
106 this duality is not new, using one point of view or the other may have contributed to past controver-
107 sies (Otto et al. 2012; Trenberth et al. 2015), although both quantities can be derived from the same
108 statistical analysis. Second, we describe how the characteristics of the event (frequency, intensity)
109 evolve with time. This allows us to describe not only the human influence to date – the main
110 diagnosis of event attribution studies – but also how a similar event will be affected by climate
111 change in the future (Christidis et al. 2015, took a first step in this direction). This type of outcome
112 is another benefit of using transient simulations, and might be very helpful for communicating the
113 relationship between an event and climate change in a comprehensive way.

114 The main goal of this paper is to describe the entire proposed statistical method, and to provide
115 a first illustration of its potential. The proposed algorithm is flexible and could be improved in
116 several ways, without significantly affecting its general structure. Such improvements could be
117 considered in future work.

118 **2. Framing and data**

119 *a. Event definition and indicators of the human influence*

120 Although a relatively wide biodiversity of events (or classes of events) has been considered in
121 event attribution, in this study we focus on simple events such as

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

122 where y is a univariate random climate variable – typically temperature, rainfall or wind speed,
123 averaged over a given time window and spatial domain – and s a predetermined threshold. We

124 assume that this event has happened at a time t_e in the factual (F) world². The attribution analy-
 125 sis involves describing the characteristics of a *similar* event happening in the counterfactual (C)
 126 world³. As we consider transient simulations where climate changes with time, describing how
 127 the characteristics of the event vary with time, e.g. in the factual world, is also of interest.

128 Changes in occurrence frequency / probability can be assessed by comparing the probability
 129 of the event E happening in (F) vs (C), considering the same threshold s . Denoting $F_{F,t}$ and
 130 $F_{C,t}$ the cumulative distribution functions of y at time t in the factual and counterfactual worlds,
 131 respectively, we define

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

132 Human influence is then typically characterized through the risk-ratio (RR) and the fraction of
 133 attributable risk (FAR, Stott et al. 2004)

$$RR(t) = \frac{p_F(t)}{p_C(t)}, \quad FAR(t) = \frac{p_F(t) - p_C(t)}{p_F(t)} = 1 - \frac{1}{RR(t)}. \quad (3)$$

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

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

$$i_C(t) = F_{F,t}^{-1}(1 - p_F), \quad i_C(t) = F_{C,t}^{-1}(1 - p_F), \quad \text{then} \quad \delta i(t) = i_F(t) - i_C(t). \quad (4)$$

138 In other words, i_F and i_C are the quantiles of order p_F of $F_{F,t}$ and $F_{C,t}$, respectively. The definition
 139 of p_F implies that $i_F(t_e) = s$. δi tells how much more/less intense the event with exactly the
 140 same frequency would have been in the counterfactual world. Note that, according to the climate

²The factual world, or *world as it is*, is the world where all external forcings, including the anthropogenic ones, have influenced climate.

³The counterfactual world, or *world that might have been*, is the world where anthropogenic influence is removed, while natural forcings still vary through time.

141 variable considered, using a relative difference, rather than absolute difference, in δi might be
142 appropriate.

143 Two important remarks can be added. First, conventional attribution studies only calculate RR ,
144 FAR or δi at time t_e , i.e. the exact date at which the event was observed. Calculation of RR , FAR
145 or δi at a different date allows us to quantify the human influence, had the event occurred at that
146 date. Second, describing how the characteristics of the event are changing through time, e.g. in
147 the factual world, is also helpful (see e.g. Christidis et al. 2015). This can be done using relative
148 indices, e.g.

$$RR_{rel}(t) = \frac{p_F(t)}{p_F(t_e)}, \quad \text{or} \quad \delta i_{rel}(t) = i_F(t) - i_F(t_e). \quad (5)$$

149 All these diagnoses are calculated and illustrated subsequently.

150 *b. Case study: 2003 European Heatwave*

151 In order to illustrate the method presented in this paper, we focus on the 2003 European Heat-
152 Wave (EHW03), an event which has long been scrutinized in event attribution studies (Stott et al.
153 2004; Schär et al. 2004; Christidis et al. 2015). We define EHW03 (variable y) as a 1-month
154 event occurring in August 2003 near Paris, France. The spatial domain considered is a $5^\circ \times 5^\circ$
155 square surrounding Paris, i.e. [45N–50N] and [0E–5E]. The choice of this space-time window is
156 debatable (see e.g. Cattiaux and Ribes 2018). A monthly value was considered – a convenient
157 choice in order to involve as many CMIP5 models as possible, and illustrate their (dis-)agreement.
158 The threshold used, s , corresponds to a 5°C anomaly with respect to the 1961-1990 mean. This
159 temperature anomaly was effectively exceeded in 2003 ($+5.38^\circ\text{C}$), but not overtaken in any other
160 year in the instrumental record.

161 As further discussed below, our technique also requires the use of a covariate x , which is as-
162 sumed to be representative of climate change magnitude over time. We consider the summer mean
163 temperature over Western Europe ([35N-70N] and [10W-30E]) in this respect.

164 *c. Data*

165 We use data from a collection of climate models from the 5th Coupled Model Intercomparison
166 Project (CMIP5) – all 24 models considered are listed in Figure 5. For each model, we combine
167 historical simulations (1850–2005) and RCP8.5 simulations (2006–2100). We use all available
168 runs in cases where ensembles have been performed – using a different number of historical and
169 RCP8.5 simulations is not problematic. Pre-industrial control simulations are also used at some
170 point to quantify internal variability and derive confidence intervals.

171 Our method also requires using observed data. We use HadCRUT4 (Morice et al. 2012,
172 <https://crudata.uea.ac.uk/cru/data/temperature/>) to provide historical summer mean
173 temperatures over Western Europe (1850–2016, [10W,30E]x[35N,70N]), and August mean tem-
174 peratures in the vicinity of Paris, France (1850–2016, [0E,5E]x[45N,50N]).

175 **3. Statistical analysis of transient simulations**

176 In this section, we consider data from one single climate model, and describe how changes in risk
177 can be calculated from such data. By construction, transient simulations exhibit a non-stationary
178 climate, so using non-stationary statistics is a key component of our approach. Therefore, we
179 consider a covariate x which is assumed to be representative of climate change magnitude over
180 time. The covariate will typically be a temperature, averaged either globally or over a large region,
181 on a seasonal or annual basis. Several studies already used the global mean temperature as such
182 a covariate (van Oldenborgh et al. 2015; van der Wiel et al. 2017). Here we use summer mean

183 European average temperature, and refer to 6 for further discussion on this choice. Once this
184 covariate x has been selected, our procedure is as follows.

185 *a. ANT and NAT contributions to changes in x*

186 As a first step, we need to estimate the forced responses in the covariate x , and in particular the
187 contributions of natural vs anthropogenic forcings to changes in x . This is typically the purpose
188 of detection and attribution techniques. However, these techniques are not usually designed to
189 provide smooth time-series as a result. We therefore propose a hybrid approach using Generalised
190 Additive Models (GAM).

191 We assume that

$$x_t = \underbrace{\mu_x + \beta e_t}_{x_t^{nat}} + \underbrace{f(t)}_{x_t^{ant}} + \varepsilon_t, \quad (6)$$

$$= x_t^{nat} + x_t^{ant} + \varepsilon_t, \quad (7)$$

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

195 Within this framework, estimation of the response to natural forcing is quite consistent with
196 usual D&A practice, as it involves the estimation of an unknown scaling factor β . The main inno-
197 vation is the consideration of the response e_t derived from an EBM, rather than a more complex
198 model. In doing this, we take advantage of the information provided by forcing time-series, and
199 avoid involving additional noise (i.e. internal variability) from a climate model run. As a result,
200 the estimated response to natural forcings is much more constrained; for instance, the impact of
201 major volcanic eruptions is clearly noticeable. A similar variant was previously used by Huber
202 and Knutti (2012). In practice, we calculate the EBM solution following Geoffroy et al. (2013).

203 Our best estimate is derived using a multi-model mean of EBM parameters. Other combinations
204 of these parameters are used to quantify uncertainty in a resampling approach.

205 Estimation of the response to anthropogenic forcing relies on the assumption that the time re-
206 sponse to anthropogenic forcing is smooth over time. This can be regarded as a sensible assump-
207 tion, as greenhouse gas and aerosols, i.e. the two dominant drivers, vary quite slowly over time.
208 In addition, anthropogenic influence has been shown to be largely dominant on mean temperature
209 changes over recent decades (Bindoff et al. 2013). Anthropogenically induced changes are com-
210 puted with respect to a reference date t_{ref} , implying that $f(t_{ref}) = 0$; we consider $t_{ref} = 1850$,
211 consistent with CMIP protocol, but another reference could be used. As the shape of f is not
212 further constrained, our estimate might be influenced by low-frequency internal variability; it will
213 be necessary to account for this component in the uncertainty analysis.

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

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

222 This decomposition procedure is illustrated in Figure 1 for one particular CMIP5 model
223 (CNRM-CM5). Response to major volcanic eruptions can be easily identified in both the factual
224 world (all forcings combined) and the counter-factual world (natural forcings only). The human
225 influence emerges from noise near 1970 in this model. This is not necessarily contradictory with
226 the fact that human influence is not attributable at that date in the instrumental record – 10 model

227 runs are used, while only one observed realization is available, implying different signal to noise
 228 ratios.

229 This decomposition produces two major outputs: the estimated response to natural forcings only,
 230 x_t^{nat} , corresponding to the expected value of x in the counterfactual world, and $x_t^{all} = x_t^{ant} + x_t^{nat}$,
 231 describing the state of x in the factual world.

232 *b. Fitting a non-stationary distribution to y_t*

233 As a second step, a non-stationary distribution is fitted to the variable of interest y . x_t^{all} is used
 234 as a covariate in this non-stationary fit. Two types of distributions are considered:

- 235 • Gaussian distribution:

$$y_t \sim N(\mu_0 + \mu_1 x_t^{all}, \sigma_0(1 + \sigma_1 x_t^{all})). \quad (8)$$

236 The parameters $\gamma = (\mu_0, \mu_1, \sigma_0, \sigma_1)$ can be estimated via maximum likelihood. However, no
 237 closed formula is available in this case, and an optimization algorithm is needed. We used
 238 the `nlminb` R routine, chosen from other possible options. Confidence regions on γ can be
 239 derived by bootstrapping (x_t^{all}, y_t) and simultaneously considering uncertainty on x_t^{all} , derived
 240 from the previous step.

- 241 • Non-parametric distributions, assuming that the quantile of order α at time t , q_t^α , satisfies:

$$q_t^\alpha = \mu_0^\alpha + \mu_1^\alpha x_t^{all}. \quad (9)$$

242 In this case, the parameters $\gamma = (\mu_0^\alpha, \mu_1^\alpha)$ can be estimated, for a collection of α , using
 243 quantile regression (Koenker and Bassett Jr 1978; Koenker and Hallock 2001). Given typical
 244 sample sizes (a few hundreds to thousands of data) and the computational cost of resampling,
 245 a fast algorithm is needed, and we used the Frisch-Newton approach after preprocessing (see
 246 Portnoy et al. 1997, implemented in R under “`pfn`”). Another potential issue comes from

247 the use of quantile regression for a set of values of α . Quantile regression is done separately
248 for each α , and there is no guarantee that the obtained quantiles are properly sorted over the
249 entire considered range of x_t^{all} . We deal with this issue by re-arranging the obtained quantiles
250 (Dette and Volgushev 2008). Searching for a regularised quantile regression able to cope with
251 this issue, e.g. where μ_0^α and μ_1^α are smooth functions of α , would be very attractive but is
252 beyond the scope of this paper. As in the Gaussian case, uncertainty on $(\mu_0^\alpha, \mu_1^\alpha)$ is assessed
253 through a bootstrap procedure.

254 This list of distributions is obviously not exhaustive, and other families might be used. For in-
255 stance, Generalized Extreme Value (GEV) distributions could be of interest when the focus is on
256 annual maxima. The use of non-parametric distribution still offers a lot of flexibility. As an ex-
257 ample, Generalized Pareto Distribution (GPD) can be adjusted to the tails of such distributions
258 in order to improve estimation of rare values⁴. In the remainder of this paper, we focus on non-
259 stationnary Gaussian distribution only for illustrating our method. Non-parametric distributions
260 were also used with success to analyze transient simulations, and they provided results consistent
261 with the Gaussian case for our case study. However, their use in subsequent steps (such as multi-
262 model synthesis and observational constraints, see Sections 4 and 5) is beyond the scope of this
263 paper.

264 The fit of a non-stationary Gaussian distribution is illustrated in Figure 2. This figure suggests
265 that x_t^{all} is an appropriate covariate for y_t , as the linear relationship is well-supported by the data.
266 More generally, this type of diagnosis can be used to check if the choice of the covariate is appro-

⁴Non-stationary GPD-distribution could be used as such to modelize threshold exceedances. 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 counter-factual world can become quite common in the course of the 21st century, requiring an estimation of the entire distribution in order to derive standard attribution statistics.

267 priate. The fact that the three regression lines (corresponding to mean and quantiles) are almost
268 parallel indicates that there is almost no changes in variance for these particular model and vari-
269 able.

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

273 *c. Estimating changes in probability / intensity*

274 Once a non-stationary distribution has been fitted on y , all attribution diagnoses introduced in
275 Section 2 can be derived easily. In particular, frequency (i.e. probability) and intensity can be
276 calculated in the factual and counterfactual world at time t_e , and RR and δi can be derived from
277 there.

278 Changes in frequency and intensity, as estimated from one particular CMIP5 model, are illus-
279 trated in Figures 3 and 4, respectively. The event frequency moves from about 10^{-4} in 1850 to
280 more than 1/2 in 2100 in the factual world. These numbers differ, but are not inconsistent with
281 Stott et al. (2004), as we consider a smaller space-time domain, implying a much smaller signal-to-
282 noise ratio. Over the same period, the magnitude of the event increases by about 6°C in response
283 to human influence. Frequency and intensity diagnoses complement each other well, and show
284 that recent changes are large in terms of risk ratios (RR near 10 in 2003) while remaining lim-
285 ited in terms of magnitude (near 1°C in 2003) in that model. The influence of natural forcings
286 is clearly discernible, and mainly driven by large volcanic eruptions. Consistent with Figure 1,
287 both frequency and intensity exhibit a discernible human influence as early as 1970 in this model.
288 Human influence becomes huge during the 21st century, with RR higher than 10^4 in 2100. Over-
289 all, confidence intervals might be found to be relatively narrow, but they are consistent with the

290 estimated changes in x (which exhibits limited uncertainty, Figure 1), and the fact that there is a
291 clear relationship between x and y (Figure 2). The latter implies that any significant change in x
292 translates into a significant change in y .

293 4. Multi-model perspective and synthesis

294 a. Results from CMIP5 models

295 In order to give a broader picture, this procedure can then be applied to other CMIP models
296 (Figure 5). This reveals that model uncertainty is large – unlike estimation (or sampling) uncer-
297 tainty which remains very limited. Models’ best estimates of RR vary from 1.8 to more than 400
298 at the date of the event. The lower bound goes down to 1 considering estimation uncertainty (i.e.
299 confidence intervals). Discrepancies among models are also very large in terms of δi , from .2 to
300 3°C in 2003. Similar findings are made on the other parameters involved: p_C, p_F, i_C, i_F – keeping
301 in mind that model biases contribute substantially to discrepancies in i_C, i_F .

302 Unlike CNRM-CM5, some individual models exhibit a significant cooling trend (e.g. FGOALS-
303 g2, ACCESS1-3, all versions of MIROC and CMCC) or warming trend (e.g. BCC-CSM1-1-M,
304 INMCM4, GISS-E2-R) in x_t^{all} prior to 1950 (Figure 6a) – a period over which the anthropogenic
305 forcings are limited. Most of this trend is interpreted as resulting from human influence (i.e., falls
306 into x_t^{ant}) according to the simple decomposition described in Sub-section 3a. Such trends typ-
307 ically result in RR (resp. δi) becoming significantly different from 1 (resp. 0) soon after 1850
308 (Figure 6d,g). At this stage it is unclear whether these trends (i) are related to low frequency
309 internal variability inappropriately taken into account, (ii) can be explained by a long-term re-
310 gional drift in (imbalanced) pre-industrial control simulations, or (iii) highlight an early onset of
311 the anthropogenic influence (i.e. appropriately taken into account), either dominated by aerosols

312 (inducing a cooling) or GHG (inducing a warming) at the regional scale. This will require further
313 investigations which go beyond the methodological scope of this paper.

314 Though large, discrepancies among models, e.g. in terms of RR and δi estimates in 2003, can be
315 easily understood. Models disagree on the magnitude of the changes in the covariate x (different
316 global or local sensitivity), the variance of y (which strongly influences the probability to exceed
317 a high threshold), and the strength of the relationship between y and x . Each model exhibits some
318 bias in one of these characteristics. This highlights the need for a multi-model synthesis.

319 *b. Building a multi-model synthesis*

320 Techniques for building a multi-model synthesis have received much attention in both the litera-
321 ture and IPCC review, due to their importance in providing climate change projections for the next
322 century, including an assessment of uncertainty (Collins et al. 2013). Literature on the subject of
323 how to use an ensemble of opportunity such as the CMIP ensemble, i.e. where no particular design
324 effort is made to cover the range of uncertainty (Tebaldi and Knutti 2007; Knutti et al. 2010a,b),
325 is particularly abundant. These attempts to combine several models into one single uncertainty
326 range have not been translated into event attribution thus far. In this section we introduce one pos-
327 sible method for conducting such a synthesis in the context of the statistical framework described
328 above. The proposed technique is similar to that outlined in Ribes et al. (2017); we review the
329 main concepts here, but refer to that publication for a more detailed discussion.

330 Following Section 3, the parameters describing the response of one single model are: $\theta =$
331 $(x_t^{all}, x_t^{nat}, \gamma)$ – all diagnoses can be derived from θ . The key idea behind the multi-model syn-
332 thesis is to assume that $(\theta_i)_{i=1, \dots, m}$ (where m is the number of models) are realizations of one
333 *multi-model* distribution. Then, it is further assumed that the truth, say θ^* , is also a realization
334 of this multi-model distribution – a paradigm known as *models are statistically indistinguishable*

335 *from the truth* (Annan and Hargreaves 2010). It is necessary to estimate this distribution in order
 336 to derive multi-model statistics such as confidence regions. In the following, this distribution is
 337 assumed to be Gaussian, but the procedure could be extended to other types of distributions.

338 In more detail, we assume that:

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

339 leading to:

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

340 where θ_i is the value of θ for model i , $\hat{\theta}_i$ its estimate, μ and Σ_m are the mean and variance of the
 341 multi-model population (i.e. Σ_m represents modelling uncertainty on θ), and $\Sigma_{\theta,i}$ describes the
 342 uncertainty related to internal variability in the estimation of θ_i . For each model, $\hat{\theta}_i$ can be derived
 343 from the estimation procedure described in Section 3. Estimates of $\Sigma_{\theta,i}$ can also be derived from
 344 the uncertainty analysis conducted there – the resampling was intended to explore uncertainty re-
 345 lated to internal variability. It is important to account for this uncertainty component because, for
 346 some class of extreme events, the signal-to-noise ratio is low, which makes the estimate $\hat{\theta}_i$ rela-
 347 tively inaccurate. In such a case, the estimation uncertainty (i.e. $\Sigma_{\theta,i}$) can substantially contribute
 348 to the spread in the estimated values $\hat{\theta}_i$ (in addition to the spread in θ_i). The next step is to esti-
 349 mate μ, Σ_m from the available sample of θ_i – we refer to (Ribes et al. 2017) for this technical step.
 350 Lastly, confidence regions for the truth θ^* can be derived from μ, Σ_m .

351 Given a collection of CMIP models such as in Figure 5, our procedure can be used to derive
 352 multi-model statistics and confidence regions (Figure 6, and 'MULTI' confidence ranges in Figure
 353 5). The fitted multi-model distribution can also be used to sample new realizations (using Monte-
 354 Carlo simulations) corresponding to virtual climate models (Figure 6) – this is a way to check that
 355 the fitted distribution is consistent with the model sample.

356 Multi-model uncertainty is found to be much larger than the sampling uncertainty related to
357 internal variability in one given model. This is not surprising for a month-long temperature event
358 such as the one investigated here, and is consistent with many other studies (e.g. Hawkins and
359 Sutton 2009). The multi-model confidence range for RR is about [1.4, 230] in 2003, which better
360 reflects the overall uncertainty than single model estimates. It is worth noting that the reported
361 multi-model confidence regions are not equal to the range of single model results. Some models
362 can be excluded from the final confidence region if they are outliers in terms of θ . And, in the
363 presence of a very large sample of models, the bounds of the multi-model confidence region would
364 converge to the corresponding quantiles of the model sample.

365 The uncertainties reported above appear larger than in previous studies. In particular, our confi-
366 dence range for RR is larger than reported by (Stott et al. 2004). Discrepancies in the methods and
367 event definitions can explain or contribute to this gap. Among these, two important features of our
368 approaches should be highlighted. First, the ensemble of models considered here is larger than in
369 any other attribution study, enabling a more comprehensive exploration of uncertainties. Second,
370 the attribution performed here is less constrained than other approaches. The most widespread
371 event attribution procedure relies on prescribed SSTs where an estimate of the anthropogenic in-
372 fluence is removed (Pall et al. 2011). The latter usually involves observations to some extent,
373 leading to a climate change signal in SSTs which is more constrained than that simulated by (un-
374 constrained) coupled models. This highlights the benefit of incorporating observed information in
375 our procedure – a path explored in the next section. However, a large part of the model spread
376 shown in Figure 5 cannot be explained by the use of coupled rather than atmospheric-only mod-
377 els. For instance, the reported spread in p_F is almost entirely related to spread in the variance
378 of y , which also strongly contributes to the spread in RR. Therefore, our results suggest that it is

379 critical to consider a large ensemble of models, with a careful assessment of uncertainty, for event
380 attribution – as for assessing many other features of climate change.

381 **5. Merging models and observations**

382 In this section we introduce two options which can be used to combine observations and infor-
383 mation provided by climate models, using the multi-model synthesis as a starting point. Among
384 other possible approaches, we focus on using observations to constrain changes in x and estimate
385 the distribution of y . Other options are briefly discussed in Section 6.

386 *a. Observed changes in x*

387 Detection and attribution studies have long illustrated that observations might be used to derive
388 information on, e.g., human induced warming to date (Bindoff et al. 2013), in particular in cases
389 where the investigated variable exhibits a high signal to noise ratio (SNR, i.e. response to anthro-
390 pogenic forcings with respect to the magnitude of internal variability). As the selected covariate x
391 typically exhibits high SNR, historical records of x are likely to be insightful with respect to both
392 past and future changes in x . Taking our case study as an example, in Figure 6a-b, models exhibit
393 large differences in the simulated regional warming to date (0.5 to 2°C in 2015). Observations
394 available over the same period of time suggest that the uncertainty range in the past warming is
395 in fact much smaller (Figure 7). It is thus natural to investigate which changes in x are consistent
396 with available observations.

397 In mathematical terms, real observations of covariable x , say \tilde{x} , can be written as

$$\tilde{x}_t = x_t^{all*} + \varepsilon_t, \quad (12)$$

398 where x_t^{all*} is the real world response to external forcings and ε_t is the contribution of internal vari-
399 ability, considered to be random. Using a Bayesian perspective, multi-model uncertainty on x^{all}

400 (derived from the multi-model synthesis, e.g., Figure 6b) can be considered as a prior distribution
 401 for x_t^{all*} , say $\pi(x_t^{all*})$. If both this prior distribution and the distribution of ε_t are known (sensible
 402 assumption), then it is possible to derive the posterior distribution of $x_t^{all*}|\tilde{x}$, using a conventional
 403 Bayesian technique. This derivation is particularly easy under the assumption made in Section 4,
 404 as all distributions involved (i.e. $\pi(x_t^{all*})$ and the distribution of ε) are assumed to be Gaussian.
 405 In fact the same technique can be employed to derive the distribution of $\theta^*|\tilde{x}$ (θ^* is the value of
 406 θ in the real world; θ^* contains x^{all*} but is larger). Then, this distribution (i.e. constrained by
 407 observations) can be used instead of that of θ^* (unconstrained) to derive all results, following the
 408 same procedure as in Section 4.

409 Application of this procedure to summer mean temperature over Europe (i.e. our covariate
 410 x) suggests that some model responses to historical forcings are inconsistent with observations
 411 (Figure 7). This phenomenon can be explained as follows. Intuitively, x_t^{all*} is the expectation of \tilde{x}_t
 412 at time t , so observed values \tilde{x}_t should be distributed around x_t^{all*} . An x_t^{all*} is not quite plausible
 413 if it lies far away from most observed values \tilde{x}_t . In Figure 7, the upper and lower bounds of
 414 the multi-model distribution $\pi(x_t^{all*})$ fall almost outside the set of observations over the beginning
 415 (before 1900) or the end (after 2000) of the observed period, suggesting some inconsistency. Using
 416 observational information therefore leads to a substantial reduction of the multi-model uncertainty
 417 in changes in x . This reduction is particularly clear over the historical period: the multi-model
 418 5–95% confidence range of total 1850–2015 warming is [0.50°C, 2.00°C] without the use of any
 419 observations (i.e. in $\pi(x_t^{all*})$), but shrinks to [0.87°C, 1.41°C] after applying the observational
 420 constraint (i.e. in $x_t^{all*}|\tilde{x}$). But the spread in future warming is also substantially reduced – [3.82°C,
 421 7.69°C] and [4.41°C, 6.88°C] for the corresponding 5–95% confidence ranges before and after
 422 applying the observational constraint, respectively.

423 The impact on event attribution diagnoses is also very clear, with a sharp reduction of uncertain-
424 ties in RR or δi (Figure 9). The lower bound of RR is particularly affected (3.1 after accounting for
425 observations, as opposed to 1.4 without), because some model responses exhibit almost no human-
426 induced warming in 2003, while observations suggest that such weak responses are unlikely. The
427 effect on δi is even bigger: uncertainty reduces from $[+0.1^{\circ}\text{C}, +2.3^{\circ}\text{C}]$ before accounting for ob-
428 servations to $[+0.5^{\circ}\text{C}, +1.5^{\circ}\text{C}]$ after. Overall, these results suggest that taking observational in-
429 formation into account is very helpful, even if done only through the covariate x , i.e. at a large
430 spatio-temporal scale.

431 *b. Observed distribution of y*

432 Another way of merging climate models outputs with real world observations is to estimate the
433 distribution of y , e.g., at the time of the event. Climate models exhibit all sorts of biases (e.g. biases
434 in the mean climate, biases in the variability, other biases affecting the tails of the distribution,
435 etc), which can make the simulated p_F (and, more generally, the entire probability distribution)
436 erroneous (Bellprat and Doblas-Reyes 2016). Figure 5 shows that estimates of p_F vary widely,
437 and are inconsistent among models. In many cases, even a limited observational record can be
438 sufficient to estimate p_F more accurately than using the multimodel synthesis. In practice, most
439 national weather services do rely on observations rather than climate model simulations in order
440 to estimate the return period of a specific event (and return periods in general, see e.g. Tank et al.
441 2009).

442 Here, we illustrate how observations of the investigated variable y , say \tilde{y} , can be used to in-
443 fer p_F and, more generally, the distribution of y at time t_e . We assume that both changes in the
444 covariate x and the non-stationary coefficients γ_1 are known from climate models, with some un-
445 certainty. Observations are therefore only used to estimate γ_0 , taking into account the influence

446 of climate change on the observational record. Note that this treatment is distinct from (and po-
 447 tentially more appropriate than) a common practice in which occurrence probability is estimated
 448 using observations, but ignoring their climate change component. In the proposed procedure, γ_0
 449 is fully determined by observations, i.e. there is no use of models' prior distribution of γ_0 ; in this
 450 respect, the proposed treatment cannot be considered as an observational constraint. But this is still
 451 a combination of models and observations, as models are used to estimate some parameters (x^{all}
 452 and γ_1) while observations are used to estimate others (γ_0).

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

$$\hat{\mu}_0 = \overline{\tilde{y}_t - \mu_1 x_t^{all}}, \quad (13)$$

455 where \bar{z} denotes the average of z , \tilde{y}_t are the available observations of y , and \tilde{y}_t, μ_1 and x_t^{all} are all
 456 known. Then, σ_0 can be estimated by

$$\hat{\sigma}_0 = \text{sd} \left(\frac{\tilde{y}_t - \mu_1 x_t^{all} - \hat{\mu}_0}{1 + \sigma_1 x_t^{all}} \right), \quad (14)$$

457 where sd is the sample standard deviation, denoting again that every term in the right hand side is
 458 known, and that $(\tilde{y}_t - \mu_1 x_t^{all} - \hat{\mu}_0) \sim N(0, \sigma_0(1 + \sigma_1 x_t^{all}))$. Note that these estimators $(\hat{\mu}_0, \hat{\sigma}_0)$ do
 459 not necessarily coincide with the Maximum Likelihood Estimators, but are however quite natural
 460 and attractive for computational reasons. Uncertainty on these parameters can be assessed by
 461 extending the bootstrap procedure to \tilde{y}_t (i.e. resampling observations \tilde{y}_t , as would be done in a
 462 stationary context), and considering simultaneously uncertainty in μ_1, σ_1 and x_t^{all} , as derived from
 463 the multimodel synthesis.

464 Our procedure is illustrated in Figure 8. As the CanESM2 model simulates a larger change $\mu_1 x_t^{all}$
 465 than CNRM-CM5, the 2003 event is relatively less abnormal according to that model, resulting
 466 in a much larger estimate of p_F . Note that changes in variance are small and do not substantially

467 influence the results in this particular example. Model discrepancies in estimating p_A are therefore
468 largely related to spread in the non-stationary term $\mu_1 x_t^{all}$ in this approach.

469 Applying this procedure to all single models and/or to the multi-model synthesis leads to much
470 reduced uncertainties in the estimate of p_F , which might be attractive for several purposes (Figure
471 9). This makes sense, as γ_0 contains the key parameters of the distribution of y , and p_F is just
472 one feature of that distribution. Estimates of p_C are similarly improved. However, the impact on
473 attribution diagnoses, i.e. RR and δi , is very limited. In fact, the proposed procedure refines
474 the estimation of the y -distribution, but does not affect the estimation of human influence, and so
475 coefficients measuring that influence are only marginally impacted.

476 *c. Applying the two constraints together*

477 The two constraints presented above can also be applied simultaneously. If so, observations are
478 used to constrain changes in x first; then parameters γ_0 are estimated using observations of y , given
479 (x, γ_1) . Therefore, observed information is used in both x and γ_0 , in addition to model information.
480 As they combine all sources of information, results obtained in this way can be considered as the
481 final results of the attribution procedure described in this paper (Figures 9 and 10 for the multi-
482 model synthesis only; results obtained applying these constraints with single-model estimates of x
483 and/or γ_1 are shown in Supplementary Material).

484 Applying the two constraints simultaneously leads to a real narrowing of uncertainties in es-
485 timating probabilities p_F or p_C (where estimation of γ_0 is critical), but also the standard human
486 influence diagnoses RR and δi (where constraining x is critical), if compared to the unconstrained
487 multi-model estimates (Figure 9). In terms of attribution diagnoses, uncertainty in RR shrinks
488 from [1.4, 230] (multimode synthesis, no use of observations) to [4,95] (applying the two con-
489 straints). Uncertainty in δi is also strongly reduced, from [+0.1°C,+2.3°C] to [+0.5°C,+1.5°C]

490 (i.e. roughly by a factor of 2). Similar findings are made for p_C and p_F . In all cases considered,
491 applying the two constraints together reduces model spread further than using one single constraint
492 or no observations at all.

493 Remarkably, time series of attribution diagnoses, RR and δi , can still be derived after applying
494 these constraints (Figure 10). Beyond estimates of RR and δi in 2003, several conclusions can
495 be derived from there. First, human influence on an event like EWH03 has been significant since
496 the mid 1980's (Figure 10a,c). Second, the odds of observing an event such as EWV03 (in the
497 sense of the same magnitude) have strongly increased since 2003; they were 3 to 9 times larger in
498 2018 than in 2003 (qualitatively consistent with Christidis et al. 2015). Third, an event similar to
499 EHW03 (in the sense of the same frequency) occurring in 2100 under an RCP8.5 scenario would
500 imply a human contribution as large as $+7.7^\circ\text{C}$ [$+4.7$, $+11.1$] (Figure 10c). Lastly, a very large
501 fraction of this human-induced warming is expected to take place after 2003: $+6.8^\circ\text{C}$ [$+4$, $+9.8$].

502 Overall, these results suggest that our approach, in addition to covering a wide range of uncer-
503 tainties through the use of a large ensemble of models, can lead to relatively constrained attribution
504 results. They also suggest that, in the particular case under consideration, the unconstrained pa-
505 rameters γ_1 do not exhibit a large spread among models.

506 6. Discussion

507 In this section we reivew several aspects of our proposed method which deserve particular at-
508 tention.

509 *(i) Choice of the covariate x* Selecting an appropriate covariate x is a key step in our method. Ob-
510 viously, the choice of this covariate is at least partly subjective, and can impact the final results. In
511 our view, using a global or regional temperature may be appropriate, as changes in many variables
512 have been described to scale with temperature (e.g. Collins et al. 2013; Tebaldi and Arblaster

513 2014). Pattern scaling, however, works better if only the GHG forcing is playing. In practice,
514 other forcings, and anthropogenic aerosols in particular, also contributed to recent changes. As the
515 strength of the aerosols forcing varies considerably over space, using a regional temperature as a
516 covariate might better reflect the regional balance between various external forcings. In any case,
517 relying on a covariate x is a strong assumption of our method, which much be properly acknowl-
518 edged. Replicating the analysis with a different covariate might be one way to explore sensitivity
519 to this choice. Incorporating a covariate uncertainty to the overall algorithm might be attractive as
520 well, but goes beyond the scope of this paper.

521 *(ii) Limitations in using non-stationary statistics* The use of non-stationary statistics is central in
522 this approach, and some limitations must be pointed out. First, a sufficiently large SNR is needed
523 in model data in order to allow fitting of the non-stationary model. The entire procedure can
524 fail if non-stationary coefficients cannot be estimated properly. In this respect, the temperature
525 event considered in this study was an easy one. The method will have to be tested on other
526 events / variables (e.g. precipitation, wind), to determine the extent of its field of application.
527 Second, further statistical developments might improve the fit of the statistical model. In the
528 current analysis, y -data were limited to a specific space-time domain – we ignore any information
529 available outside this domain. Using further spatial (e.g. a broader region than that of the event) or
530 temporal (e.g. modelling the an entire seasonal cycle) information might be particularly attractive,
531 but would involve a sharp increase in the complexity of the statistical model and inference.

532 *(iii) Climate model evaluation and reliability* Using CMIP models, the resolution of which is
533 typically limited, brings into question the model's reliability in simulating events comparable to
534 the one under scrutiny – given that model biases do impact event attribution results (Bellprat and
535 Doblus-Reyes 2016). The model reliability issue has sometimes been tackled through implement-

536 ing model selection (e.g. King et al. 2015) – an approach which could become ineffective for
537 non-temperature small-scale events. A common and more general recommendation is to use high-
538 resolution models (National Academies of Sciences and Medicine 2016, and references therein).
539 However, in both cases, limiting the analysis to a small number of models can have undesirable
540 effects on the uncertainty analysis.

541 In our study, we did not perform any model evaluation nor selection. However, at least two options
542 can be mentioned to cope with this issue. First, our method could be easily applied to Cordex-style
543 experiments, offering the possibility of exploring the modelling uncertainty using higher-resolution
544 models. Second, bias correction techniques could be applied to model outputs before implement-
545 ing the statistical analysis. Some very simple bias correction is in fact already applied in our
546 method – although not explicitly stated – through defining the event as an anomaly with respect
547 to a given reference period. In our view, using more formal bias correction techniques might
548 reconcile event attribution with the use of coarse resolution CMIP-style models.

549 *(iv) Uncertainty quantification and modelling uncertainty* One key outcome of our analysis is
550 that considering modelling uncertainty is critical in event attribution. Uncertainty ranges vary
551 greatly in size if derived using one model only vs a multi-model ensemble, with ranges far too
552 narrow in the former case. The technique used to build the multi-model synthesis is very simple
553 and could be improved in many ways, e.g. by using a link function for some parameters in θ , a
554 non-Gaussian dependence structure, another paradigm than the model truth exchangeability, etc.
555 In the current form, the uncertainty derived from the multi-model synthesis is typically relatively
556 large, but it might still be necessary to check that it is consistent with observations – no such check
557 was implemented here.

558 *(v) Role of observations* In the case of the EHW03, observations are found to substantially re-
559 duce uncertainties in RR and δi estimates (Section 5). However, this case might be quite unusual
560 because (i) a long observational record is available, and (ii) SNR is relatively large. In many other
561 cases, observational constraints will have much less impact on the final results. Attribution re-
562 sults might even be made artificially sensitive to the region where the event occurred through the
563 influence of the length of the historical record.

564 *(vi) Additional observational constraints* Section 5 explored two possible ways to use observa-
565 tions to constrain attribution results. With respect to y , observations are only used to estimate the
566 *stationary* parameters γ_0 . A natural extension of this work would be to constrain both γ_0 and γ_1 ,
567 i.e. to use observational information to constrain the magnitude of changes in y . Among other pos-
568 sible implementations, this could be done using a Bayesian approach, in which the mutli-model
569 uncertainty on γ is used as a prior distribution. Other strategies for blending model information
570 with observations could also be valid.

571 *(vii) Consistency with other approaches* Assessing the consistency of our results with previous
572 studies which also focused on the EWH03 event is not easy, primarily because variations in the
573 event definition can contribute to discrepancies in the results (Cattiaux and Ribes 2018). Roughly
574 speaking, our results in terms of RR lie somewhere between Stott et al. (2004) and Christidis et al.
575 (2015) – two studies which used similar methods but led to quite different results. Results in terms
576 of δi seems consistent with the figures in Christidis et al. (2015), although this diagnosis was not
577 given explicitly. Providing closer comparisons between our approach and other event attribution
578 methods, potentially single model results, will be of primary interest in the future.

579 *(viii) Consistency with IPCC reports* Several similarities between our approach and typical IPCC
580 practice can be noted. Indeed, our results are based on a collection of CMIP models, and the treat-

581 ment of modelling uncertainty is consistent with Collins et al. (2013). Furthermore, uncertainty
582 plumes describing future changes under a given emission scenario, such as shown in Figure 6h,
583 are qualitatively similar to typical IPCC projections. In this respect, this approach could be con-
584 sidered to be an adaptation of IPCC projections for a singular weather or climate event. It is also
585 important to note that the model information used after applying observational constraints, i.e. γ_1 ,
586 is very similar to that used in Collins et al. (2013, see, e.g., their Figure 12.27 for precipitation
587 extremes) to describe long-term changes in extremes.

588 **7. Conclusion**

589 This study describes and illustrates a new statistical method for event attribution which can be
590 decomposed into three steps. First, event attribution diagnoses are derived from transient CMIP-
591 style experiments using non-stationary statistics and an appropriate covariate. Single model results
592 derived from this step typically exhibit large discrepancies. Second, a multi-model synthesis is
593 performed, assuming model / truth exchangeability. Evidence suggests that this synthesis might
594 represent uncertainty better than single model analyses. Third, multi-model information is com-
595 bined with historical observations in order to account for all sources of information available. This
596 blending typically reduces uncertainty in the final attribution diagnoses, while providing a more
597 comprehensive description of the event and human influence on it.

598 This study illustrates that it is possible to perform event attribution using available CMIP-style
599 simulations, although this goal might have been considered challenging at first glance. This is
600 an important result because the use of such experiments offers several advantages. In particular,
601 it offers the possibility of characterizing the human influence on a singular event in a way that
602 is consistent with long-term projections, i.e. using the same data and a similar quantification of
603 uncertainty. The calculation of uncertainty plumes covering both the past and future also provides

604 a new perspective on the human influence on a singular event. And, obviously, re-using available
605 simulations could save much of computation time and efforts.

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

611 *Acknowledgments.* Part of this work was supported by the French Ministère de la Transition
612 Énergétique et Solidaire (through *Extremoscope* and *Convention pour les Services Climatiques*
613 grants), the European Research Area for Climate Services (through the Eupheme project), the
614 French National Research Agency (ANR, through the REMEMBER project, contract ANR-12-
615 SENV-001).

616 The authors are supported by Météo France (AR), CEA (ST) and CNRS (JC).

617 R scripts used to make the analysis described in this paper will be made publicly available before
618 publication of this manuscript.

619 **References**

620 Allen, M., 2003: Liability for climate change. *Nature*, **421 (6926)**, 891.

621 Annan, J., and J. Hargreaves, 2010: Reliability of the cmip3 ensemble. *Geophysical Research*
622 *Letters*, **37 (L02703)**, doi:10.1029/2009GL041994.

623 Bellprat, O., and F. Doblas-Reyes, 2016: Attribution of extreme weather and climate events over-
624 estimated by unreliable climate simulations. *Geophysical Research Letters*, **43 (5)**, 2158–2164.

625 Bindoff, N., and Coauthors, 2013: *Detection and Attribution of Climate Change: from Global*
626 *to Regional*. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working*
627 *Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*
628 [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V.
629 Bex and P.M. Midgley (eds.)], Cambridge University Press, Cambridge, United Kingdom and
630 New York, NY, USA.

631 Cattiaux, J., and A. Ribes, 2018: Defining single extreme weather events in a climate perspective.
632 *BAMS*, doi:10.1175/BAMS-D-17-0281.1, on-line.

633 Cattiaux, J., R. Vautard, C. Cassou, P. Yiou, V. Masson-Delmotte, and F. Codron, 2010: Winter
634 2010 in europe: A cold extreme in a warming climate. *Geophysical Research Letters*, **37** (20).

635 Christidis, N., G. S. Jones, and P. A. Stott, 2015: Dramatically increasing chance of extremely hot
636 summers since the 2003 european heatwave. *Nature Climate Change*, **5** (1), 46.

637 Christidis, N., P. A. Stott, A. A. Scaife, A. Arribas, G. S. Jones, D. Copsey, J. R. Knight, and W. J.
638 Tennant, 2013: A new hadgem3-a-based system for attribution of weather-and climate-related
639 extreme events. *Journal of Climate*, **26** (9), 2756–2783.

640 Ciavarella, A., and Coauthors, 2018: Upgrade of the hadgem3-a based attribution system to high
641 resolution and a new validation framework for probabilistic event attribution. *Weather and Cli-*
642 *mate Extremes*.

643 Collins, M., and Coauthors, 2013: Long-term climate change: Projections, commitments and
644 irreversibility. *Climate change 2013: the physical science basis. Contribution of Working Group*
645 *I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge
646 University Press, 1029–1136.

647 Dette, H., and S. Volgushev, 2008: Non-crossing non-parametric estimates of quantile curves.
648 *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **70 (3)**, 609–627.

649 Dong, B., R. T. Sutton, L. Shaffrey, and N. P. Klingaman, 2017: Attribution of forced decadal
650 climate change in coupled and uncoupled ocean–atmosphere model experiments. *Journal of*
651 *Climate*, **30 (16)**, 6203–6223.

652 Eden, J. M., K. Wolter, F. E. Otto, and G. J. van Oldenborgh, 2016: Multi-method attribution
653 analysis of extreme precipitation in boulder, colorado. *Environmental Research Letters*, **11 (12)**,
654 124 009.

655 Geoffroy, O., D. Saint-Martin, D. J. Olivié, A. Voldoire, G. Bellon, and S. Tytéca, 2013: Transient
656 climate response in a two-layer energy-balance model. part i: Analytical solution and parameter
657 calibration using cmip5 aogcm experiments. *Journal of Climate*, **26 (6)**, 1841–1857.

658 Hauser, M., and Coauthors, 2017: Methods and model dependency of extreme event attribution:
659 the 2015 european drought. *Earth’s Future*, **5 (10)**, 1034–1043.

660 Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predic-
661 tions. *Bulletin of the American Meteorological Society*, **90 (8)**, 1095–1107.

662 Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis, 2010: Probing
663 the fast and slow components of global warming by returning abruptly to preindustrial forcing.
664 *Journal of Climate*, **23 (9)**, 2418–2427.

665 Huber, M., and R. Knutti, 2012: Anthropogenic and natural warming inferred from changes in
666 earth’s energy balance. *Nature Geoscience*, **5 (1)**, 31–36.

667 King, A. D., G. J. van Oldenborgh, D. J. Karoly, S. C. Lewis, and H. Cullen, 2015: Attribution of
668 the record high central england temperature of 2014 to anthropogenic influences. *Environmental*
669 *Research Letters*, **10** (5), 054 002.

670 Knutti, R., G. Abramowitz, M. Collins, V. Eyring, P. J. Gleckler, B. Hewitson, and L. Mearns,
671 2010a: *Good Practice Guidance Paper on Assessing and Combining Multi Model Climate Pro-*
672 *jections*. In: *Meeting Report of the Intergovernmental Panel on Climate Change Expert Meet-*
673 *ing on Assessing and Combining Multi Model Climate Projections* [Stocker, T.F., D. Qin, G.-K.
674 Plattner, M. Tignor, and P.M. Midgley (eds.)], IPCC Working Group I Technical Support Unit,
675 University of Bern, Bern, Switzerland.

676 Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl, 2010b: Challenges in combining
677 projections from multiple climate models. *Journal of Climate*, **23** (10), 2739–2758.

678 Koenker, R., and G. Bassett Jr, 1978: Regression quantiles. *Econometrica: journal of the Econo-*
679 *metric Society*, 33–50.

680 Koenker, R., and K. F. Hallock, 2001: Quantile regression. *Journal of economic perspectives*,
681 **15** (4), 143–156.

682 Lewis, S. C., and D. J. Karoly, 2013: Anthropogenic contributions to australia’s record summer
683 temperatures of 2013. *Geophysical Research Letters*, **40** (14), 3705–3709.

684 Massey, N., and Coauthors, 2015: weather@ homedevelopment and validation of a very large
685 ensemble modelling system for probabilistic event attribution. *Quarterly Journal of the Royal*
686 *Meteorological Society*, **141** (690), 1528–1545.

687 Morice, C., J. Kennedy, N. Rayner, and P. Jones, 2012: Quantifying uncertainties in global and
688 regional temperature change using an ensemble of observational estimates: The hadcrut4 data
689 set. *Journal of Geophysical Research*, **117 (D8)**, doi:10.1029/2011JD017187.

690 National Academies of Sciences, E., and Medicine, 2016: *Attribution of extreme weather events*
691 *in the context of climate change*. National Academies Press.

692 Otto, F. E., N. Massey, G. Van Oldenborgh, R. Jones, and M. Allen, 2012: Reconciling two
693 approaches to attribution of the 2010 russian heat wave. *Geophysical Research Letters*, **39 (4)**.

694 Paciorek, C. J., D. A. Stone, and M. F. Wehner, 2018: Quantifying statistical uncertainty in the
695 attribution of human influence on severe weather. *Weather and Climate Extremes*, **20**, 69–80.

696 Pall, P., T. Aina, D. A. Stone, P. A. Stott, T. Nozawa, A. G. Hilberts, D. Lohmann, and M. R.
697 Allen, 2011: Anthropogenic greenhouse gas contribution to flood risk in england and wales in
698 autumn 2000. *Nature*, **470 (7334)**, 382.

699 Peterson, T. C., P. A. Stott, and S. Herring, 2012: Explaining extreme events of 2011 from a
700 climate perspective. *Bulletin of the American Meteorological Society*, **93 (7)**, 1041–1067.

701 Portnoy, S., R. Koenker, and Coauthors, 1997: The gaussian hare and the laplacian tortoise: com-
702 putability of squared-error versus absolute-error estimators. *Statistical Science*, **12 (4)**, 279–300.

703 Ribes, A., F. W. Zwiers, J.-M. Azaïs, and P. Naveau, 2017: A new statistical approach
704 to climate change detection and attribution. *Climate Dynamics*, **48**, 367–386, doi:10.1007/
705 s00382-016-3079-6.

706 Risser, M. D., D. A. Stone, C. J. Paciorek, M. F. Wehner, and O. Angéllil, 2017: Quantifying the
707 effect of interannual ocean variability on the attribution of extreme climate events to human
708 influence. *Climate Dynamics*, **49 (9-10)**, 3051–3073.

709 Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller, 2004: The
710 role of increasing temperature variability in european summer heatwaves. *Nature*, **427 (6972)**,
711 332.

712 Stott, P., D. Stone, and M. Allen, 2004: Human contribution to the european heatwave of 2003.
713 *Nature*, **432**, 610, doi:10.1038/nature03089.

714 Tank, A. M. K., F. W. Zwiers, and X. Zhang, 2009: *Guidelines on analysis of extremes in a*
715 *changing climate in support of informed decisions for adaptation*. WMO.

716 Tebaldi, C., and J. Arblaster, 2014: Pattern scaling: Its strengths and limitations, and an update on
717 the latest model simulations. *Climatic Change*, **122**, 459–471, doi:10.1007/s10584-013-1032-9.

718 Tebaldi, C., and R. Knutti, 2007: The use of the multi-model ensemble in probabilistic climate pro-
719 jections. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical*
720 *and Engineering Sciences*, **365 (1857)**, 2053–2075.

721 Trenberth, K. E., J. T. Fasullo, and T. G. Shepherd, 2015: Attribution of climate extreme events.
722 *Nature Climate Change*, **5 (8)**, 725.

723 Uhe, P., F. Otto, K. Haustein, G. van Oldenborgh, A. King, D. Wallom, M. Allen, and H. Cullen,
724 2016: Comparison of methods: Attributing the 2014 record european temperatures to human
725 influences. *Geophysical Research Letters*, **43 (16)**, 8685–8693.

726 van der Wiel, K., and Coauthors, 2017: Rapid attribution of the August 2016 flood-inducing
727 extreme precipitation in south Louisiana to climate change. *Hydrol. Earth Syst. Sci.*, **21 (2)**,
728 897–921, doi:10.5194/hess-21-897-2017.

729 van Oldenborgh, G. J., F. E. Otto, K. Haustein, and H. Cullen, 2015: Climate change increases the
730 probability of heavy rains like those of storm Desmond in the UK an event attribution study in
731 near-real time. *Hydrology & Earth System Sciences Discussions*, **12** (12).

732 Vautard, R., and Coauthors, 2015: Extreme fall 2014 precipitation in the Cévennes mountains.
733 *Bulletin of the American Meteorological Society*, **96** (12), S56–S60.

734 Wehner, M., D. Stone, H. Shiogama, P. Wolski, A. Ciavarella, N. Christidis, and H. Krishnan,
735 2018: Early 21st century anthropogenic changes in extremely hot days as simulated by the
736 c20c+ detection and attribution multi-model ensemble. *Weather and Climate Extremes*, **20**, 1–8.

737 **LIST OF FIGURES**

738 **Fig. 1. ANT and NAT influences in covariate x .** Changes in covariate x (raw data are shown as
739 black points) are decomposed into changes related to all forcings combined (top, red), an-
740 thropogenic forcings only (middle, green), and natural forcings only (bottom, blue). Shaded
741 area corresponds to 5–95% confidence regions. All data are from CNRM-CM5, including
742 all runs available from this model (10 historical, 5 RCP8.5). 38

743 **Fig. 2. Fit of a non-stationary Gaussian distribution.** Illustration of the non-stationary fit
744 for data from the CNRM-CM5 model, assuming Gaussian distribution. Black points: data
745 (x_t^{all}, y_t) , where x_t^{all} has been estimated following Section a, and y_t is the raw y data. Several
746 simulations are considered, leading to several values of y_t at each time t . Many points lie in
747 the bottom-left corner, corresponding to the quasi-stationary climate of the period before
748 1980. Red lines: change in the mean of y_t (solid) or 5% - 95% quantiles (dashed lines) as
749 estimated from the non-stationary fit. 39

750 **Fig. 3. Changes in Frequency.** Changes in the frequency of the event, as diagnosed from the
751 analysis of transient historical and RCP8.5 scenarios, in terms of occurrence probability in
752 the factual world (p_F , top panel), occurrence probability in the counter-factual world (p_C ,
753 middle top), RR/FAR ($RR = \frac{p_F}{p_C}$, middle bottom), and RR relative to year 2003 ($RR_{rel} =$
754 $\frac{p_F(t)}{p_F(t_e)}$), i.e. the year on which the event occurred (vertical bar). Shaded areas correspond to
755 5%–95% confidence regions. All results are for the CNRM-CM5 model only. 40

756 **Fig. 4. Changes in Intensity.** Changes in the intensity of the event, as diagnosed from the analysis
757 of transient historical and RCP8.5 scenarios, in terms of intensity in the factual world (i_F , top
758 panel), intensity in the counter-factual world (i_C , middle top), difference between these two
759 ($\delta i = i_F(t) - i_C(t)$, middle bottom), or difference in the factual world with respect to
760 2003 ($\delta i_{rel} = i(t) - i(t_e)$, bottom), i.e. the year on which the event occurred (vertical bar).
761 Shaded areas correspond to 5%–95% confidence regions. All results are for the CNRM-
762 CM5 model only. 41

763 **Fig. 5. Diagnoses in year 2003.** 42

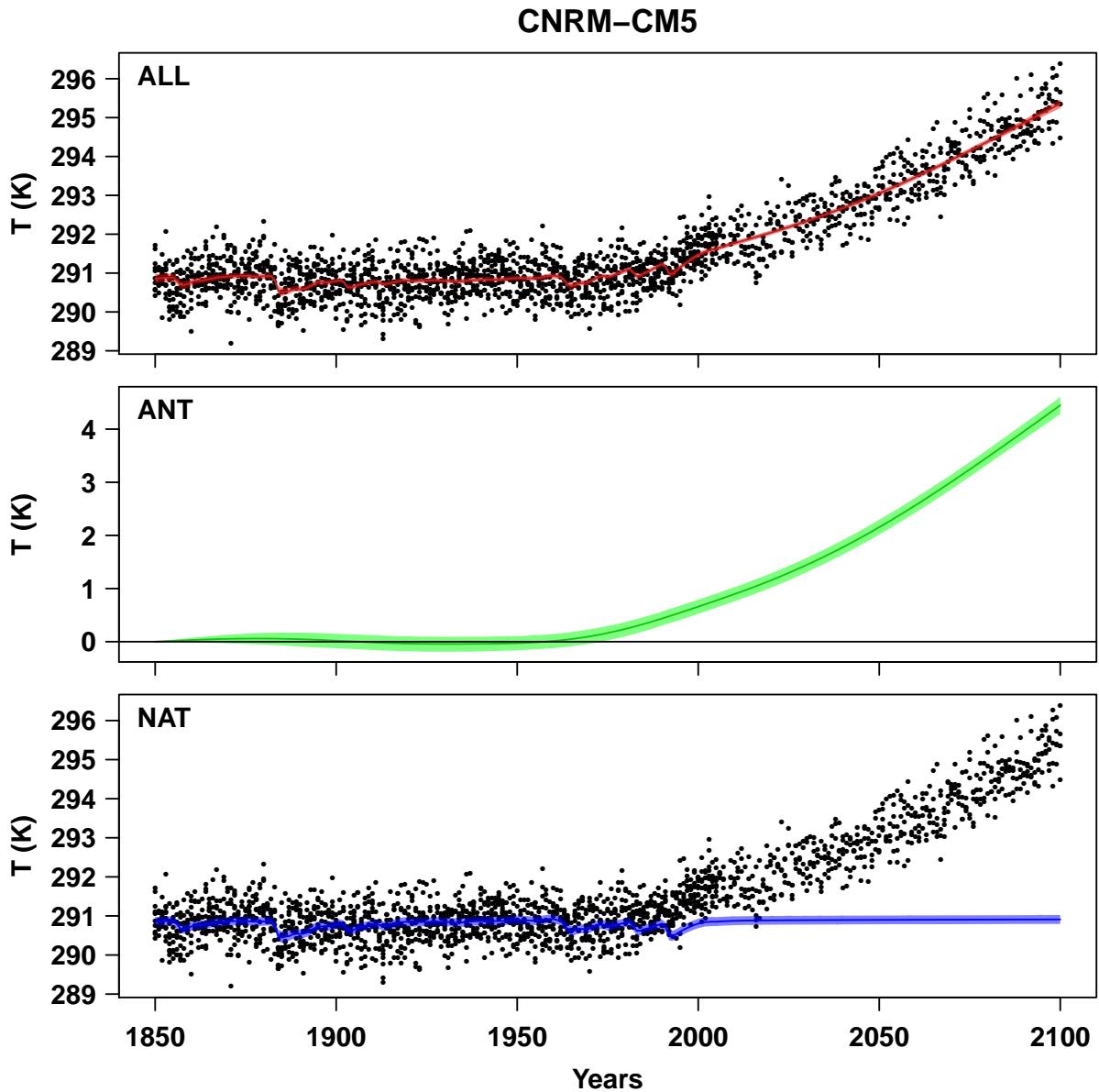
764 **Fig. 6. Multi-model statistics and synthesis.** Results for the 23 individual CMIP5 models consid-
765 ered (left) are shown in terms of changes in covariate x (degrees with respect to the reference
766 year 2003, top), risk ratio (RR, middle), and δi (bottom). The multimodel distribu-
767 tion estimated from this sample of models is illustrated through: the mean and 90% confidence
768 region (i.e. confidence region of θ^* , or resulting diagnoses, using the “models are statisti-
769 cally indistinguishable from the truth” paradigm; center), or new realisations drawn from
770 that distribution (which can be interpreted as virtual climate models; right). 43

771 **Fig. 7. Observational constraint on covariate x .** Observation of covariate x (here European sum-
772 mer mean temperature since 1850, black points), are compared to the multi-model distribu-
773 tion of externally forced changes in x , i.e. $\pi(x_t^{all*})$ (light pink, 5–95% confidence region,
774 identical to Figure 6b). Uncertainty in changes in x decreases after applying the observa-
775 tional constraint, i.e. $x_t^{all*} | \bar{x}$ (dark pink, 5–95% confidence region). Best estimates before
776 (light brown) and after (brown) applying the observational constraint are almost indistin-
777 guishable in this case, as observations are consistent with the multi-model mean estimate.
778 All values are temperature anomalies with respect to the 1961-1990 period. 44

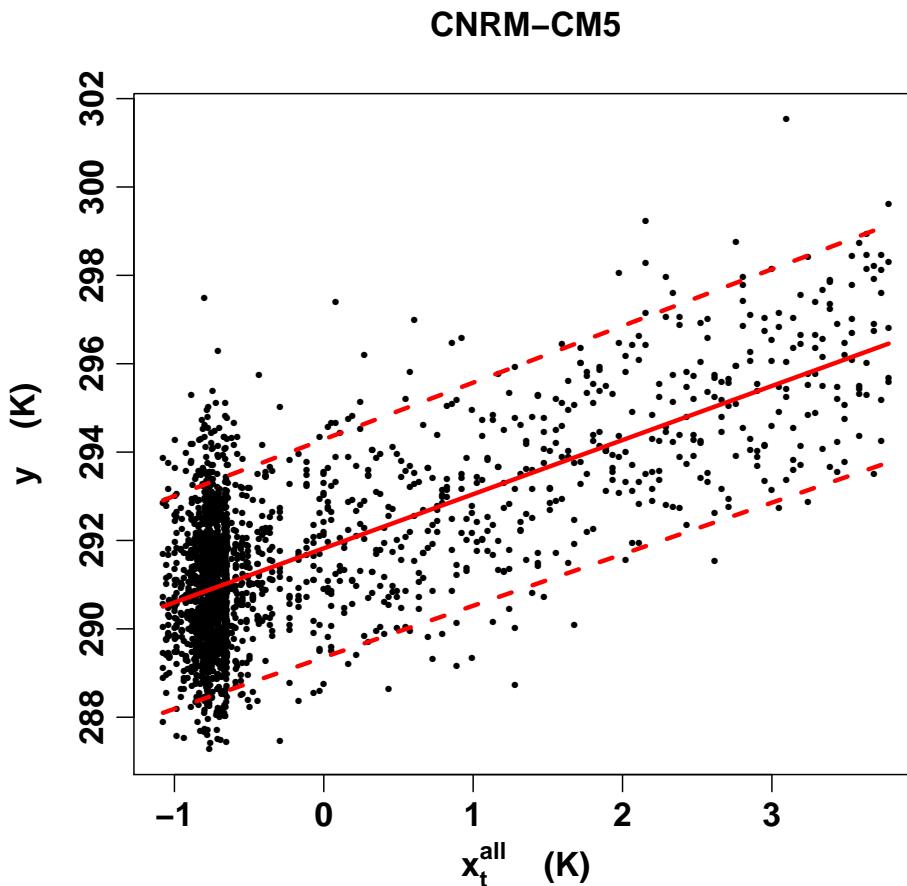
779 **Fig. 8. Use of corrected observations to estimate p_F .** Top panel: observed record \widetilde{y}_t (black line)
 780 is compared to the changes in y (i.e. $\mu_1 x_t^{all}$) as simulated by two climate models (CNRM-
 781 CM5 and CanESM2, color lines). Temperatures are anomalies with respect to the 1961-
 782 1990 average. Other panels: Observed time-series after correction for the mean change
 783 simulated by the models, i.e. $\widetilde{y}_t - \mu_1 x_t$. Correction is made such that the 2003 value is
 784 unchanged. p_F denotes the probability of exceeding the threshold (dotted line), as estimated
 785 from these corrected records. Uncertainty analysis and correction for changes in variance
 786 are not represented. 45

787 **Fig. 9. Impact of (combined) constraints on p_C (a), p_F (b), RR/FAR (c) and δi (d).** All results
 788 correspond to the multi-model synthesis (Section 4). NO OBS: only model outputs (no
 789 observations) are used; results are identical to those presented in Figure 5 for the MULTI
 790 synthesis. C0: γ_0 is estimated from observations \widetilde{y} (see subsection 5.b). CX: Estimates
 791 of changes in x are constrained by observation \widetilde{x} (see subsection 5.a). CX+C0: the two
 792 constraints are applied simultaneously. 46

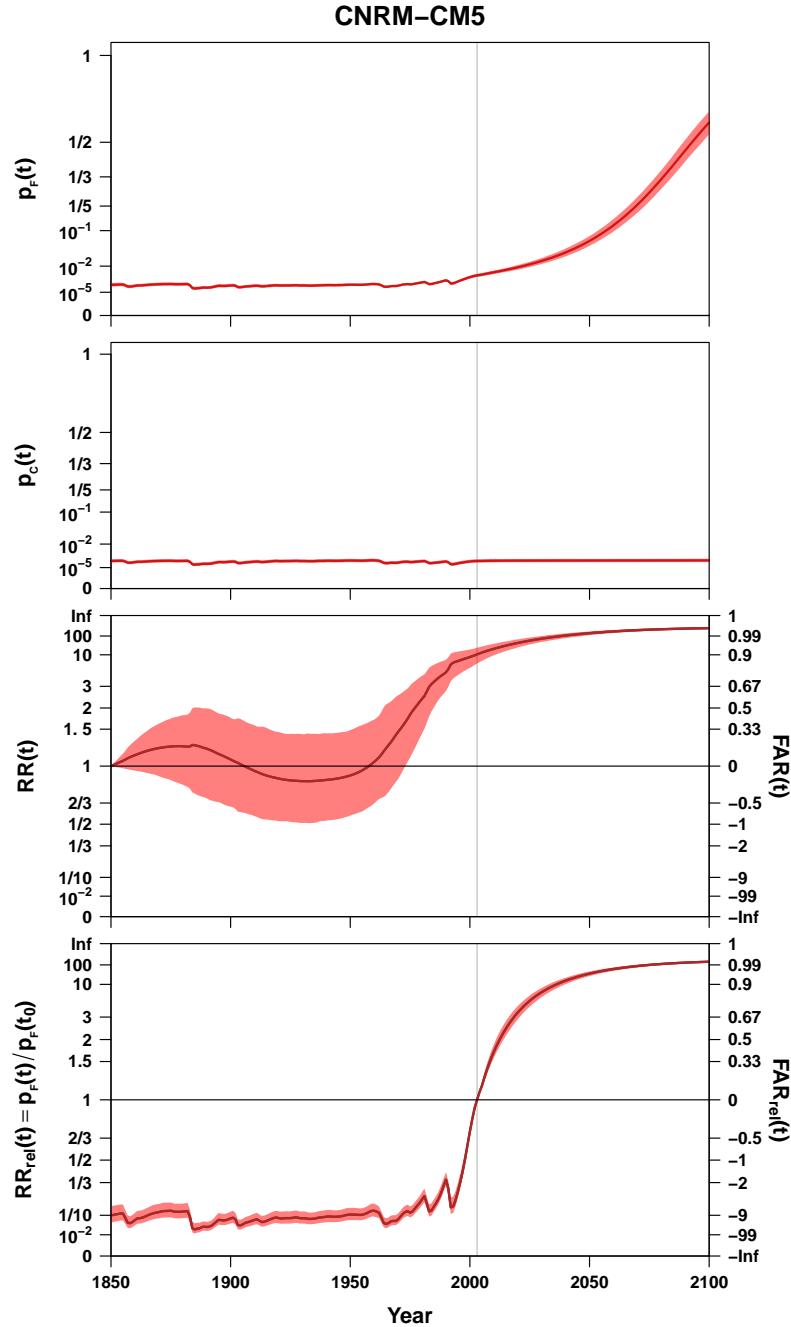
793 **Fig. 10. Attribution diagnoses as estimated after applying observational constraints CX and**
 794 **C0.** Changes in frequency (RR and RR_{rel} , left) and intensity (δi and δi_{rel} , right) are used to
 795 quantify the human influence on an EHW03-like event. All results are shown as a function
 796 of time, had the event occurred at that time. The vertical bar indicates the year 2003. Results
 797 are for the multimodel synthesis only, after applying the two observational constraints CX
 798 and C0 simultaneously. 47



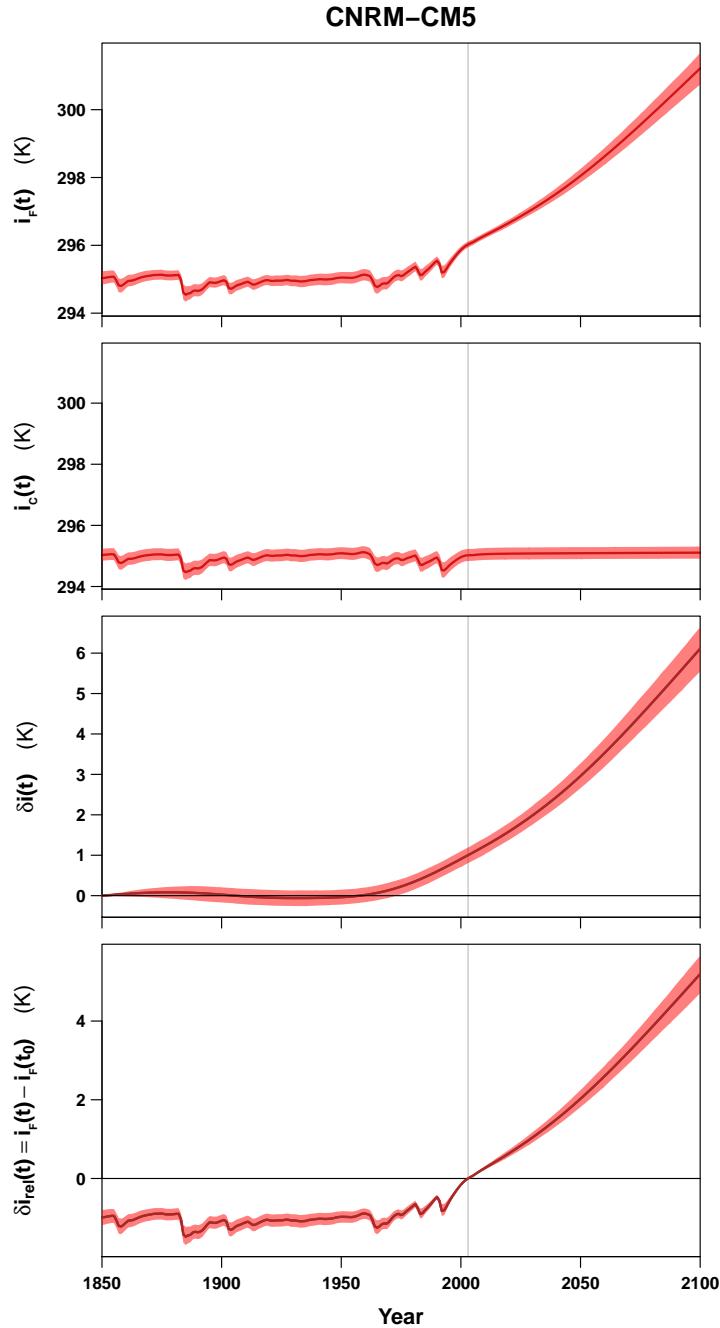
799 FIG. 1. **ANT and NAT influences in covariate x .** Changes in covariate x (raw data are shown as black points)
 800 are decomposed into changes related to all forcings combined (top, red), anthropogenic forcings only (middle,
 801 green), and natural forcings only (bottom, blue). Shaded area corresponds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



803 **FIG. 2. Fit of a non-stationary Gaussian distribution.** Illustration of the non-stationary fit for data
 804 from the CNRM-CM5 model, assuming Gaussian distribution. Black points: data (x_t^{all}, y_t) , where x_t^{all} has been
 805 estimated following Section a, and y_t is the raw y data. Several simulations are considered, leading to several
 806 values of y_t at each time t . Many points lie in the bottom-left corner, corresponding to the quasi-stationary
 807 climate of the period before 1980. Red lines: change in the mean of y_t (solid) or 5% - 95% quantiles (dashed
 808 lines) as estimated from the non-stationary fit.



809 FIG. 3. **Changes in Frequency.** Changes in the frequency of the event, as diagnosed from the analysis of
 810 transient historical and RCP8.5 scenarios, in terms of occurrence probability in the factual world (p_F , top panel),
 811 occurrence probability in the counter-factual world (p_C , middle top), RR/FAR ($RR = \frac{p_F}{p_C}$, middle bottom), and
 812 RR relative to year 2003 ($RR_{rel} = \frac{p_F(t)}{p_F(t_e)}$), i.e. the year on which the event occurred (vertical bar). Shaded areas
 813 correspond to 5%–95% confidence regions. All results are for the CNRM-CM5 model only.



814 **FIG. 4. Changes in Intensity.** Changes in the intensity of the event, as diagnosed from the analysis of
 815 transient historical and RCP8.5 scenarios, in terms of intensity in the factual world (i_f , top panel), intensity in
 816 the counter-factual world (i_c , middle top), difference between these two ($\delta i = i_f(t) - i_c(t)$, middle bottom), or
 817 difference in the factual world with respect to year 2003 ($\delta i_{rel} = i(t) - i(t_e)$, bottom), i.e. the year on which the
 818 event occurred (vertical bar). Shaded areas correspond to 5%–95% confidence regions. All results are for the
 819 CNRM-CM5 model only.

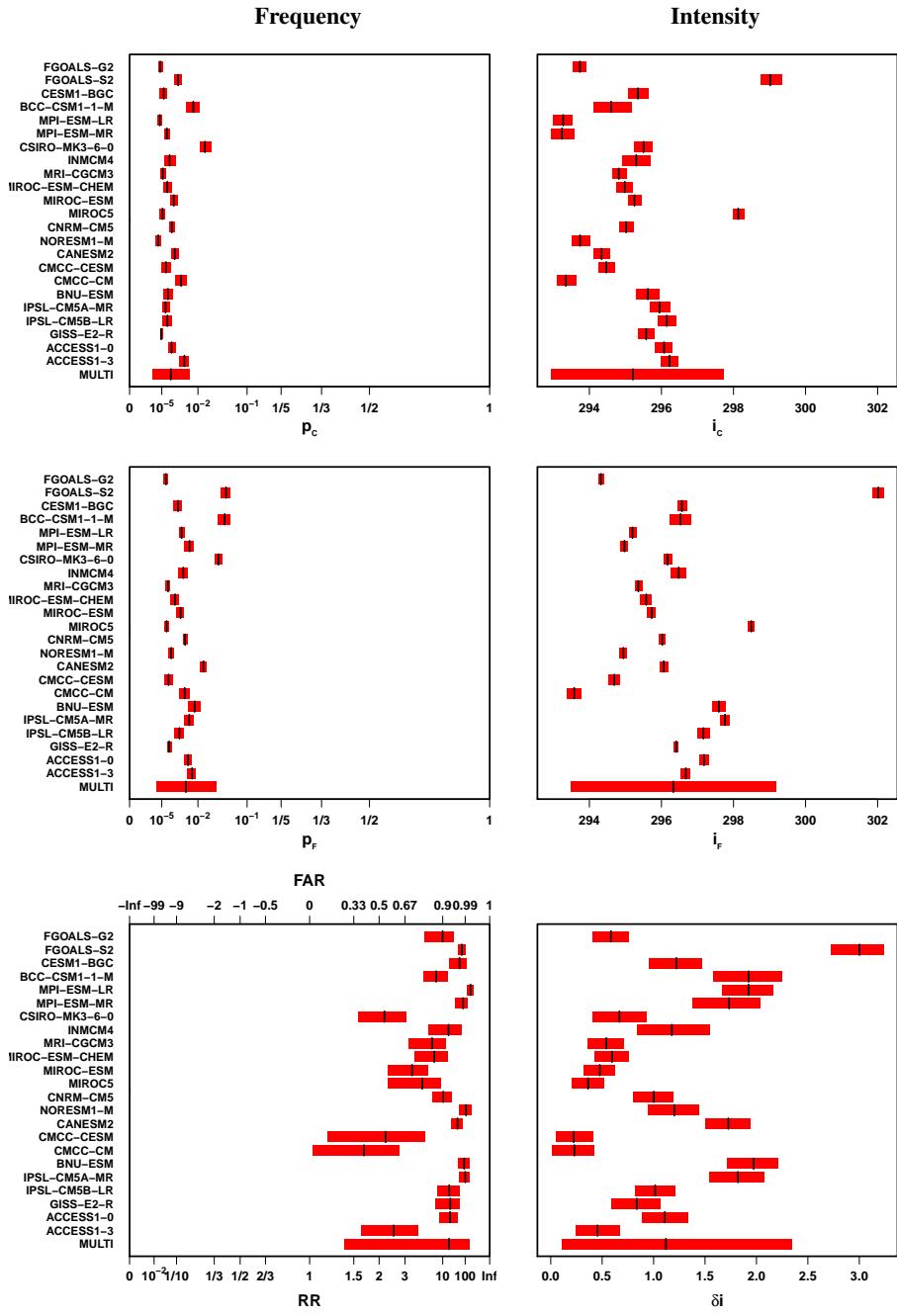
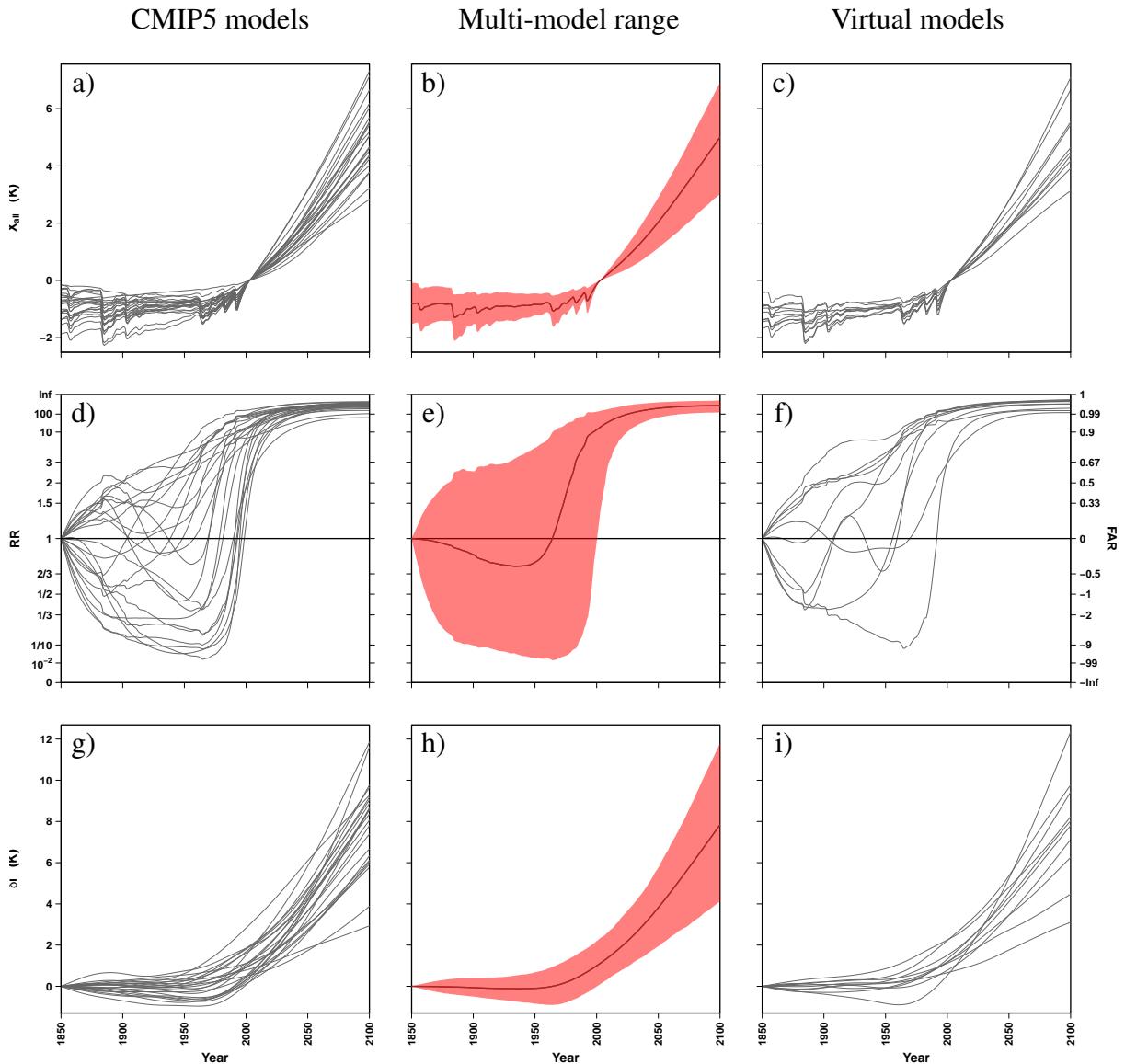
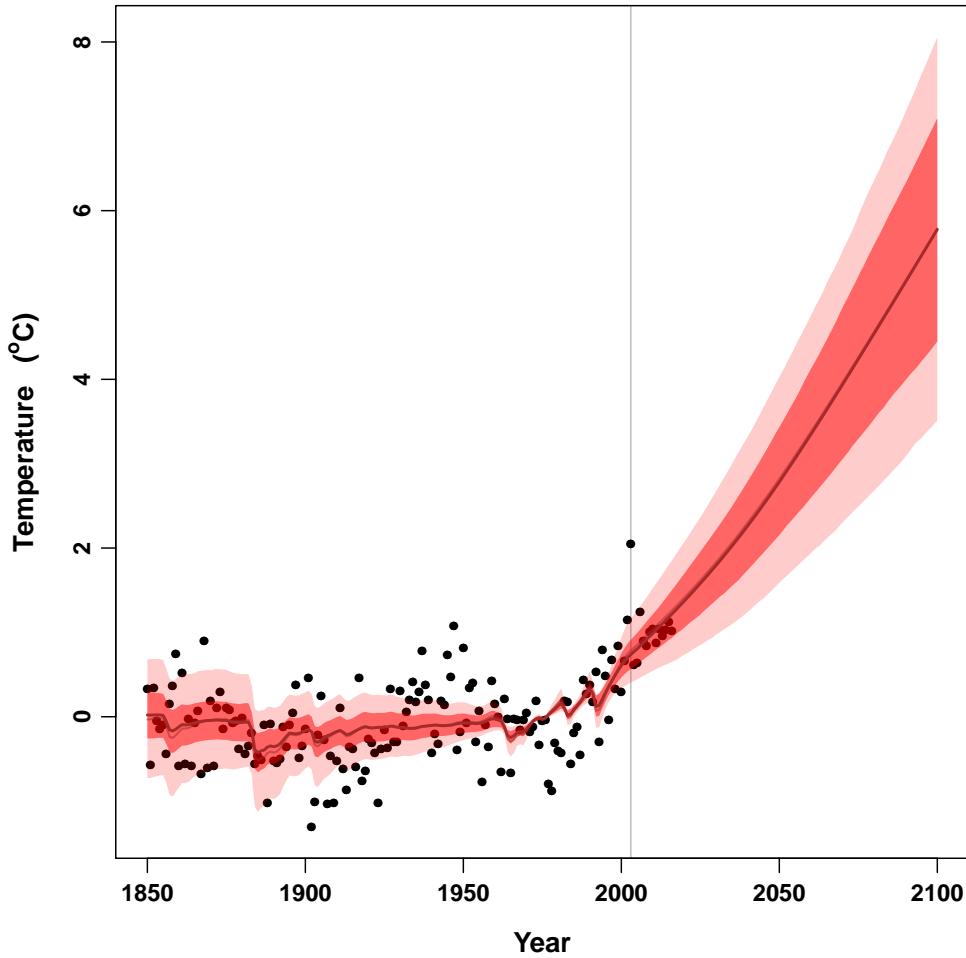


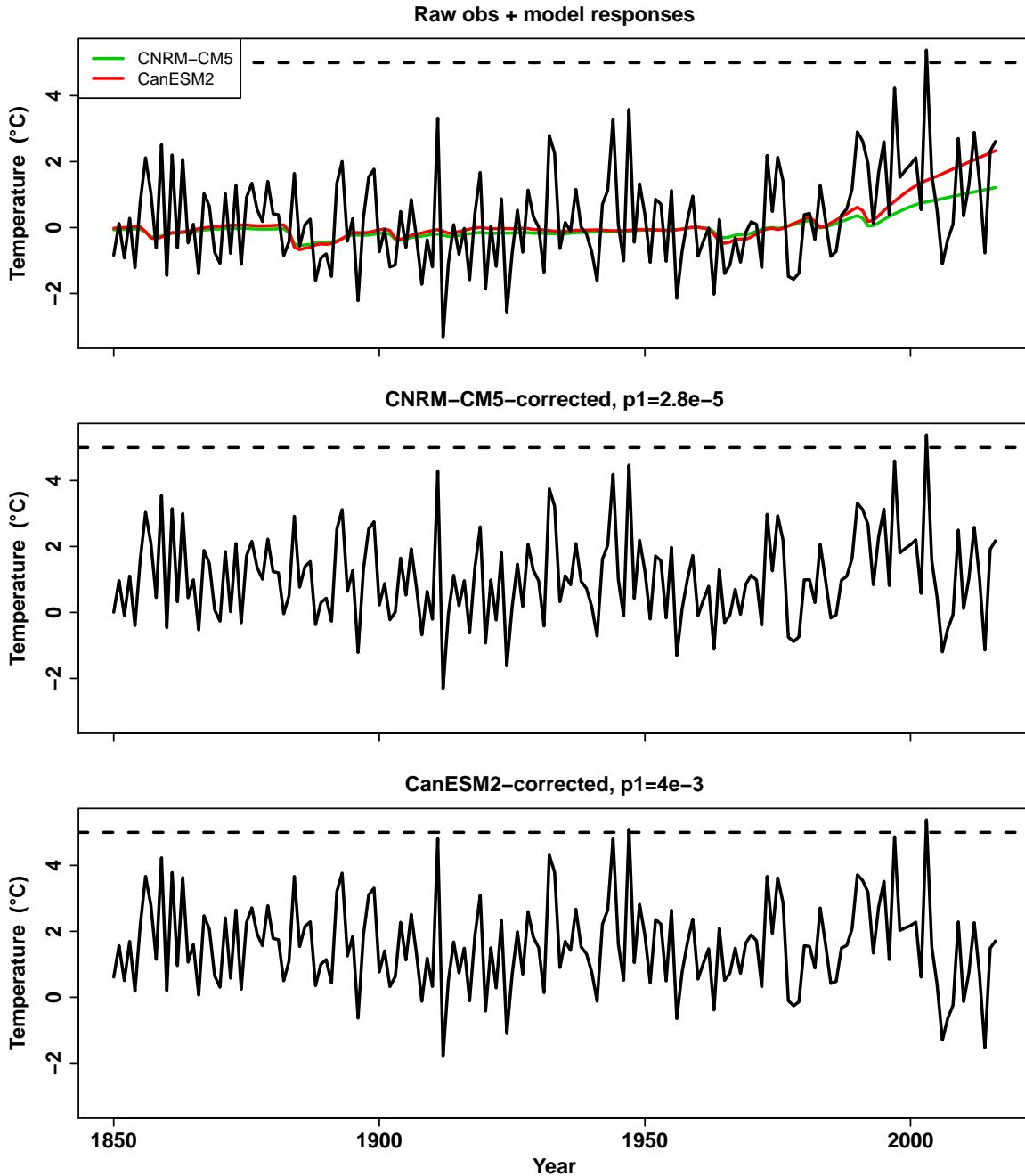
FIG. 5. Diagnoses in year 2003.



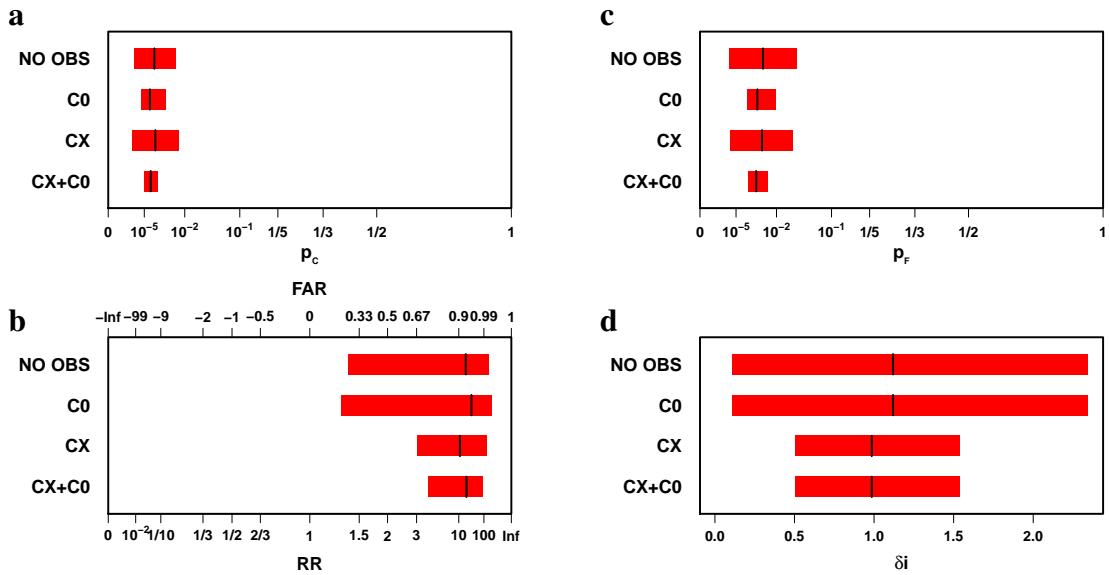
820 **FIG. 6. Multi-model statistics and synthesis.** Results for the 23 individual CMIP5 models considered (left)
 821 are shown in terms of changes in covariate x (degrees with respect to the reference year 2003, top), risk ratio
 822 (RR, middle), and δi (bottom). The multimodel distribution estimated from this sample of models is illustrated
 823 through: the mean and 90% confidence region (i.e. confidence region of θ^* , or resulting diagnoses, using the
 824 “models are statistically indistinguishable from the truth” paradigm; center), or new realisations drawn from that
 825 distribution (which can be interpreted as virtual climate models; right).



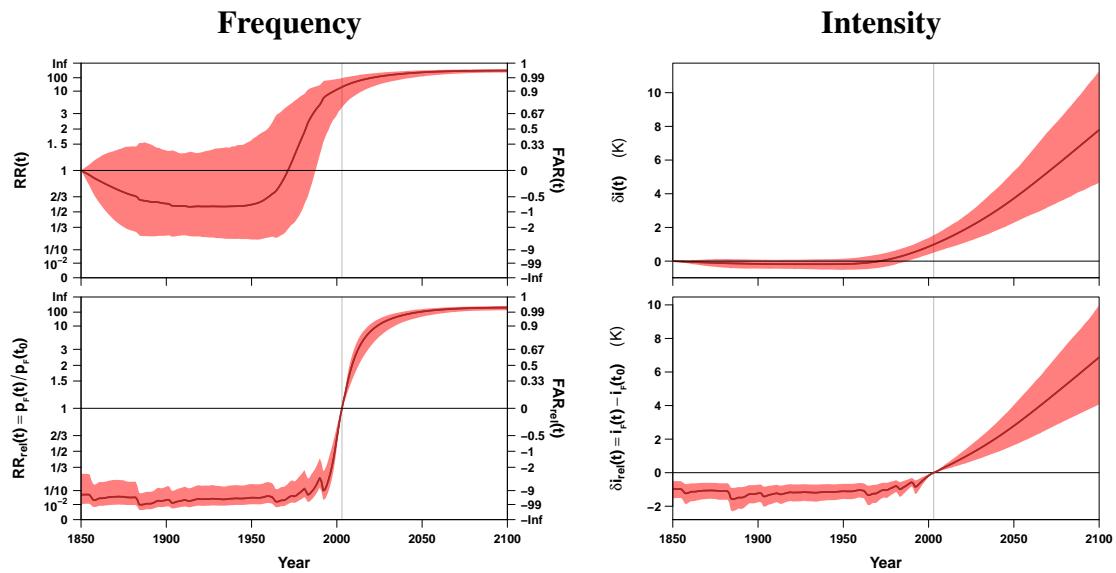
826 FIG. 7. **Observational constraint on covariate x .** Observation of covariate x (here European summer mean
 827 temperature since 1850, black points), are compared to the multi-model distribution of externally forced changes
 828 in x , i.e. $\pi(x_t^{all*})$ (light pink, 5–95% confidence region, identical to Figure 6b). Uncertainty in changes in x de-
 829 creases after applying the observational constraint, i.e. $x_t^{all*} | \tilde{x}$ (dark pink, 5–95% confidence region). Best
 830 estimates before (light brown) and after (brown) applying the observational constraint are almost indistinguish-
 831 able in this case, as observations are consistent with the multi-model mean estimate. All values are temperature
 832 anomalies with respect to the 1961-1990 period.



833 FIG. 8. Use of corrected observations to estimate p_F . Top panel: observed record \tilde{y}_t (black line) is compared
 834 to the changes in y (i.e. $\mu_1 x_t^{all}$) as simulated by two climate models (CNRM-CM5 and CanESM2, color lines).
 835 Temperatures are anomalies with respect to the 1961-1990 average. Other panels: Observed time-series after
 836 correction for the mean change simulated by the models, i.e. $\tilde{y}_t - \mu_1 x_t$. Correction is made such that the 2003
 837 value is unchanged. p_F denotes the probability of exceeding the threshold (dotted line), as estimated from these
 838 corrected records. Uncertainty analysis and correction for changes in variance are not represented.



839 FIG. 9. **Impact of (combined) constraints on p_C (a), p_F (b), RR/FAR (c) and δ_i (d).** All results correspond
 840 to the multi-model synthesis (Section 4). NO OBS: only model outputs (no observations) are used; results are
 841 identical to those presented in Figure 5 for the MULTI synthesis. C0: γ_0 is estimated from observations \tilde{y} (see
 842 subsection 5.b). CX: Estimates of changes in x are constrained by observation \tilde{x} (see subsection 5.a). CX+C0:
 843 the two constraints are applied simultaneously.



844 **FIG. 10. Attribution diagnoses as estimated after applying observational constraints CX and C0.**
 845 Changes in frequency (RR and RR_{rel} , left) and intensity (δi and δi_{rel} , right) are used to quantify the human
 846 influence on an EHW03-like event. All results are shown as a function of time, had the event occurred at that
 847 time. The vertical bar indicates the year 2003. Results are for the multimodel synthesis only, after applying the
 848 two observational constraints CX and C0 simultaneously.