Can the Issuance of Hazardous-Weather Warnings Inform the Attribution of Extreme Events to Climate Change?

Dale R. Durran

ABSTRACT: When extreme weather occurs, the question often arises whether the event was produced by climate change. Two types of errors are possible when attempting to answer this question. One type of error is underestimating the role of climate change, thereby failing to properly alert the public and appropriately stimulate efforts at adaptation and mitigation. The second type of error is overestimating the role of climate change, thereby elevating climate anxiety and potentially derailing important public discussions with false alarms. Long before societal concerns about global warming became widespread, meteorologists were addressing essentially the same trade-off when faced with a binary decision of whether to issue a warning for hazardous weather. Here we review forecast–verification statistics such as the probability of detection (POD) and the false alarm ratio (FAR) for hazardous-weather warnings and examine their potential application to extreme-event attribution in connection with climate change. Empirical and theoretical evidence suggests that adjusting tornado-warning thresholds in an attempt to reduce FAR produces even larger reductions in POD. Similar tradeoffs between improving FAR and degrading POD are shown to apply using a rubric for the attribution of extreme high temperatures to climate change. Although there are obviously significant differences between the issuance of hazardous-weather warnings and the attribution of extreme events to global warming, the experiences of the weather forecasting community can provide qualitative guidance for those attempting to set practical thresholds for extreme-event attribution in a changing climate.

<https://doi.org/10.1175/BAMS-D-20-0026.1>

Corresponding author: Dale R. Durran, drdee@uw.edu

In final form 25 March 2020

©2020 American Meteorological Society

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).

AFFILIATION: Durran—Department of Atmospheric Sciences, University of Washington, Seattle, Washington

There can be no question that Earth's average temperature is increasing (Sánchez-Lugo et al. 2019). Reasons for this are well understood as a consequence of increasing greenhouse gases altering the global energy balance. Because regional climate change is not solely governed by radiative energy balance, there is less certainty about the extent to which changes in regional climate are the result of natural variability or changes in greenhouse gas concentrations. For example, ensembles of climate-model simulations with essentially the same global-warming trajectories can show significant regional variations in climate change among the individual members (Deser et al. 2012). Changes in extreme events are even harder to unambiguously attribute to global warming. For many phenomena, the observational record is inadequate for establishing baselines for their occurrence in the past and the extent to which they may be changing in the current climate. Ensembles of climate-model simulations are typically of limited use because extremes occur too infrequently to be well sampled by the ensemble. Moreover, many extreme events involve mesoscale processes that are particularly challenging to incorporate in climate simulations owing to limitations in numerical resolution and uncertainties in the physical parameterizations of clouds, precipitation, and boundary layer processes.

The science of extreme-event attribution has grown rapidly in an effort to address the challenge of quantifying the extent to which climate change is contributing to increases in the frequency or severity of extreme events and to potentially guide efforts toward mitigation and adaptation. Two approaches that have been used in the attribution of extreme events to climate change are the *risk-based* and the *storyline* approaches. In the risk-based approach, the goal is to quantify the extent to which anthropogenic influences have altered the probability of occurrence of a particular type of extreme event (e.g., Stott et al. 2016). In contrast, the storyline approach is not focused on the probabilistic estimation of the impact of climate change, but rather attempts to estimate the contribution of climate change to the various physical processes responsible for an extreme event (e.g., Shepherd 2016). Rather than being completely different methods, the risk-based and storyline approaches have also been regarded as limiting cases along a spectrum of event-attribution assessments in which conditional probabilities play an increasingly significant role as the method becomes closer to the storyline or Bayesian framework (Stott et al. 2017).

Using terminology from statistical hypothesis testing, a common null hypothesis in the risk-based approach is that a particular extreme event is the product of natural variability. Two types of errors can arise when testing this hypothesis. A type I error (rejection of a true null hypothesis) occurs when an extreme event is produced by natural variability and incorrectly attributed to climate change. A type II error (failure to reject a false null hypothesis) occurs when climate change is responsible for producing that extreme, but the evidence is deemed inadequate to reject the claim that it arose from natural variability. Because it is not possible to avoid both types of error, we must weigh the likelihood and seriousness of type I and type II errors in formulating our assessment. It has been suggested that using the risk-based approach to test the preceding hypothesis tends to understate anthropogenic contributions to extreme events (type II errors), whereas the storyline approach tends to overstate them (type I errors) (Lloyd and Oreskes 2018).

One way to avoid errors that arise from simple yes–no answers is to offer more nuanced probabilistic information. In the context of frequentist risk-based attribution, this can be provided by measures such as the *risk ratio* RR , which is the ratio of the probabilities of occurrence of events above a given threshold in the presence of, to that in the absence of, anthropogenic climate change, or alternatively, the *fraction of attributable risk*: $1 - RR^{-1}$ (Stott et al. 2004; Stone and Allen 2005). Such numeric values, along with the associated probability distributions for extreme events in the baseline and current climates, might potentially be weighted by the estimated costs of reducing the risks from extremes to arrive at optimal strategies for mitigation and adaptation. Nevertheless, specific frameworks for environmental risk

assessment are extremely difficult to formulate and implement for several reasons, including that the relevant probabilities and the costs are both poorly constrained (Jones 2001), and because the population that would most benefit from risk avoidance may be different from the population that bears its cost.

The political will to incur any cost is highly dependent on stakeholder perceptions of the risks from anthropogenic climate change. At present, the American public's perception of these risks is still influenced by yes–no answers to the basic question: was a given extreme weather event made more likely or more severe by climate change? Even if scientists express their conclusions in probabilistic terms, the takeaway message conveyed to the public by the media is often simplified to “yes” or “no.” The provision of yes–no answers to a series of hypothesis tests also constituted a basic methodology used in a recent article in this journal on the detection and attribution of the influence of climate change on tropical cyclones (Knutson et al. 2019).

While being cognizant of its limitations, in the remainder of this paper we will therefore focus on considerations related to binary yes–no decision-making with the goal of highlighting connections between two problems that are nominally very different: the issuance of hazardous-weather warnings and the attribution of extreme weather events to climate change. *An important goal in both cases is to properly alert the public about atmospheric phenomena that can have dramatic impacts on life and property.*

Weather forecasters have long dealt with the tradeoffs between over- and under-warning for hazardous-weather events; forecasting metrics and historical performance data for U.S. National Weather Service (NWS) tornado warnings will, therefore, be presented in the second section. The application of similar metrics in the attribution of extreme high temperatures to climate change is examined in the third section, along with the relation between the weather-forecast metrics and other measures more traditionally used in extreme-event-attribution studies. Forecast metrics for warnings issued for other types of hazardous-weather events will be reviewed in the fourth section. The fifth section contains the conclusions.

To warn or not to warn

One important weather hazard is the tornado; as with the impacts of global warming over a longer time scale, lives are at stake when considering when and whether to sound the alarm. Consider three forecasters, Jane, Jennifer, and Jaclyn, who are deciding whether to issue tornado warnings during a period when 80 tornadoes actually occur.¹ Jane issues a total of 200 warnings during this period, 52 of which verify. Jennifer is much more conservative, issuing only 10 warnings when she is highly confident, and all 10 of these warnings verify. Jaclyn tries a more middle-of-the-road approach, issuing 130 warnings, 40 of which verify. Who is the best forecaster?

Two key measures of forecast quality are the probability of detection (POD) and the false alarm ratio (FAR). These metrics are defined by considering the possible outcomes involving yes/no forecast decisions:

- A warning is issued, and the event occurs. (Let a denote the number of these cases.)
- A warning is issued, but the event does not occur. (Let b denote the number of these cases.)
- No warning is issued, but the event does occur. (Let c denote the number of these cases.)
- No warning is issued, and no event occurs. (Let d denote the number of these cases.)

The POD is the ratio of the number of correctly warned events to all the actual hazardous events and is defined mathematically

¹ The NWS issues *warnings* when the public should take action and there is imminent danger to life or property. Warnings should not be confused with watches, which are issued at a lower threat level and advise the public to be prepared for the possibility of a tornado.

as $POD = a/(a + c)$. The FAR is the ratio of the number of warnings issued for events that did not occur to the total number of warnings;² it is mathematically defined as $FAR = b/(a + b)$. The performance of our three forecasters is evaluated in terms of POD and FAR in Table 1. Jane, with the most correctly forecast events, has the highest POD of 0.65, but she also has the highest FAR. Jennifer's cautious approach has yielded a perfect FAR of 0, but a very poor POD of 0.125. Unsurprisingly, Jaclyn's results lie between those of the other two. Jaclyn reduced her FAR relative to Jane by 0.05; note however, that in reducing her FAR, she also reduced her POD relative to Jane by 0.15. That is, Jaclyn's improvement in FAR was accompanied by 3 times more degradation in POD.

Whose approach is best? If we temporarily leave aside the tornado-forecasting context and attempt to answer this question in the context of extreme-event attribution and climate change, one might deduce from the current debate that it is largely a matter of personal opinion about acceptable levels of error. Some claim we must be very conservative, essentially taking Jennifer's approach. Others might claim we should be following Jane, being sure the alarm is sounded at essentially every opportunity.

Let us return to the question of tornado forecasting and consider how the actual practice of the NWS compares with the performance of our hypothetical forecasters. Figure 1 shows POD and FAR scores for nationwide NWS tornado warnings issued between 1990 and 2017.³ Perhaps surprisingly, the FAR has remained near 0.75 during this entire 28-yr period. Roughly 3 out of every 4 tornado warnings are false alarms! In contrast the changes in POD are rather large: rising from about 1/3 early in the period to values ranging between 0.64 and 0.72

between 2002 and 2011, before finally slipping back to roughly 1/2 after 2012. What is responsible for these changes in the POD and why are they so much larger than the changes in the FAR?

The gradual improvement in POD between 1990 and 2002 can largely be attributed to the deployment of the WSR-88D Doppler radar network and to improvements in the scientific understanding of tornadoes, the numerical guidance, and the training of forecasters in the use these assets (Brooks 2004). Improvements in POD are of course a welcome development, but what happened after 2011? The death toll from U.S. tornadoes in 2011, 553, was unusually high, with 161 fatalities resulting from a single tornado on 22 May in Joplin, Missouri. A subsequent investigation of the Joplin tornado identified high false-alarm ratios for NWS tornado warnings in the Joplin area as a concern and possible contributor to the high number of fatalities (Kuligowski et al. 2014). The drop in POD

Table 1. Forecaster performance as measured by the probability of detection (POD; 1 is perfect) and false alarm ratio (FAR; 0 is perfect) in a hypothetical period with 80 observed tornadoes.

Forecaster	Hits/warnings	POD	FAR
Jane	52/200	0.65	0.74
Jennifer	10/10	0.125	0.0
Jaclyn	40/130	0.5	0.69

Fig. 1. Nationally averaged probability of detection (POD; blue) and false alarm ratio (FAR; red) for tornado warnings issued in advance by the NWS as a function of year. The red dashed line at 0.75 is a reference.

² The false alarm ratio should not be confused with terminology in signal detection theory (Mason 1982; Maxmillan and Creelman 1991), in which the probability of a false alarm is defined as the ratio of the nonevents for which warnings are issued to all nonevents or $b/(b + d)$.

³ The POD data are from Brooks and Corriea (2018), and are for warnings issued before the appearance of a tornado; the complementary FAR data were provided by H. Brooks (2019, personal communication).

evident in Fig. 1 after 2011 can be attributed to “an apparent emphasis on reducing FAR that led to a change in the threshold for issuing warnings, such that fewer warnings were issued” (Brooks and Corriea 2018). Similar to the differences between Jane and Jaclyn, the actual degradation in the NWS forecast POD from the period 2002–2011 to 2012–2017 is 3 times the corresponding improvement in FAR.

Returning to our hypothetical forecasters, we see that Jane’s performance metrics were roughly similar to the NWS’s performance over the period 2002–2011, whereas Jaclyn’s numbers are similar to those of the NWS between 2012 and 2017. In the context of tornado warnings, both Jane and Jaclyn are performing at levels consistent with the NWS in the several years prior to, or after, 2011. Jennifer on the other hand, is following a strategy completely at variance with operational practice—a practice that is nominally set by societal values of acceptable risk when lives are at stake.

The large changes in POD relative to FAR in Fig. 1 can be modeled using signal detection theory. Brooks and Corriea (2018) estimated that for the quality of the NWS observing and forecasting system in the years 2008–2011, a change in the tornado warning threshold that produces a decrease in FAR should produce somewhat more than twice the decrease in POD. Reducing FAR without significantly impacting POD would require significant improvements in the discriminating capabilities of the operational forecast system.

Extreme-temperature attribution

There are obvious differences between deciding whether to issue a tornado warning and deciding whether to attribute some aspect of an extreme-weather event to climate change. Forecasts of hazardous weather need to occur in real time or near-real time to inform emergency response and decisions about personal and public safety, whereas climate-change event attribution tends to be post hoc and intended to inform the societal responses to the longer-term consequences of anthropogenic climate change. The actions that the public are advised to take in response to a tornado warning, and the cost of those actions, are certainly different from those that might be expected to arise from the confident attribution of extreme events to global warming.

Nevertheless, it is worth noting that one similarity between tornado warnings and extreme-event attribution is that those in disadvantaged communities with limited financial resources are often unable to do much in response to the information. Almost 20% of the population in the southeastern United States lives in mobile homes (Crockett 2019). It is not safe to remain in a mobile home during a tornado. People are advised to evacuate to a storm shelter or a friend’s permanent house—yet many on the lowest rungs of the economic ladder lack access to such refuge.⁴ Similarly, there is often little that disadvantaged communities can do to mitigate or adapt to climate change (Dulal et al. 2010) despite their substantial exposure to its impacts (Mendelsohn et al. 2006).

Perhaps *the most fundamental goal, common to both tornado warnings and extreme-event attribution, is to communicate simply and clearly what the science is telling us about the state of the atmosphere*, and in this context, let us consider how the preceding considerations about acceptable values of POD and FAR in tornado warnings would translate to the problem of attributing increasing daily maximum temperatures to climate change. The first obvious problem with the application of metrics from hazardous-weather warnings to extreme-event attribution is that, in principle, it is relatively straightforward to verify every hazardous-weather warning and thereby accumulate statistics for POD and FAR. In contrast, the whole question

⁴ Longer-time-scale efforts at adaptation to tornadoes are not conditioned on tornado warnings themselves, but they can also have surprising parallels to political considerations that arise when proposing adaptations to climate change. For example, on 23 May 2014, one year and three days after seven children were killed when a tornado hit an elementary school in Moore, Oklahoma, the Oklahoma legislature rejected a proposal “for a statewide vote on allowing school districts, with local voter approval, to increase their bonding authority once to build tornado shelters” (Green 2014).

of extreme-event attribution arises because it is not possible to say with absolute certainty whether any particular extreme event occurred “because” of climate change. Our strategy will, therefore, be to consider attribution rubrics for which we can evaluate the expected average values of POD and FAR provided we are able to determine statistical distributions for the probability with which the event would occur both a world without anthropogenic warming and in the current climate. These are the same statistical distributions that would be required to evaluate the risk ratio and the fraction of attributable risk for traditional risk-based attribution.

We will use an idealized risk-based approach motivated by the analysis of Hansen et al. (2012) of Northern Hemisphere summer-season temperature extremes over land. They found the normalized and spatially aggregated distribution was approximately normal and exhibited a gradual, decade-by-decade, shift to warmer values relative to the 1951–80 reference period, as illustrated in Fig. 2. A variety of practical details must be considered when analyzing actual data to obtain plots like Fig. 2 (Rhines and Huybers 2013; Sippel et al. 2015). For the purposes of this discussion we will assume we have arrived at an accurate representation

Fig. 2. Frequency of occurrence of local June–August temperature anomalies (relative to the 1951–80 mean) for Northern Hemisphere land in units of local standard deviation (horizontal axis). Temperature anomalies in 1951–80 closely match the normal distribution (green curve), which is used to define cold (blue), typical (white), and hot (red) seasons, each with probability 33.3%. Also indicated is the area greater than 3σ (deep red). Computations use the same set of 1,886 stations that were reporting over the entire period. Adapted from Hansen et al. (2013).

of the probability distribution for the daily maximum temperatures in a given month in the natural world without anthropogenic influences (the control climate), and a second distribution representing corresponding maximum temperatures in the current climate. For simplicity, both sets of temperature observations are assumed to be normally distributed with the same standard deviation σ . The mean daily maximum temperature in the warmer world is assumed to have shifted toward warmer temperatures by $\sigma/2$ from the mean in the control climate μ_c . An extreme warm event will be defined as a daily maximum temperature that is more than 2σ warmer than μ_c , or equivalently, the warmest 2.3% of the climatological distribution of the daily maximums for the given month in the control climate. According to this definition, an extreme daily maximum temperature in a given month in the control climate would occur, on average, about four times in a 6-yr period.

Consider a family of attribution rubrics in which global warming is deemed to have produced an extreme event if the daily maximum temperature exceeds μ_c plus various multiples of σ . Let us begin by considering a rubric in which every extreme event is attributed to global warming (i.e., the attribution threshold is $\mu_c + 2\sigma$). Because such events obviously also occurred in the control climate, this rubric might be considered biased toward producing type I errors, that is, toward “crying wolf.”⁵ What is the false alarm ratio associated with this rubric?

The FAR can be computed from the probability distributions for the control and current climates illustrated in Fig. 3. The

⁵ The extent to which people become desensitized to warnings because of false alarms is a matter of debate and continuing research; see, for example, LeClerc and Joslyn (2015) or Lim et al. (2019).

probability of incorrectly attributing the extreme temperature event to global warming is given by the area of the region labeled b , whereas the probability of correctly attributing the event to global warming is given by area a . As was the case with the tornado warnings, $\text{FAR} = b/(a + b)$, which evaluates to 0.34, a value far less than the false alarm ratios associated with NWS tornado warnings. What is the probability of detection under this rubric? Since all extreme high temperatures are attributed to global warming, the POD is a perfect 1.0.

The preceding analysis is nothing more than a reframing of standard risk-based assessment in terms of FAR and POD. The false alarm ratio is closely related to the two parameters commonly used in risk-based extreme-event attribution mentioned in the introduction: the risk ratio RR and the fraction of attributable risk, which will be denoted as $\mathcal{FAR}$ to distinguish it from the false alarm ratio. From Fig. 3, the probability that events exceeding the prescribed attribution threshold will occur in the current climate is $a + b$, while the probability they would have occurred in the control climate is b , so the risk ratio is $\text{RR} = (a + b)/b = 1/\text{FAR}$. The fraction of attributable risk, $\mathcal{FAR}$, equals $1 - \text{RR}^{-1}$, and therefore satisfies $\mathcal{FAR} = 1 - \text{FAR}$. A threshold for extreme-event exceedance giving a risk ratio is 3 and a fraction of attributable risk of 2/3, may alternatively be characterized as giving a false alarm ratio of 1/3, which is approximately the situation illustrated in Fig. 3, and as previously noted, is far better than the false alarm ratio for tornado warnings.

The threshold for attribution of extremes due to climate change need not, however, be identical to the cutoff criteria used to define an extreme. One might suppose, for example, that we could more confidently attribute extreme maximum temperature events to climate change if we use a higher attribution threshold. Consider, therefore, how FAR and POD change if we shift the rubric to attribute only those events with maximum temperatures greater than 2.5σ above μ_c to climate change. (The definition of an extreme event as a temperature exceeding $\mu_c + 2\sigma$ remains unchanged). The probabilities for this case are illustrated in Fig. 4. As expected, limiting the attribution to an more extreme subset of all 2σ extremes decreases the FAR,

Fig. 3. Hypothetical probability distributions for the daily maximum temperature perturbation $T'_{\max}$ in the control climate (blue) and the current climate (red). Perturbations are computed with respect to the mean in the control climate μ_c . Both probabilities are normally distributed with the same standard deviation σ and plotted as a function of normalized temperature perturbations $T'_{\max}/\sigma$. The mean in the current climate is $\sigma/2$ warmer than μ_c . Areas a and b illustrate the contributions to POD and FAR using a rubric in which all normalized temperatures greater than $\mu_c + 2\sigma$ are attributed to global warming.

Fig. 4. As in Fig. 3, except that the rubric illustrated attributes all normalized temperatures greater than $\mu_c + 2.5\sigma$ to global warming.

which drops to 0.27. What happens to the POD? Under this new rubric, the probability that an extreme event is produced by climate change, but *not* correctly attributed to that change, is given by area *c*. Analogous to the tornado warning problem, the probability of detection is $POD = a/(a + c)$, which evaluates to 0.38. Shifting to the more restrictive rubric produced a decrease in the POD that is 9 times the decrease in the FAR!

If one were to consider a FAR of 0.27 to nevertheless be excessive, an even more restrictive attribution rubric might be adopted in which only those extremes with temperatures greater than 3σ above μ_c were attributed to climate change. The FAR and POD for this rubric, along with those for the previous cases, are given in Table 2. Increasing the attribution threshold to 3σ produces a further modest decrease in FAR to 0.21 while driving the POD all the way down to 0.11. *As in the tornado warning system, tightening attribution thresholds produces reductions in both the FAR and POD, with substantially larger reductions in the POD.*

Of course for each scenario in Table 2, we can shift the POD back to 1.0 if we move the goalposts by shifting the cutoff defining an extreme maximum temperature for a given month from a value that occurs in the control climate an average of four times in six years (the 2σ case) to a value that occurs once every six years (the 2.5σ case) or once every 25 years (the 3σ case). The most extreme events, as well as compound events at lower temperature thresholds (Baldwin et al. 2019), are typically the most impactful, so very extreme events are important. But that does not mean they should dominate the public discussion of the attribution question. It would be helpful if, with the help of social scientists, we could better understand what the public envisions as a practically important threshold for an extreme maximum temperature. Another difficulty in focusing on very extreme events is that their expected frequency is highly sensitive to the specification of the tails in the probability distributions defining those events, and in many applications where those probabilities might be described by generalized extreme value (GEV) distributions, it can be challenging to determine the tails with confidence.

Finally, note that the FAR and POD also characterize the probability of error arising when assessing the null hypothesis that extreme events above a given threshold are produced by natural variability. The FAR is the probability of a type I error, and the probability of a type II error is $1 - POD$. In contrast to the situation for standard risk-based attribution, which Lloyd and Oreskes (2018) suggest is biased toward type II errors, under the yes–no attribution metric illustrated in Fig. 3 the probability of a type II error drops to zero, and method is biased toward type I errors.

Relation to other hazardous-weather events

Tornadoes pose the most difficult severe-weather-warning challenge. What are the characteristic values of POD and FAR for warnings issued for other hazards? Table 3 lists these metrics for severe thunderstorms, flash floods, high winds, and winter storms using NWS data for fiscal year 2019. The 2019 values are representative of the performance over the last few years. Unsurprisingly, the FAR are lower, and the POD are higher, for all these hazards than for tornadoes, yet these FAR values,

Table 2. Impact on FAR and POD of changing the threshold temperature in an attribution rubric where an extreme high temperature is deemed to be boosted by global warming if the temperature exceeds the control climate mean μ_c by various multiples of one standard deviation σ . Calculations assume the temperatures are normally distributed and that the mean in the warmer climate has shifted warmer by $\sigma/2$, while the variance remains unchanged.

Metric	2σ	2.5σ	3σ
FAR	0.34	0.27	0.22
POD	1.0	0.38	0.11

Table 3. Nationwide average FAR, POD, and lead-time statistics for additional NWS warnings from FY 2019, courtesy of the National Weather Service.

Metric	Severe thunderstorms	Flash floods	High winds	Winter storms
FAR	0.51	0.44	0.56	0.45
POD	0.82	0.78	0.69	0.82
Lead time	18 min	64 min	12 h	21 h

which range from 0.44 to 0.56, are still much higher than those associated with the climate-attribution rubrics considered in the preceding section. Also plotted in Table 3 are the average lead times at which warnings were issued for each type of hazard. Decreases in forecast lead time can produce decreases in FAR, and in fact the decrease in FAR for tornadoes before and after 2011 shown in Fig. 1 was associated with a decrease in average tornado warning lead time from roughly 13 to about 9 min.

A rather different trade-off between FAR and lead time occurred for high-wind warnings between 2008 and 2019. As shown in Fig. 5, there was little change in the high-wind POD and FAR over this period, but a significant improvement in the forecast lead time, which increased from 6.1 to 11.6 h. Rather than attempting to reduce FAR values in excess of 0.5, the weather service effectively prioritized getting earlier alerts out to the public. In the context of extreme-event attribution in a warming world, getting information out earlier would amount to using the rubrics described in the “Extreme-temperature attribution” section while the shift in the mean between the control and current climate remains relatively small.

Values of POD as a function of FAR for the extreme-attribution example presented in the “Extreme-temperature attribution” section are plotted as the blue line in Fig. 6 for a family of attribution metrics in which extreme events (again defined as values larger than $\mu_c + 2\sigma$) are identified as arising from climate change when the variable in question exceeds $\mu_c + a\sigma$ as a varies continuously between 2 and 4. The specific values for the cases illustrated in Figs. 3 and 4 are indicated by blue dots. Corresponding curves are also

Fig. 5. Nationally averaged probability of detection (POD, blue), false alarm ratio (FAR, red), and forecast lead time (green) for high-wind warnings issued by the NWS as a function of fiscal year.

Fig. 6. POD as a function of FAR for the maximum temperature attribution metrics in the “Extreme-temperature attribution” section with thresholds $\mu_c + a\sigma$ for all values a between 2 and 4. The curves show the value for normal distributions where the mean has warmed relative to the control climate by $\sigma/2$ (blue), σ (orange), and $3\sigma/2$ (green). Points plotted as symbols are the 2019-averaged FAR and POD metrics for NWS warnings for high wind (HW), severe thunderstorms (ST), winter storms (WS), and flash floods (FF). T1 and T2 plot the average (FAR, POD) values from Fig. 1 for tornado warnings from 2003 to 2011, and 2014 to 2017, respectively. The (POD, FAR) values for the cases illustrated in Figs. 3 and 4 are indicated by blue dots.

plotted for shifts of the mean of the normal distribution in the current climate, relative to the control climate, of σ (orange) and 1.5σ (green). In all cases, the POD decreases very rapidly if α is increased in an effort to reduce the FAR. Also plotted in Fig. 6 are the 2019 averaged POD and FAR metrics for NWS winter-storm, flash-flood, high-wind, and severe-thunderstorm warnings from Table 3, and for tornado warnings over the periods from 2003 to 2011 and 2014 to 2017 from Fig. 1. Note that the seemingly naive rubric of attributing every extreme event to global warming (i.e., choosing $\alpha = 2$ for any of the three curves) yields a POD of 1.0, thereby avoiding all type II errors, while incurring FAR values much smaller than those associated with the NWS severe weather warnings.

Discussion and conclusions

When it is possible to determine probability distributions for the occurrence of some extreme event in the current climate and in a control climate uninfluenced by anthropogenic forcing, those two distributions themselves contain the most complete information about changes in the occurrence of the event in the warmer world. Using methods adapted from epidemiology (Stone and Allen 2005), this information can be distilled into important measures of the warming-induced changes in the probability of events exceeding a given threshold, such as the risk ratio and the fraction of attributable risk: measures that have proved very useful in previous attribution studies. Here we have considered how the same information about shifts in the probability distribution of some event under climate change might be used in connection with attribution rubrics whose expected average performance can be assessed using familiar weather-forecast-verification metrics like the probability of detection and the false alarm ratio.

In our maximum-temperature example, we considered a modest one-half standard-deviation shift in the mean from the control to the current climate. As shown in Fig. 6, when larger shifts occur, the false alarm ratios are substantially reduced, and the contributions from climate change can become rather clear cut. If the shift in the mean is a full standard deviation, as suggested by the rightmost panel in Fig. 2, the FAR and the probability of a type I error arising from the seemingly incautious strategy of attributing all temperatures greater than $\mu_c + 2\sigma$ to global warming decreases to just 0.14. The probability of detection in this case is a perfect 1.0; the probability of a type II error is zero, and the POD and FAR metrics would be far better than those achievable by meteorologists forecasting hazardous weather.

Nevertheless, as discussed previously, the parallels between the issuance of hazardous-weather warnings and the attribution of extreme events under climate change are limited. The goal in this discussion is not to suggest specific attribution thresholds, but to share information about the issuance of severe weather warnings with those working on climate change attribution. Given the time lag between greenhouse gas emissions and the steady-state warming they induce in Earth's climate, it is clearly advantageous to inform the public about events attributable to global warming as early as possible. When the signal from such warming is small, errors involving over-attribution will be associated with any reasonable attribution threshold. *As we have illustrated, both in the context of tornado warnings and the attribution of extreme maximum temperatures, attempts to reduce the false alarm ratio are accompanied by much larger reductions in the probability of detection.* To allow the NWS forecasting system to achieve its current probabilities of detection for the hazardous-weather conditions considered in this paper, their forecasts also generate lots of false alarms, with FAR values ranging between 0.44 to 0.75. Global warming is another grave threat to mankind, and its impact will grow in the future. When the public asks if some extreme weather event is linked to global warming, and atmospheric scientists formulate their answer, they are faced with tradeoffs reminiscent of those faced by forecasters issuing warnings for hazardous weather. The gravity of the threat from global warming, and the relatively small signal currently providing

evidence for such warming, suggest that we may need to follow the practice of operational meteorologists and tolerate a significant false alarm ratio to properly alert the public (thereby avoiding excessive type II error).

One might argue that there are political considerations that should take precedence over the science. The damage done by severe weather is plainly obvious, whereas opinions about the severity of the threat from global warming are strongly linked to an individual's political viewpoint. Political considerations are indeed important in any effort to enact policies to counter global warming, but they should not be conflated with science. Jennifer's overly cautious tornado forecasting strategy might be argued by some to be smart politics in the context of attributing extreme events to global warming, but it is inconsistent with the way meteorologists warn for a wide range of hazardous weather and arguably with the way society expects to be warned about threats to property and human life.

There is one important additional factor to consider when comparing decision making in tornado forecasting to that in attributing extreme events to global warming. If a forecaster fails to warn for a tornado there may be serious consequences and loss of life, but missing the forecast does not make next year's tornadoes more severe. On the other hand, every failure to alert the public about those extreme events actually influenced by global warming facilitates the illusion that mankind has time to delay the actions required to address the source of that warming. Because the residence time of CO₂ in the atmosphere is many hundreds to thousands of years (National Research Council 2011, p. 75), the cumulative consequences of such type II errors can have a very long lifetime.

Acknowledgments. The author has greatly benefited from conversations with Harold Brooks, Susan Solomon, Alexandra Anderson-Frey, Jane Baldwin, Brad Colman, Andrew DeLaFrance, and from comments by anonymous reviewers. NWS forecast verification metrics were kindly provided by Charles Kluepfel.

References

- Baldwin, J. W., J. B. Dessy, G. A. Veechi, and M. Oppenheimer, 2019: Temporally compound heat wave events and global warming: An emerging hazard. *Earth's Future*, **7**, 411–427, <https://doi.org/10.1029/2018EF000989>.
- Brooks, H. E., 2004: Tornado-warning performance in the past and future: A perspective from signal detection. *Bull. Amer. Meteor. Soc.*, **85**, 837–843, <https://doi.org/10.1175/BAMS-85-6-837>.
- , and J. Corriea, 2018: Long-term performance metrics for National Weather Service tornado warnings. *Wea. Forecasting*, **33**, 1501–1511, <https://doi.org/10.1175/WAF-D-18-0120.1>.
- Crockett, C., 2019: Tornado warnings don't adequately prepare mobile home residents. *Eos, Trans. Amer. Geophys. Union*, **100**, <https://doi.org/10.1029/2019EO123717>.
- Deser, C., R. Knutti, S. Solomon, and A. S. Phillips, 2012: Communication of the role of natural variability in future North American climate. *Nat. Climate Change*, **2**, 775–779, <https://doi.org/10.1038/nclimate1562>.
- Dulal, H. B., G. Brodnig, H. K. Thakur, and C. Green-Onoriose, 2010: Do the poor have what they need to adapt to climate change? A case study of Nepal. *Local Environ.*, **15**, 621–635, <https://doi.org/10.1080/13549839.2010.498814>.
- Green, R., 2014: Oklahoma school tornado shelter measure: Fails, passes, then fails again. *Oklahoman*, <https://oklahoman.com/article/4850541/oklahoma-school-tornado-shelter-measure-fails-passes-then-fails-again>.
- Hansen, J., M. Sato, and R. Ruedy, 2012: Perception of climate change. *Proc. Natl. Acad. Sci. USA*, **109**, E2415–E2423, <https://doi.org/10.1073/pnas.1205276109>.
- , —, and —, 2013: Reply to Rhines and Huybers: Changes in the frequency of extreme summer heat. *Proc. Natl. Acad. Sci. USA*, **110**, E547–E548, <https://doi.org/10.1073/pnas.1220916110>.
- Jones, R. N., 2001: An environmental risk assessment/management framework for climate change impact assessments. *Nat. Hazards*, **23**, 197–230, <https://doi.org/10.1023/A:1011148019213>.
- Knutson, T., and Coauthors, 2019: Tropical cyclones and climate change assessment: Part I. Detection and attribution. *Bull. Amer. Meteor. Soc.*, **100**, 1987–2007, <https://doi.org/10.1175/BAMS-D-18-0189.1>.
- Kuligowski, E. D., F. T. Lombardo, L. T. Phan, M. L. Levitan, and D. P. Jorgensen, 2014: Technical investigation of the May 22, 2011 tornado in Joplin, Missouri. National Institute of Standards and Technology Tech. Rep. NCSTAR 3, 428 pp.
- LeClerc, J., and S. Joslyn, 2015: The cry wolf effect and weather-related decision making. *Risk Anal.*, **35**, 385–395, <https://doi.org/10.1111/risa.12336>.
- Lim, J. R., B. F. Liu, and M. Egnoto, 2019: Cry wolf effect? Evaluating the impact of false alarms on public responses to tornado alerts in the southeastern United States. *Wea. Climate Soc.*, **11**, 549–563, <https://doi.org/10.1175/WCAS-D-18-0080.1>.
- Lloyd, E., and N. Oreskes, 2018: Climate change attribution: When is it appropriate to accept new methods? *Earth's Future*, **6**, 311–325, <https://doi.org/10.1002/2017EF000665>.
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- Maxmillan, N. A., and C. D. Creelman, 1991: *Detection Theory: A Users Guide*. Cambridge University Press, 407 pp.
- Mendelsohn, R., A. Dinar, and L. Williams, 2006: The distributional impact of climate change on rich and poor countries. *Environ. Dev. Econ.*, **11**, 159–178, <https://doi.org/10.1017/S1355770X05002755>.
- National Research Council, 2011: *Climate Stabilization Targets: Emissions, Concentrations, and Impacts over Decades to Millennia*. National Academies Press, 298 pp.
- Rhines, A., and P. Huybers, 2013: Frequent summer temperature extremes reflect changes in the mean, not the variance. *Proc. Natl. Acad. Sci. USA*, **110**, E546, <https://doi.org/10.1073/pnas.1218748110>.
- Sánchez-Lugo, A., P. Berrisford, C. Morice, and J. P. Nicolas, 2019: Global surface temperature [in "State of the Climate in 2018"]. *Bull. Amer. Meteor. Soc.*, **100** (9), S11–S14, <https://doi.org/10.1175/2019BAMSSTATEOFTHECLIMATE.1>.
- Shepherd, T. G., 2016: A common framework for approaches to extreme event attribution. *Curr. Climate Change Rep.*, **2**, 28–38, <https://doi.org/10.1007/s40641-016-0033-y>.
- Sippel, S., J. Zscheischler, M. Heimann, F. E. L. Otto, J. Peters, and M. D. Mahecha, 2015: Quantifying changes in climate variability and extremes: Pitfalls and their overcoming. *Geophys. Res. Lett.*, **42**, 9990–9998, <https://doi.org/10.1002/2015GL066307>.
- Stone, D. A., and M. R. Allen, 2005: The end-to-end attribution problem: From emissions to impacts. *Climatic Change*, **71**, 303–318, <https://doi.org/10.1007/s10584-005-6778-2>.
- Stott, P. A., D. A. Stone, and M. R. Allen, 2004: Human contribution to the European heatwave of 2003. *Nature*, **432**, 610–614, <https://doi.org/10.1038/nature03089>.
- , and Coauthors, 2016: Attribution of extreme weather and climate-related events. *Wiley Interdiscip. Rev.: Climate Change*, **7**, 23–41, <https://doi.org/10.1002/WCC.380>.
- , D. J. Karoly, and F. W. Zwiers, 2017: Is the choice of statistical paradigm critical in extreme event attribution studies? *Climatic Change*, **144**, 143–150, <https://doi.org/10.1007/s10584-017-2049-2>.