CrossMark

# Is the choice of statistical paradigm critical in extreme event attribution studies?

Peter A. Stott[1,2] · David J. Karoly[3,4] · Francis W. Zwiers[5]

**Abstract** The science of event attribution meets a mounting demand for reliable and timely information about the links between climate change and individual extreme events. Studies have estimated the contribution of human-induced climate change to the magnitude of an event as well as its likelihood, and many types of event have been investigated including heatwaves, floods, and droughts. Despite this progress, such approaches have been criticised for being unreliable and for being overly conservative. We argue that such criticisms are misplaced. Rather, a false dichotomy has arisen between "conventional" approaches and new alternative framings. We have three points to make about the choice of statistical paradigm for event attribution studies. First, different approaches to event attribution may choose to occupy different places on the conditioning spectrum. Providing this choice of conditioning is communicated clearly, the value of such choices depends ultimately on their utility to the user concerned. Second, event attribution is an estimation problem for which either frequentist or Bayesian paradigms can be used. Third, for hypothesis testing, the choice of null hypothesis is context specific. Thus, the null hypothesis of human influence is not inherently a preferable alternative to the usual null hypothesis of no human influence.

This comment refers to the article available at doi:10.1007/s10584-017-2048-3

✉ Peter A. Stott
    peter.stott@metoffice.gov.uk

1    College of Engineering, Mathematics and Physical Sciences, University of Exeter, North Park Road, Exeter EX4 4QF, UK

2    Met Office Hadley Centre, Fitzroy Road, Exeter EX1 3PB, UK

3    Environmental Change Institute, Oxford Martin School, Oxford, UK

4    School of Earth Sciences and ARC Centre of Excellence for Climate System Science, University of Melbourne, Melbourne 3010, Australia

5    Pacific Climate Impacts Consortium, University of Victoria, University House, 1700, Stn CSC, Victoria, British Columbia V8W 2Y2, Canada

## 1 Introduction

Allen (2003) first proposed the concept of attributing individual extreme weather events, using a paradigm in which attribution is based on estimates of the extent to which human-induced climate change has increased or decreased the likelihood of an event such as a heatwave or flood. The first application of the concept was an attribution of causes of the extreme European summer temperatures of 2003 (Stott et al. 2004). This study showed that it was very likely (greater than 90% chance) that human-induced climate change had more than doubled the likelihood of occurrence of such extreme temperatures.

Research on event attribution has subsequently flourished, responding to a mounting demand for reliable and timely information about the links between climate change and individual extreme events. Although we will focus on the likelihood of events in this article, many studies have also estimated the contribution of human-induced climate change to the magnitude of an event, as well as to its likelihood. In many cases, the magnitude of the event is likely to be more strongly affected by natural variability than its likelihood. For example, the likelihood of the Russian heatwave of 2010 was found to have increased substantially with global warming (Rahmstorf and Coumou 2011) while its magnitude was found to be mainly natural in origin (Dole et al. 2011). These different results reflect the different ways of framing the attribution question (Otto et al. 2012) and illustrate that it is worth considering both likelihood and magnitude in event attribution.

Many types of events in many regions have now been investigated. Human-induced climate change is estimated to have increased the likelihood of the Australian record summer temperature of 2013 (Lewis and Karoly 2013), to have contributed to flood risk that led to the devastating inundations in England and Wales in Autumn 2000 (Pall et al. 2011), and to have increased the chances of the record low Arctic sea ice extent seen in 2012 (Kirchmeier-Young et al. 2017). Annual reports in the *Bulletin of the American Meteorological Society* assess the extent to which anthropogenic climate change has affected the strength and likelihood of recent individual extreme events. Human-induced climate change has increased the intensity or likelihood of almost all the heat-related events examined, and has affected many other events including tropical cyclones, forest fires, and cold events.

As a result of such work, the value of event attribution is now becoming clear. Stakeholders from different sectors (insurance, policy making, media, legal) have a variety of different uses for such information (Stott et al. 2016). The US National Academy of Sciences recently assessed the validity of event attribution (NAS 2016) and concluded that "it is now often possible to make and defend quantitative statements about the extent to which human-induced climate change (or another causal factor, such as a specific mode of natural variability) has influenced either the magnitude or the probability of occurrence of specific types of event or event classes." The report noted that event attribution science "has advanced a great deal in recent years and is still evolving rapidly." This was due to two main reasons: "one, the understanding of the climate and weather mechanisms that produce extreme events is improving, and two, rapid progress is being made in the methods that are used for event attribution."

Despite this progress, the approaches described above have been criticised for being flawed in two basic respects (Trenberth et al. 2015; Mann et al. 2017). First, it is argued that the reliance of event attribution studies on climate models makes standard approaches unreliable (Trenberth et al. 2015). Instead, it is argued, attribution studies should consider how anthropogenic climate change has altered the effects of extreme weather events, while considering the meteorological structure of the event, such as the state of the atmospheric circulation, as

being fixed. For example, global warming may have increased the quantity of snow falling in the "snowmaggeden" storm in the Eastern United States in February, 2010 (Trenberth et al. 2015). In such a *storyline* approach (Shepherd 2016), rather than trying to estimate how the overall likelihood of such a snowstorm may have been altered by climate change, it is suggested that the state of the atmospheric circulation that produced the storm should be considered as a given and that the impact of climate change on the amount of snow delivered by the storm should be estimated. While understanding how the overall risks of extreme events are changing requires both a thermodynamic perspective and an understanding of changes in atmospheric circulation (Otto et al. 2016), the counter argument is that this alternative storyline approach may be less prone to error because it does not depend on a climate model's ability to simulate variability and change in atmospheric circulation (Trenberth et al. 2015; Shepherd 2016).

The second flaw, it is argued, is that standard event attribution approaches require significance testing in which the null hypothesis of no human influence must first be ruled out at a sufficiently high significance level (typically 5%) (Trenberth et al. 2015; Mann et al. 2017). Mann et al. 2017 (see also Trenberth et al. 2015) suggest that such a testing-based approach is "conservative" and argue that an approach in which event likelihoods are estimated via Bayesian methods would be better both empirically and ethically.

In this article, we argue that a false dichotomy has arisen between "conventional" approaches and new alternative approaches. All attribution studies depend on certain conditions remaining equal between the factual and counter-factual worlds being compared. The storyline approach is a special case of such conditioning. We do not accept that Bayesian approaches to event attribution are inherently more accurate than frequentist approaches. There is no evidence to suggest that "conventional" approaches are inherently biased either towards anthropogenic influence or against.

## 2 Conditioning in event attribution

The original formulation of Allen (2003) is to compare the probability of an event today (P1) with the probability of the event in the counterfactual conditions (P0) under which a particular driver of climate, such as human influence, is missing. This allows calculation of the fraction of attributable risk (1−P0/P1 = (P1−P0)/P1) or risk ratio (P1/P0). A quadrupling of the risk ratio equates to a FAR equal to 0.75.

Calculating P0 and P1 requires estimating the probability of an event in two different conditions, all other things being equal. For example, in many attribution studies, the same solar and volcanic forcings on climate are included in both factual (P1) and counterfactual (P0) conditions to ensure that differences between P1 and P0 reflect only the additional impact of human influence. A climate model is required for both P0, because these are counterfactual conditions that have not existed in reality, and for P1, because we live in a non-stationary climate that makes estimates inappropriate when based on observational information under the assumption of stationarity. While statistical models based on observations are sometimes used, studies predominantly estimate these probabilities from dynamical climate models. Climate model evaluation is critical in both cases since if a model fails to capture features salient to calculating these two probabilities, its estimate of attributable risk will be in error.

The challenge of model evaluation has been taken up by the event attribution community since an understanding of the fidelity of models is an important component in evaluating and

communicating confidence in event attribution results (Stott et al. 2016; Bellprat and Doblas-Reyes 2016; Lott and Stott 2016). Climate models are not perfect but the relevant question for any application is whether they are fit for purpose. To take an example, a recent evaluation of the HadGEM3-A model being used for event attribution has shown the model has a good representation of circulation characteristics relevant to European extreme events such as heatwaves and droughts when run at seasonal forecast resolution (Vautard et al. 2017). Such evaluation of model performance helps assess confidence in a model-based assessment of changing risk (Stott et al. 2016).

All event attribution studies calculate the effects of a causal factor or factors while controlling for the effects of other factors. The analysis of European seasonal temperatures in 2003 by Stott et al. (2004) compared simulations of European temperatures with and without anthropogenic climate change while assuming that other factors, such as natural forcings, were the same. More stringent conditions can be imposed. For example, uncoupled atmosphere only climate models can be used to evaluate the odds of a particular event with and without anthropogenic forcing under the pattern of sea surface temperatures (SST) observed at the time of the event (e.g. Pall et al. 2011). These odds therefore become SST-pattern dependent. This allows, for example, an assessment of how anthropogenic influence changed the probability of Australian extreme rainfall in 2010/2011 under the La Niña conditions that prevailed at the time (Christidis et al. 2013; King et al. 2013). Even more stringent conditions can be imposed by evaluating the change in probability of an event given specific aspects of the circulation. This allows, for example, an assessment of how anthropogenic influence changed the probability of a Western European cold surge during north-easterly flow in 2009/2010 (Cattiaux et al. 2010).

An event attribution assessment becomes closer to that of the storyline approach (Trenberth et al. 2015; Shepherd 2016) as the degree of conditioning becomes increasingly stringent. When we talk of an event in the attribution context, we are actually referring to a class of events that satisfy a specific criterion, such as the exceedance of a particular temperature or precipitation threshold. This is also true of the storyline approach even if in this case the class of event is much more closely tied to the observed evolution of the event, such as its synoptic evolution. Risk-based approaches to event attribution and the storyline approach differ primarily in occupying different places on this conditioning spectrum. They both depend on climate models of some sort to estimate the appropriate counterfactual since the counterfactual is fundamentally unobservable. Potentially, they both have value. Ultimately, such a value depends on the utility of the information derived for the users interested in exploiting it.

# 3 Frequentist versus Bayesian approaches to event attribution

Event attribution is fundamentally an estimation problem in which the probabilities of an event in the factual and counterfactual worlds (P1, P0, respectively) are calculated from models, either statistical or climatological, under the two alternative situations. Frequentist and Bayesian approaches are different ways of tackling this estimation problem.

Frequentists estimate probabilities conventionally, often by maximizing the likelihood function or seeking estimates with other desirable properties such as unbiasedness or minimum variance under repeated sampling (e.g. Cox and Hinkley 1974). Bayesians proceed similarly, except they calculate a posterior likelihood function, which is a convolution between the likelihood function and a prior distribution based on Bayes theorem that summarizes

information about P0 and P1 that might have been available prior to conducting the analysis. A philosophical distinction is that probability distributions are used to describe uncertainty in P0 and P1, whereas frequentists consider only the variation in the estimates of P0 and P1 that would have occurred under repeated sampling of the observed system.

A potential criticism of Bayesian approaches is that the choice of prior distribution can strongly influence the posterior distribution, which creates a source of uncertainty that may be difficult to evaluate (e.g. Gelman 2008). This could lead to a large bias in the estimation of the human influence on some extreme events if the posterior distribution is unduly influenced by a strong prior belief in a human-induced effect, or the lack of a human-induced effect. Overestimation could lead to poor adaptation decisions by, for example, investing in infrastructure to protect against an increased frequency of events that are not in reality being made more likely by human-induced climate change. Conversely, underestimation could lead to a failure to make needed investments.

## 4 The role of hypothesis testing

It is often supposed (e.g. Trenberth et al. 2015; Mann et al. 2017) that frequentist approaches to event attribution require the null hypothesis of no human influence to be tested before an attribution result is reported. Rather, as pointed out above, event attribution is primarily an estimation problem—that of estimating the probabilities P1 and P0. In making this supposition, a criticism of frequentist approaches to event attribution is that they do not exploit prior information that there is high confidence in a dominant human-induced component to global warming. Additionally, the suggestion is made that this leads to a conservative bias, which is claimed to be favoured by scientists who are keen to protect their reputations and fearful of their results being discredited.

Mann et al. (2017) introduce a medical analogy in their criticism of frequentist approaches, arguing that climate scientists should seek to disprove the null hypothesis that human-induced climate change has made a particular extreme event worse or more likely, just as pharmaceutical companies have to disprove the null hypothesis that drugs do harm. But any such choice needs to be made with a clear-sighted appreciation of its impact on the relevant benefit/loss function. This will be different for different users of the information.

In medicine, the interests of the patient may be different from that of the medical practitioner, or society as a whole. For climate change, an assumption that anthropogenic climate change influenced a particular event in the absence of any further evidence may be appropriate for some international policy makers (where there is strong evidence that such extreme events globally are being affected) but not for regional planners for whom the local specifics of the event are relevant to whether adaptation measures are appropriate.

Ultimately, the question of ethics needs to be addressed by considering the differing interests of different users of event attribution findings and the nature of the decision-making framework that will use event attribution results. If climate change adaptation is still largely a matter of individual actors making local and regional decisions in the face of scarce adaptation resources, then they might expect to be asked to demonstrate that P1 is significantly greater than P0. On the other hand, if it has been decided at a global (or national) political level that the greater good dictates that adaptation will be undertaken, then we might expect to see provisions that will allow local or regional authorities to opt out under some circumstances. One of those circumstances might involve having to demonstrate that P1 is not significantly greater than P0.

Regarding ethics, whether these questions are addressed with Bayesian or frequentist methods is a secondary consideration (either approach can be used for both null hypotheses). The choice of approach should focus primarily on the method that is most appropriate for the inference problem at hand. In instances where the prior is not controversial, a Bayesian method may be preferable from both an estimation and testing perspective. But in other instances where the prior is highly contentious, a Bayesian approach may have little relevance except in those cases where the available evidence overwhelms the choice of prior.

An important point to consider in event attribution is the potentially limited relevance of prior information about the causes of global climate change to the regional event attribution problem. While it is generally accepted that a warmer atmosphere will lead to higher atmospheric moisture content and heavier extreme precipitation globally, there are a number of locations where a prior belief that this expectation applies locally could lead to an incorrect conclusion about anthropogenic influence on climate events at regional scales. Some specific examples include projected declines in extreme rainfall in the parts of the sub-tropics associated with tropical circulation changes (Kharin et al. 2013; Pfahl et al. 2017), observed winter rainfall in the southwest of Australia, which is already showing significant declines due to climate change (Delworth and Zeng 2014), record low October rainfall in October 2015 in Tasmania, Australia (Karoly et al. 2016), and recent unexpected increases in frost risk in some parts of southern Australia (Dittus et al. 2014).

In such cases, such prior expectations might lead to an inappropriate rejection of the alternative null hypothesis proposed by Mann et al. (2017), namely that there is an anthropogenic influence on the event in question. For example, southwest Australian winter rainfall has decreased rather than increased as prior expectations based only on global thermodynamic considerations would suggest. Thus, replacing the null hypothesis of no human influence with its opposite, as suggested by Mann et al. (2017), would not necessarily improve the reliability of hypothesis testing.

The question of ethics and its relation to the question about how to formulate the null hypothesis for testing is not fundamentally a question of a choice between Bayesian and frequentist approaches. Instead, whether posed in a Bayesian or frequentist manner, we return to the point that event attribution problem is an estimation problem. Given that changes locally can be very different to global expectations, as a result for example of dynamically induced changes over-coming thermodynamically induced ones, great care must be taken in using prior expectations derived from global considerations. In some cases, the inappropriate use of such prior information could reach too liberal conclusions. In other cases, the neglect of relevant prior information could lead to overly conservative conclusions.

# 5 Conclusion

In summary, we have three points to make about the choice of statistical paradigm for event attribution studies. First, different approaches to event attribution may choose to occupy different places on the conditioning spectrum. Providing this choice of conditioning is communicated clearly, the value of such choices depends ultimately on their utility to the user concerned. Second, event attribution is an estimation problem for which either frequentist or Bayesian paradigms can be used. Third, for hypothesis testing, the choice of null hypothesis is context specific. Thus, the null hypothesis of human influence is not inherently a preferable alternative to the usual null hypothesis of no human influence.

Finally, we make a remark about ethical practice as it relates to event attribution. Ethical practice should include such considerations as being clear about methods and assumptions (including priors), rigorously assessing tools and uncertainties, and being clear on which hypotheses are being tested and why a particular testing formulation is suitable for the circumstances being considered. This view of what constitutes ethical practice for a practitioner should not be controversial and should be kept distinct from considerations of what constitutes ethical practice for policy makers, business leaders, and politicians.

# References

Allen M.R.(2003) Liability for climate change. **421**, 891–892.

Bellprat O, Doblas-Reyes F (2016) Attribution of extreme weather and climate events overestimated by unreliable climate simulations. Geophys Res Lett 43:2158–2164

Cattiaux J, Vautard R, Cassou C, Yiou P, Masson-Delmotte V, Codron F (2010) Winter 2010 in Europe: a cold extreme in a warming climate. Geophys Res Lett. doi:https://doi.org/10.1029/2010GL044613

Christidis, N., P. A. Stott, D. J. Karoly and A Ciavarella (2013) An attribution study of the heavy rainfall over eastern Australia in March 2012. In explaining extreme events of 2012 from a climate perspective", Bull Amer Meteor Soc, 94 (9), S58-S61.

Cox DR, Hinkley DV (1974) Theoretical statistics. Chapman and Hall, London

Delworth TL, Zeng F (2014) Regional rainfall decline in Australia attributed to anthropogenic greenhouse gases and ozone levels. Nat Geosci 7:583–587

Dittus AJ, Karoly DJ, Lewis SC, Alexander LV (2014) An investigation of some unexpected frost day increases in southern Australia. Aust Met Ocean J 64:261–271

Dole R et al (2011) Was there a basis for anticipating the 2010 Russian heatwave? Geophys Res Lett 38:L06702. doi:https://doi.org/10.1029/2010GL046582

Gelman A (2008) Objections to Bayesian statistics. Bayesian Anal 3:445–450

Karoly DJ, Black MT, Grose MR, King AD (2016) The roles of climate change and El Niño in the record low rainfall in October 2015 in Tasmania, Australia [in "explaining extremes of 2015 from a climate perspective"]. Bull Am Met Soc 97:S127–S130. doi:https://doi.org/10.1175/BAMS-D-16-0139.1

Kharin VV, Zwiers FW, Zhang X, Wehner M (2013) Changes in temperature and precipitation extremes in the CMIP5 ensemble. Clim Chang 119:345–357

King AD, Lewis SC, Perkins SE, Alexander LV, Donat MG, Karoly DJ, Black MT (2013) Limited evidence of anthropogenic influence on the 2011–12 extreme rainfall over Southeast Australia. In "explaining extreme events of 2012 from a climate perspective". Bull. Amer. Meteor. Soc. 94(9):S55–S58

Kirchmeier-Young MC, Zwiers FW, Gillett NP (2017) Attribution of extreme events in Arctic sea ice extent. J Clim 30:553–571

Lewis SC, Karoly DJ (2013) Anthropogenic contributions to Australia's record summer temperature of 2013. Geophys Res Lett 40:3705–3709. doi:https://doi.org/10.1002/grl.50673

Lott FC, Stott PA (2016) Evaluating simulated fraction of attributable risk using climate observations. J Clim. doi:https://doi.org/10.1175/JCLI-D-15-0566.1

Mann M.E., E. A. Lloyd, N. Oreskes (2017) Assessing climate change impacts on extreme weather events: the case for an alternative (Bayesian) approach. In Press.

National Academies of Sciences, Engineering and Medicine (2016) Attribution of extreme weather events in the context of climate change. The National Academies Press, Washington, DC. doi:10.17226/21852

Otto FEL, Massey N, van Oldenborgh GJ, Jones RG, Allen MR (2012) https://doi.org/10.1029/2011GL050422. Geophys Res Lett 39(L04702)

Otto FEL, van Oldenborgh GJ, Eden J, Stott PA, Karoly DJ, Allen MR (2016) The attribution question. Nature Clim Ch 6:813–816

Pall P, Aina T, Stone DA, Stott PA, Nozawa T, Hilberts AGJ, Lohmann D, Allen MR (2011) Nature 470:382–385

Pfahl S, O'Gorman PA, Fischer EM (2017) Understanding the regional pattern of projected future changes in extreme precipitation. Nat Clim Chang 7(6):423–427

Rahmstorf S, Coumou D (2011) Increase of extreme events in a warming world. PNAS 44:17905–17909

Shepherd TG (2016) A common framework for approaches to extreme event attribution. Curr Clim Change Rep 2:28–38

Stott PA, Stone DA, Allen MR (2004) Human contribution to the European heatwave of 2003. Nature 432:610–614

Stott PA, Christidis N, Otto FEL, Sun Y, Vanderlinden J-P, van Oldenborgh GJ, Vautard R, von Storch H, Walton P, Yiou P, Zwiers FW (2016) Attribution of extreme weather and climate-related events. WIRES Climate Change 7:23–41

Trenberth KE, Fasullo JT, Shepherd TG (2015) Attribution of climate extreme events. Nat Clim Chang 5:725–730

Vautard, R., N. Christidis, A. Ciavarella, C. Alvarez-Castro, O. Bellprat, B. Christiansen, I. Colfescu, T. Cowan, F. Doblas-Reyes, J. Eden, M. Hauser, G. Hegerl, N. Hemplemann, K. Klehmet, F. Lott, C. Nangini, R. Orth, S. Radanovics, S. Seneviratne, G. J. van Oldenborgh, P. Stott, S. Tett (2017) Evaluation of the Had GEM3-A simulations in view of climate and weather event human influence attribution in Europe. Climate Dynamics, Submitted.