

Assessing climate change impacts on extreme weather events: the case for an alternative (Bayesian) approach

Michael E. Mann¹ · Elisabeth A. Lloyd² · Naomi Oreskes³

Received: 5 June 2016 / Accepted: 19 January 2017 / Published online: 28 August 2017
© Springer Science+Business Media B.V. 2017

Abstract The conventional approach to detecting and attributing climate change impacts on extreme weather events is generally based on frequentist statistical inference wherein a null hypothesis of no influence is assumed, and the alternative hypothesis of an influence is accepted only when the null hypothesis can be rejected at a sufficiently high (e.g., 95% or “ $p = 0.05$ ”) level of confidence. Using a simple conceptual model for the occurrence of extreme weather events, we show that if the objective is to minimize forecast error, an alternative approach wherein likelihoods of impact are continually updated as data become available is preferable. Using a simple “proof-of-concept,” we show that such an approach will, under rather general assumptions, yield more accurate forecasts. We also argue that such an approach will better serve society, in providing a more effective means to alert decision-makers to potential and unfolding harms and avoid opportunity costs. In short, a Bayesian approach is preferable, both empirically and ethically.

1 Introduction

There is a growing body of research dealing with the issue of detecting and attributing the impact of anthropogenic climate change on various types of weather and climate phenomena (e.g., Stott et al. 2013; Bindoff et al. 2013). In recent years, attention has turned in particular to the impact climate change may be having on extreme weather phenomena (Trenberth 2011;

A comment to this article is available at doi:[10.1007/s10584-017-2049-2](https://doi.org/10.1007/s10584-017-2049-2)

Electronic supplementary material The online version of this article (doi:[10.1007/s10584-017-2048-3](https://doi.org/10.1007/s10584-017-2048-3)) contains supplementary material, which is available to authorized users.

✉ Michael E. Mann
mann@psu.edu

¹ Department of Meteorology and Atmospheric Science, Penn State University, University Park, PA 16802, USA

² History and Philosophy of Science Department, Indiana University, 652 Ballantine Hall, Bloomington, IN 47405, USA

³ History of Science Department, Harvard University, Cambridge, MA 02138, USA

Allen 2011; Blanchet and Davison 2011; Peterson et al. 2012; Otto et al. 2012; Cooley et al. 2007; NRC 2016). The traditional approach in these studies involves the use of parallel model simulations that alternately have and have not been subjected to anthropogenic increases in greenhouse gas concentrations. If a particular phenomenon of interest (e.g., record heat, heat wave duration, droughts, floods, or active tornado or hurricane seasons) is found to occur sufficiently more often in the latter case than in the former, it is concluded that a change has been detected and can be attributed to the impact of climate change. Among other assumptions, this approach assumes that the models capture the full range of processes that may impact the occurrence of extreme weather events. That assumption has been disputed by some leading scientists (Francis and Vavrus 2012; Trenberth 2011, 2012; Trenberth et al. 2015) and has long been recognized as a problem in all modeling (Oreskes et al. 1994).

Some researchers have proposed a Bayesian framework for climate change detection and attribution (Berliner et al. 2000), while other researchers have used approaches (see Hegerl et al. 2010) that sidestep the detection step or, e.g., employ fractional attributed risk (“FAR”; see, e.g., Bellprat and Doblas-Reyes 2016) which can avoid the assumptions of frequentist statistics in climate change attribution. Nonetheless, numerous recent studies (IPCC 2013; Allen 2011; Herring et al. 2015) have continued to invoke a frequentist detection and attribution approach wherein a null hypothesis of no impact is invoked, and rejection of the null hypothesis and alleged detection and attribution of a climate change impact demands rejection of the null hypothesis at a high level of likelihood (e.g., $p = 0.05$ or $p = 0.10$). To quote the most recent IPCC chapter on Detection and Attribution (emphasis added by us), “*Attribution results are typically expressed in terms of conventional ‘frequentist’ confidence intervals or results of hypothesis tests: when it is reported that the response to anthropogenic GHG increase is very likely greater than half the total observed warming, it means that the null hypothesis that the GHG-induced warming is less than half the total can be rejected with the data available at the 10% significance level. Expert judgment is required in frequentist attribution assessments, but its role is limited to the assessment of whether internal variability and potential confounding factors have been adequately accounted for, and to downgrade nominal significance levels to account for remaining uncertainties. Uncertainties may, in some cases, be further reduced if prior expectations regarding attribution results themselves are incorporated, using a Bayesian approach, but this is not currently the usual practice.*”

This philosophical approach to hypothesis testing—i.e., the frequentist framework—is widespread in the scientific community and it is common to many physical and social scientific disciplines. Indeed, it is so pervasive that some practitioners conflate it with “the scientific method,” and consider it inappropriate even to question whether it provides an appropriate interpretational framework for all scientific hypotheses (see Curry 2011; Allen 2011). Yet, despite the sense that it is deeply engrained in the history of science, its roots are rather shallow, with widespread adoption of the practice dating back only to the 1940s (Gigerenzer et al. 1989). Scientists in several fields are increasingly re-examining the appropriateness of the frequentist approach to hypothesis testing (Nuzzo 2014).

Some have argued that this philosophical framework is a by-product of the intrinsic conservatism of the scientific discipline, reflecting a tendency among scientists for “least drama” in drawing inferences and conclusions (Brysse et al. 2013; Anderegg et al. 2014). In practice, this approach embraces (see Lloyd and Oreskes, in review) a subjective preference for risking type 2 errors of statistical inference (failure to reject a false null hypothesis, i.e., “false negative”) over type 1 errors (rejecting a true null hypothesis, i.e., “false positive”). In the context of climate science, it is likely that attacks on scientists by critics, and the ensuing phenomenon of “seepage,”

whereby scientists are induced to be cautious for fear of being a target of politically-motivated criticism (Lewandowsky et al. 2015), reinforces this latent tendency.

However, in many areas of biomedicine and in pharmaceutical testing, where the principle of “first, do no harm” prevails, the standard practice (and in some cases legal requirement) is to assume harm until safety can be demonstrated (Gigerenzer and Edwards 2003). With human-caused climate change, there is a potential for greatly increased damage and loss of life, and inaction comes at a large potential societal cost. In this regard, one may argue that climate change attribution is more like biomedicine than it is, for example, like experimental psychology. Because of the potential for harm, the overly conservative frequentist framework is ethically concerning.

An alternative *Bayesian* framework (Berliner et al. 2000) has been proposed wherein one employs as a statistical *prior*, the evidence that exists—either theoretical or observational in nature—that climate change may be impacting the underlying statistical distribution of the climate variable/s of interest. (Indeed, the alternative Bayesian framework was advocated decades ago in the field of clinical pharmacology, e.g., Berry 1987). In the case of impacts on extreme weather, some researchers (Trenberth 2011, 2012, Trenberth et al. 2015; Lloyd and Oreskes (in review)) have proposed that one employs a conditional approach, by accepting a priori relevant physical principles regarding the relationship between climate and extreme weather. For example, the demonstrated warming of the planet has increased the likelihood of daily heat extremes (Meehl et al. 2007). Moreover, it has fundamentally intensified the global hydrological cycle, increasing overall levels of atmospheric humidity and the potential for extreme precipitation events (Trenberth et al. 2015). These considerations provide theoretical and mechanistic reasons—one might say *priors*—to believe that climate change is likely to be impacting extreme weather events. In sum, the Bayesian framework has the advantage, in detection and attribution practices, of having its conclusions be more clearly traceable to the underlying assumptions and dependence of results on prior assumptions.

The impacts of climate change may be compound in nature for certain types of extreme weather events. For example, with extreme precipitation events, there is both a thermodynamic component (warmer temperatures favor greater atmospheric humidity) and a dynamical component (upward vertical motion is required for condensation of moisture). In the case of tornadoes, there are likewise thermodynamic (warmer temperatures favor increased moist energy in the atmosphere) and dynamical factors (greater atmospheric wind shear favors the twisting of the atmosphere required for tornadic circulation) involved.

The projected changes in the underlying thermodynamic factors are typically better known than those in the dynamical factors, and some researchers have argued that we can draw inferences about extreme heat and extreme rainfall events based on the thermodynamic considerations (e.g., Shepherd 2014; Trenberth 2011, 2012; Trenberth et al. 2015). Others (e.g., Stott et al. 2016) have argued that as long as projected dynamical factors remain uncertain, it is not possible to draw reliable inferences about the impact that climate change will have on these events.

However, the latter argument implicates its advocates in a fallacy: that because we do not know all underlying factors with certainty, we are unable to say anything about the impact of climate change on a particular type of weather event. Consider, for simplicity, the case where the impacts of the two factors (thermodynamic and dynamical) are multiplicative and independent. In such a case, if we know with some considerable confidence (e.g., 90% likelihood) that the thermodynamic factors will favor an increase in the frequency of the extreme weather events in question, while dynamical factors are considered a toss-up (i.e., 50% likelihood), then the joint probability (70% likelihood of an increase) is far from a toss-up. Such considerations are explored in more detail by Shepherd (2016) and are implicit in the recent work of Diffenbaugh et al. (2015).

Here, we attempt to compare the two competing philosophical approaches (frequentist vs. Bayesian) to assessing climate change impacts on extreme weather phenomena. We make use of a simple model of extreme weather events wherein the occurrence of events can be classified on an annual basis as less active/below normal (“–”) or more active/above normal (“+”). We assume N years of annual/seasonal observations available at M independent locations over which we can form aggregate metrics of activity.

In a neutral (i.e., unaltered) climate, both active and inactive seasons (as measured relative to some appropriately-defined climatological mean) are equally likely. By contrast, we suppose that in a climate altered by anthropogenic increases in greenhouse gas concentrations, the distribution of events will be shifted. While this characterization could be applied to any number of types of extreme weather phenomena, there are a number of salient examples. One could consider the number of record-breaking maximum daily temperatures over the course of the year across all locations in North America, Eurasia, or the Northern Hemisphere (e.g., Meehl et al. 2007). Alternatively, one could consider the number of extreme rainfall events over the course of the year over these regions. In each of these examples, there are a priori physical reasons, as noted earlier, to anticipate that climate change will on balance *increase* the incidence of these events.

We investigate the relative performance of the two philosophically different approaches with respect to the simple conceptual model described above. We use Monte Carlo simulations to generate L realizations of a process consisting of time series of length $N = 64$ years at M sites under the assumption of both neutral (i.e., equal likelihood of less active “–” and more active “+” years) and climate change “biased” statistics (higher likelihood of “+” years).

We then examine error rates for each approach for each situation, allowing for the evaluation of both type 1 and type 2 errors of statistical inference. We conclude with a discussion of the larger implications of our findings with regard to climate risk assessment.

2 Methods

We assume that the extreme weather events of interest have some long-term climatological average rate of occurrence, and that there is equal likelihood of either fewer than average (“–”) or greater than average (“+”) numbers of events in any given year or season. This situation can be characterized by a discrete, binary-valued statistical process that is statistically equivalent to coin flipping. The probability distribution is given by the binomial distribution for N events (years in this case):

$$P(\theta, N, k) = \binom{N}{k} \theta^k (1-\theta)^{N-k} \quad (1)$$

where $k = 0, 1, 2, \dots, N$, and where θ represents the fractional probability of a “+” year ($1-\theta$ represents the fractional probability of a “–” year), and where

$$\binom{N}{k} = \frac{N!}{k!(N-k)!} \quad (2)$$

is the binomial coefficient for N events taken k at a time.

In the absence of any climate change impact, we assume an equal probability of less active (“–”) and more active (“+”) years, with an active year fraction $\theta = \theta_0 = 0.5$. We are also

interested in the alternative hypothesis that climate change has led to an *increase* in the occurrence of events. That situation is characterized by a probability $\theta = \theta_0 > 0.5$ for active (“+”) years. Note that for cases (like heat extremes, heat waves, flood frequency) where climate change can be a priori assumed to lead to an increase in likelihood, we are dealing one-sided statistical inferences and a one-tailed analysis of any change in the underlying probability distribution

In the conventional, frequentist approach to detection and attribution, we adopt a null hypothesis H_0 of an equal probability of active and inactive years ($\theta = \theta_0 = 0.5$). We reject it in favor of the alternative hypothesis H_1 of a bias toward more active years ($\theta = \theta_0 > 0.5$) only when we are able to achieve rejection of H_0 at a high (we choose the conventional critical value $p = 0.05$) level of confidence. Once we determine that the null hypothesis can be rejected at the $p = 0.05$ level, we replace the prior assumed value $\theta = 0.5$ with the value of θ determined from the accumulated raw data for the site. We do this for each of the $M = 100$ sites independently and only in the end aggregate the statistical results to define a domain-wide metric of occurrence frequency.

In the alternative, Bayesian approach, we assume (1) to represent a likelihood function for θ conditional on the available observations. As a prior, we assume an unbiased process ($\theta = 0.5$), admitting both informative (e.g., pre-specified binomial distribution centered at $\theta = \theta_0 = 0.5$) or uninformative (e.g., uniform over $[0, 1]$) prior distributions. As increasingly large numbers of data N become available, we obtain a posterior distribution of the form of (1) via Bayes Theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3)$$

where $P(A)$ is the prior, $P(B|A)/P(B)$ is the likelihood function given the data B , and $P(A|B)$ is the resulting posterior distribution, and where the revised estimate θ is defined by the peak of posterior distribution. Once again, we do the updating for each of the M sites independently, and aggregate results in the end.

For the purpose of the analysis procedure, we generated via Monte Carlo simulations of a binary-valued process of length $N_{\max} = 64$ for both (a) the unbiased case $\theta_0 = 0.5$, (b) the *modestly* biased case $\theta_0 = 0.6$, and (c) the *strongly* biased case $\theta_0 = 0.75$. The latter two cases correspond to a 20% higher and 50% higher likelihood, respectively, of active (“+”) years vs. inactive (“-”) years. Given, for example, that the rate of record-breaking warmth has doubled (i.e., exhibited a 100% increase) over the past half century (Meehl et al. 2007), our use of a 20 and even 50% increase is, at least for some extreme weather phenomena, conservative.

For each experiment, we performed parallel frequentist and Bayesian estimation of expected values for increasingly large subsets of the data series of length N_{tot} , iteratively refining our estimates of the posterior distribution and bias ($b = \theta_0 - 0.5$). We performed six sub-experiments that consist of using the first $N_{\text{tot}} = 2, 4, 8, 16, 32$, and 64 years of the total of $N_{\max} = 64$ years of data for each site. These six experiments introduce, sequentially $N = 2, 2, 4, 8, 16$, and 32 new years of data, respectively. For each set, we computed the expected ($\tilde{N}+$) number of active years based on updated estimates of θ_0 derived from the posterior distribution of the previous experiment.

For both methods, we defined an error function ε as the average over all M sites of the absolute difference between the observed ($N+$) and predicted ($\tilde{N}+$) number of active years, allowing us to objectively compare the performance of the frequentist and Bayesian

approaches. The absolute error is a reasonable proxy for quantities, such as total accrued damage, that might be of interest in the context of climate change risk.

For a single binary process with nearly equal probabilities (i.e., $\theta_0 = 0.5$ to 0.6) for the two (less active and more active) states, the signal-to-noise ratios for the bias $b = \theta_0 - 0.5$ are relatively low. However, aggregating over a set of independent sites M leads to a considerably better estimate of the signal.

We show results based on both (a) a large ($M = 100$) spatial array of sites that can be thought of, conceptually, as an average over a continental (e.g., US-wide) domain, and (b) a smaller ($M = 5$) spatial array that can be thought of conceptually as representing an average over a local region (e.g., central Pennsylvania). In both cases, we average the spatial-mean statistics over a large ($L = 100$) number of independent realizations to get representative estimates. For the Bayesian analysis, we have employed a modestly informative prior corresponding to the binomial distribution $P(\theta = 0.5, k = 2, N = 4)$. Similar results are obtained for the uninformative uniform prior [0 1].

Our approach represents a simplification of the real world impact of climate change on extreme weather, and this is indeed the point. The simplicity of the proof-of-concept we provide speaks to the generality of the underlying approach and the broad likely applicability of the conclusions drawn. We suggest that there are a number of potential extensions of the approach that are worth pursuing. Among them is the use of a steadily changing climate (rather than the simple binary biased/unbiased before/after approach taken), including allowing for scenarios such as accelerating climate change impacts and/or tipping point-like transitions.

3 Results and discussion

In the frequentist analysis, the distribution of p values for rejecting the null remains centered at $p = 0.5$ in the case of the unbiased distribution, as it should (Fig. 1). The small number of false positives (~ 2 per site breach the $p = 0.05$ level on average for the $M = 100$ site experiments) are consistent with chance expectations. In the case of the biased distribution, by contrast, there is instead a clear trend toward rejecting the null as N_{tor} increases. As we approach $N_{\text{tor}} = 64$, we find that the median of the distribution just breaches the $p = 0.05$ threshold for the modestly biased case ($\theta_0 = 0.6$), indicating that the null hypothesis of an unbiased distribution is now being rejected at roughly half the sites. For the strongly biased case ($\theta_0 = 0.75$), we observe much more rapid convergence toward rejection of the null, with more than 50% of the sites breaching the $p = 0.05$ threshold at $N_{\text{tor}} = 16$ and more than 75% of the sites at $N_{\text{tor}} = 32$. The smaller array of sites ($M = 5$) yields slightly slower convergence toward the rejection of the null, e.g., the median still lies above the $p = 0.05$ threshold for the modestly biased case at $N_{\text{tor}} = 64$. The frequentist approach, in short, yields the expected results.

In the Bayesian analysis (Fig. 2), by contrast, we find that posterior distributions efficiently converge toward the correct values for both the unbiased ($\theta_0 = 0.5$) and biased ($\theta_0 = 0.6$ and $\theta_0 = 0.75$) cases, as the number of years increases to $N_{\text{tor}} = 32$ and then $N_{\text{tor}} = 64$ years. The distributions both approach the correct mean value and narrow in width and uncertainty as more data become available. For the modestly biased case $\theta_0 = 0.6$ and $M = 100$ sites, only a small fraction ($\sim 5\%$) of the total area under the distribution lies to the left of $\theta_0 = 0.5$ for $N_{\text{tor}} = 64$, consistent with the observation above (Fig. 1) that the median p value breaches the $p = 0.05$ significance level for $N_{\text{tor}} = 64$.

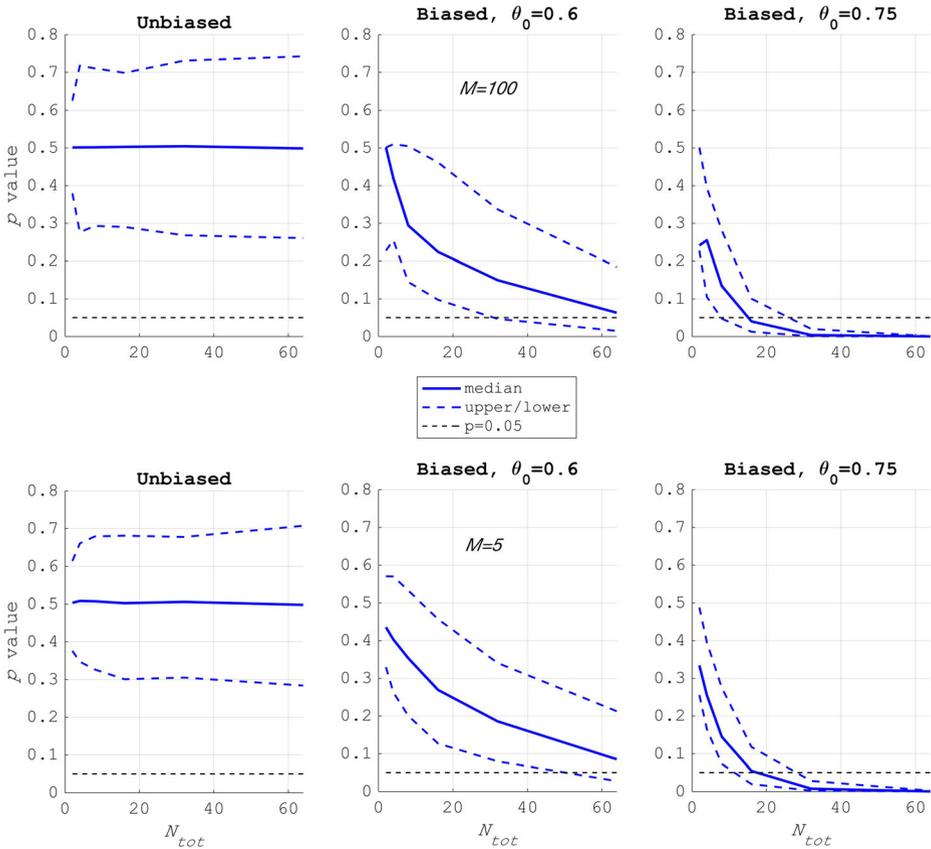


Fig. 1 p value from frequentist analysis for rejecting null hypothesis of unbiased weather statistics vs. total years of data (N_{tot}) for (left) neutral climate, (center) biased climate with $\theta_0 = 0.6$, and (right) biased climate with $\theta_0 = 0.75$. Shown are median (solid) and lower and upper interquartile range (dashed) for (upper) $M = 100$ sites and (bottom) $M = 5$ sites. As in subsequent figures, statistics have been averaged over a set of $L = 100$ realizations

Now compare the estimates of the active year fraction θ_0 for both biased and unbiased distributions, for both approaches (Fig. 3). We see that the Bayesian approach, once again, readily converges toward the correct estimates of bias in both cases. In the frequentist approach, we update θ for a site, replacing the null value ($\theta_0 = 0.5$) with an estimated θ_0 based on the available data for the site, when and only when the critical value ($p = 0.05$) for rejecting the null has been reached for that site. The estimates for the biased cases using frequentist updating converge far more slowly toward the true values than with Bayesian updating. This effect is most pronounced for the modestly biased case ($\theta_0 = 0.6$) and for the small ($M = 5$) spatial array, but is evident in all experiments. In short, frequentist updating is too conservative, and too slow, to recognizing the bias toward the active state that exists in the underlying data. As a consequence, the frequentist updating consistently underpredicts the number of active years, while the Bayesian approach yields estimates that closely match the observed numbers (see Supplementary Fig. S1).

Finally, we can compare the errors for the two approaches (Fig. 4). Consider the case where the true distribution is biased. The baseline for comparison then is the steadily

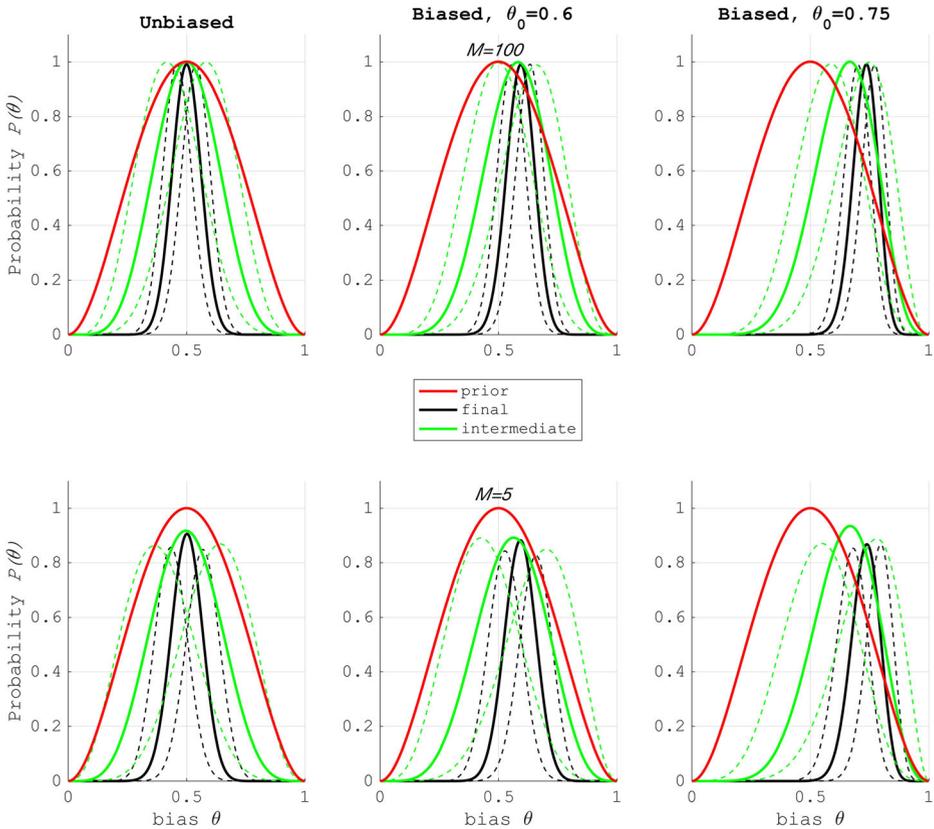


Fig. 2 Prior and posterior probability distributions from Bayesian analysis for the active year fraction (θ) for (left) neutral climate, (center) biased climate with $\theta_0 = 0.6$, and (right) biased climate with $\theta_0 = 0.75$. Shown are prior distribution (red), and posterior distributions for intermediate $N_{\text{tot}} = 32$ (green) and final $N_{\text{tot}} = 64$ (black) distributions. Shown are median (solid) and lower and upper interquartile range (dashed) for (top) $M = 100$ sites and (bottom) $M = 5$ sites

increasing absolute error associated with the null prediction of an unbiased distribution. For the strongly biased case $\theta_0 = 0.75$ with large ($M = 100$) spatial array, where convergence toward the true distribution is efficient for either method, the errors for both frequentist and Bayesian updating level off rapidly to the asymptotic constant background value. For the biased case $\theta_0 = 0.6$, however, the errors using frequentist updating continue to grow with sample size through $N = 32$, while the errors for the Bayesian approach remain substantially smaller and nearly level as the posterior distribution becomes tighter and more accurate with increasing N . The errors for the Bayesian prediction also remain lower than those for frequentist updating for the small spatial ($M = 5$) spatial array in both modestly and strongly biased cases.

While the Bayesian updating reduces errors when a bias is present, such reduced error might be offset by greater error when there is no bias present. Such increased error could result from more frequent false positives (i.e., spurious estimates of θ that depart from the true value of 0.5). However, we see that this effect is modest, as errors are similar for the two approaches in the unbiased case. That conclusion holds most strongly for the large array ($M = 100$) of sites (Fig. 4).

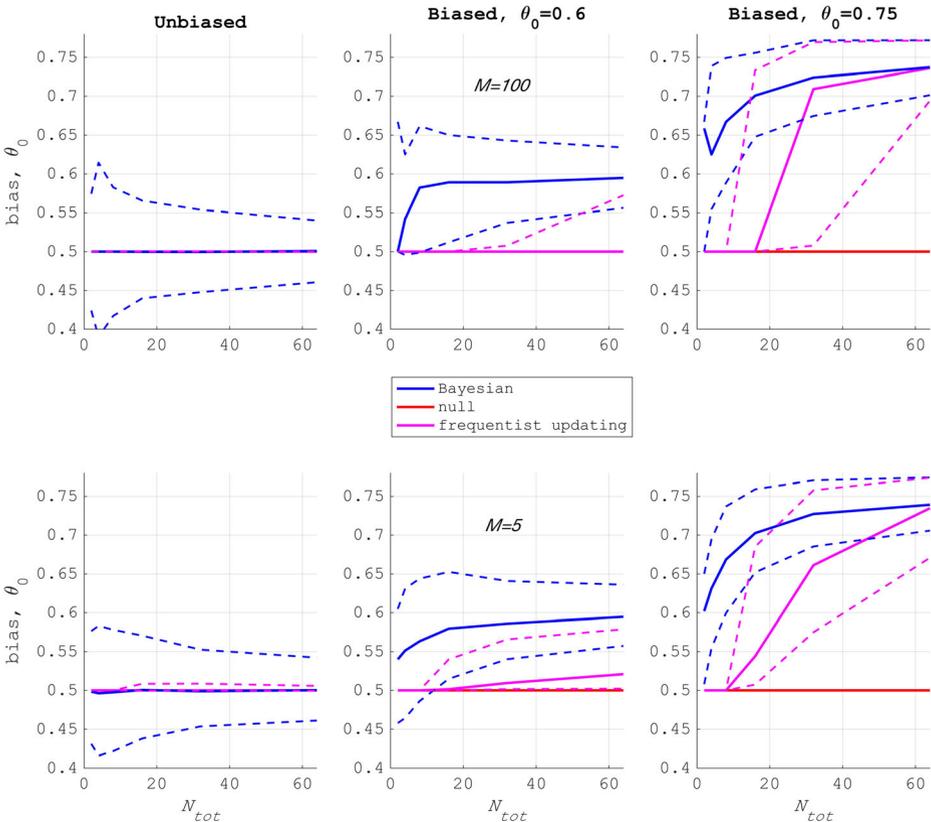


Fig. 3 Estimated active year fraction (θ_0) for (left) neutral climate, (center) biased climate with $\theta_0 = 0.6$, and (right) biased climate with $\theta_0 = 0.75$ as a function of total years of data (N_{tot}). Results are shown for frequentist updating with $p = 0.05$ (magenta), as well as Bayesian updating (blue). Shown are median (solid) and lower and upper interquartile range (dashed) for (upper) $M = 100$ sites and (bottom) $M = 5$ sites

4 Ethical considerations in climate change attribution

We have argued for the Bayesian approach on the intellectual grounds that it offers a greater likelihood of producing empirically accurate results in detection and attribution studies. This argument is reinforced by the ethical considerations of climate change attribution. It is well established that the conventional frequentist approach poses a greater risk of type 2 rather than type 1 error. In climate change detection and attribution, this means that we will tend to underestimate the danger represented by extreme events that have been worsened, or made more probable, by an anthropogenic component (Anderegg et al. 2014). This in turn means that society may underprepare for real threats, increasing the likelihood that property damage, injury, and loss of life will ensue (Lloyd and Oreskes, in review).

One might therefore argue that scientists should “err on the side of caution” and take steps to ensure that we are not underestimating climate risk and/or underestimating the human component of observed changes. Yet, as several workers have shown (Rahmstorf et al. 2007; Brysse et al., 2013), the opposite is the case in prevailing practice. Available

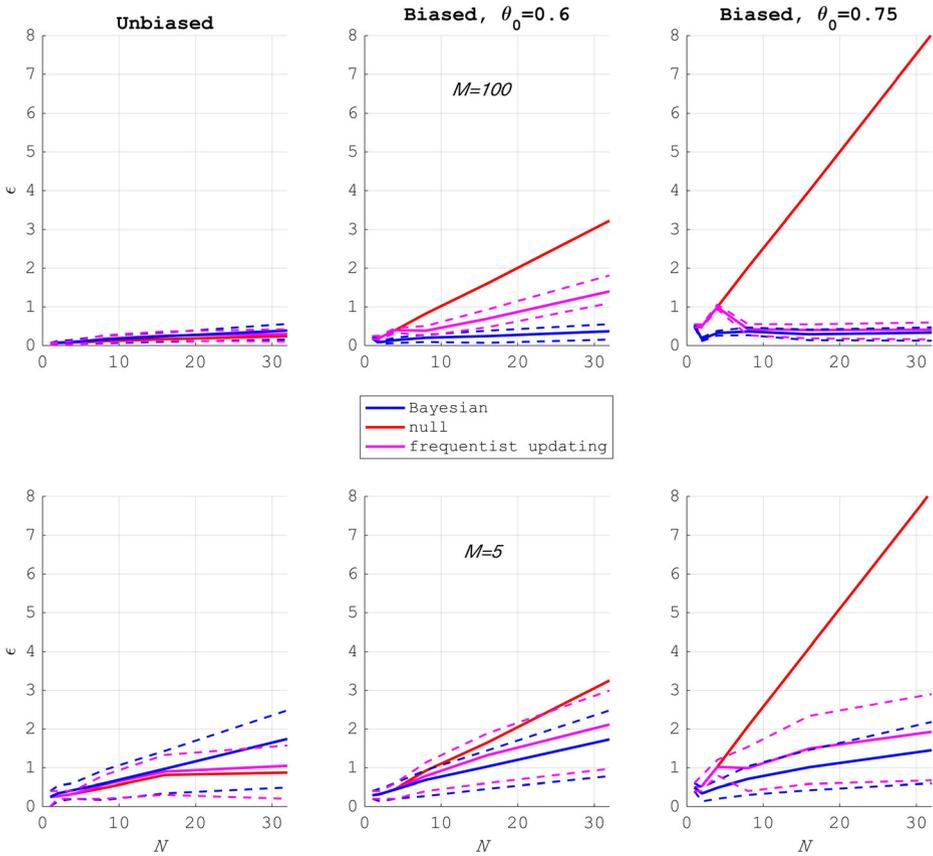


Fig. 4 Absolute error (ϵ) between average over M sites of the predicted and observed number of active years ($N+$) vs. years of data (N) for (left) neutral climate, (center) biased climate with $\theta_0 = 0.6$, and (right) biased climate with $\theta_0 = 0.75$. Results are shown for null ($\theta_0 = 0.5$) prediction (red) and null prediction using frequentist updating with $p = 0.05$ (magenta), as well as Bayesian updating (blue). Shown are median (solid) and lower and upper interquartile range (dashed) for (top) $M = 100$ sites and (bottom) $M = 5$ sites

evidence shows a tendency among climate scientists to underestimate key parameters of anthropogenic climate change, and thus, implicitly, to understate the risks related to that change. This underestimation is linked to the use of the frequentist approach with a null hypothesis of “no effect” and a high level of required confidence.

Why have scientists taken this approach, given that other approaches are not only possible, but, as we have shown, may produce more accurate results? One often offered answer is that “conservatism” is crucial for protecting the credibility and reputation of climate science, where conservatism is defined as not “crying wolf.” (Brysse et al. 2012; Oppenheimer et al. 2017). However, it is also possible to lose credibility by missing threats. More important, if we fail to alert society to unfolding harms, or understate what we know about those harms, then civil society will underprepare and people will be unnecessarily hurt. Conventional practice effectively puts protecting the reputation of climate science ahead of human safety and well-being.

Of course, there are serious risks associated with overreacting. However, given the unequivocal evidence that anthropogenic climate change is underway, and the mounting evidence that at least some hazardous phenomena have been made more likely by it, and

given the observed slow response of civil society to act to prevent those harms, it seems reasonable to conclude that the risks of underreaction to climate change are now greater than the risks of overreaction. We suggest that in such a situation, it is ethically preferable to embrace an approach that avoids understating what we know. Fortunately, as we have shown, this approach is intellectually preferable as well.

5 Conclusions

Using a simple conceptual model for the occurrence of extreme weather events, we show that the traditional frequentist approach to hypothesis testing, under a fairly general set of assumptions, suffers from a tendency to underestimate potential climate change impacts on the occurrence of extreme weather events when such an impact is present in the data. That underestimation, and the error associated with it, tends to increase substantially over time. An alternative Bayesian updating approach does not suffer from this problem.

The dominant current paradigm used in the field of detection and attribution is to employ a frequentist approach, requiring a very high threshold of significance (e.g., $p = 0.05$) for rejecting the null hypothesis of no impact. We argue that this paradigm is conceptually flawed, empirically damaging, and ethically questionable. It comes at a significant cost to the empirical accuracy of detection, leads to delayed detection and a weakened ability to respond to such detection, and increases the overall likelihood of underpreparedness for climate-related harms.

If the objective is to minimize forecast error and potential damage, an alternative Bayesian approach wherein likelihoods of impact are continually updated as data become available is preferable. We have shown that such an approach will, under rather general assumptions, yield more accurate forecasts. Indeed, a proof-of-concept for how scientists might employ a Bayesian approach to detection and attribution has already been outlined in the literature (Berliner et al. 2000). It is our recommendation that such an approach be pursued more vigorously in future work. Such an approach would be better both empirically and ethically.

Acknowledgements We thank James V. Stone, Psychology Department, Sheffield University, Sheffield, England for kindly posting the Bayesian coin flipping routine (MatLab code version 7.5. downloaded from <http://jim-stone.staff.shef.ac.uk/BayesBook/Matlab>). We thank two anonymous reviewers for the helpful comments on the initial draft of this article.

References

- Allen M (2011) In defense of the traditional null hypothesis: remarks on the Trenberth and Curry WIREs opinion articles. *WIREs Clim Change* 2:931–934. doi:10.1002/wcc.145
- Anderegg WR, Callaway ES, Boykoff MT, Yohe G, Root TYL (2014) Awareness of both type 1 and 2 errors in climate science and assessment. *Bull Am Meteorol Soc* 95:1445–1451
- Bellprat O, Doblas-Reyes F (2016) Attribution of extreme weather and climate events overestimated by unreliable climate simulations. *Geophys Res Lett* 43:2158–2164
- Berliner LM, Levine RA, Shea DJ (2000) Bayesian climate change assessment. *J Clim* 13:3805–3820
- Berry DA (1987) Interim analysis in clinical trials: the role of the likelihood principle. *Amer Stat* 41:117–122
- Bindoff NL, Stott PA, Achuta Rao KM, Allen MR, Gillett N, Gutzler D, Hansingo K, Hegerl G, Hu Y, Jain S, Mokhov II, Overland J, Perlwitz J, Sebbari R, Zhang X (2013) Detection and attribution of climate change:

- from global to regional. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge
- Blanchet J, Davison AC (2011) Spatial modeling of extreme snow depth. *Ann Appl Stat* 5:1699–1725
- Bryse K, Oreskes N, O'Reilly J, Oppenheimer M (2013) Climate change prediction: erring on the side of least drama? *Glob Environ Chang* 23(1):327–337
- Cooley D, Nychka D, Naveau P (2007) Bayesian spatial modeling of extreme precipitation return levels. *J Am Stat Assoc* 102(479):824–840, Applications and case studies. doi:10.1198/016214506000000780
- Curry JA (2011) Nullifying the climate null hypothesis. *WIREs Clim Change* 2:919–924. doi:10.1002/wcc.141
- Diffenbaugh NS, Swain DL, Touma D (2015) Anthropogenic warming has increased drought risk in California. *Proc Natl Acad Sci U S A* 112(13):3931–3936. doi:10.1073/pnas.1422385112
- Francis JA, Vavrus SJ (2012) Evidence linking Arctic amplification to extreme weather in mid-latitudes. *Geophys Res Lett* 39:L06801. doi:10.1029/2012GL051000
- Gigerenzer G, Edwards A. (2003) Simple tools for understanding risks: from innumeracy to insight, *British Medical Journal*, 327, 741–744
- Gigerenzer G, Switjink Z, Porter T, Daston L, Beatty J, Kruger L (1989) *The empire of chance: how probability changed science and everyday life*. Cambridge University Press, Cambridge
- Hegerl GC, Hoegh-Guldberg O, Casassa G, Hoerling MP, Kovats RS, Parmesan C, Pierce DW, Stott PA (2010) Good practice guidance paper on detection and attribution related to anthropogenic climate change. In: Stocker TF, Field CB, Qin D, Barros V, Plattner GK, Tignor M, Midgley PM, Ebi KL (eds) Meeting report of the intergovernmental panel on climate change expert meeting on detection and attribution of anthropogenic climate change. IPCC Working Group I Technical Support Unit, University of Bern, Bern p 8
- Herring SC, Hoerling MP, Kossin JP, Peterson TC, Stott PA (2015) Explaining extreme events of 2014 from a climate perspective. *Bull Am Meteorol Soc* 96(12):S1–S172
- IPCC (2013) Summary for policymakers. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge
- Lewandowsky S, Oreskes N, Risbey JS, Newell BR, Smithson M (2015) Seepage: climate change denial and its effect on the scientific community. *Glob Environ Chang* 33:1–13. doi:10.1016/j.gloenvcha.2015.02.013
- Meehl GA, Arblaster JM, Tebaldi C (2007) Contributions of natural and anthropogenic forcing to changes in temperature extremes over the U.S. *Geophys Res Lett* 34:L19709. doi:10.1029/2007GL030948
- National Research Council (2016) Attribution of extreme weather events in the context of climate change. National Academies Press, Washington, DC
- Nuzzo R (2014) Scientific method: statistical errors. *Nature* 506:150–152
- Oppenheimer M, Oreskes N, Jamieson D, Bryse K, O'Reilly J, Shindell M, (2017) *Assessing assessments: scientific knowledge for public policy*, University of Chicago Press. (forthcoming)
- Oreskes N, Shrader-Frechette K, Belitz K (1994) Verification, validation, and confirmation of numerical models in the earth sciences. *Science* 263:641–646
- Otto FEL, Massey N, van Oldenborgh GJ, Jones RG, Allen MR (2012) Reconciling two approaches to attribution of the 2010 Russian heat wave. *Geophys Res Lett* 39:L04702. doi:10.1029/2011GL050422
- Peterson T, Stott P, Herring S (2012) Explaining extreme events of 2011 from a climate change perspective. *Bull Am Meteorol Soc* 93(7):1041–1067
- Rahmstorf S, Cazenave A, Church JA, Hansen JE, Keeling RF, Parker DE, Somerville RC (2007) Recent climate observations compared to projections. *Science* 316(5825):709–709
- Shepherd TG (2014) Atmospheric circulation as a source of uncertainty in climate change projections. *Nat Geosci* 7:703–708
- Shepherd TG (2016) A common framework for approaches to extreme event attribution. *Curr Clim Change Rep* 2(1):28–38. doi:10.1007/s40641-016-0033-y
- Stott PA, Allen M, Christidis N, Dole RM, Hoerling M, Huntingford C, Pall P, Perlwitz J, Stone D (2013) Attribution of weather and climate-related events. In: Asrar GR, Hurrell JW (eds) *Climate science for serving society*. Springer, Haarlem, pp 307–337
- Stott P, 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. *Wiley Interdiscip Rev Clim Chang* 7(1):23–41. doi:10.1002/wcc.380
- Trenberth KE (2011) Attribution of climate variations and trends to human influences and natural variability. *WIREs Clim Change* 2:925–930
- Trenberth KE (2012) Framing the way to relate climate extremes to climate change. *Clim Chang* 115:283–290
- Trenberth KE et al (2015) Attribution of climate extreme events. *Nat Clim Chang* 5:725–730