Unavoidable Errors

A Weather Warnings Perspective on Climate Attribution

Adapted from "Can the Issuance of Hazardous-Weather Warnings Inform the Attribution of Extreme Events to Climate Change?" by **Dale R. Durran** (University of Washington). Published online in *BAMS*, August 2020. For the full, citable article, see [DOI:10.1175/BAMS-D-20-0026.1](https://doi.org/10.1175/BAMS-D-20-0026.1).

The science of extreme-event attribution has grown rapidly in an effort to quantify the extent to which climate change is altering the frequency or severity of extreme events and to potentially guide efforts toward mitigation and adaptation.

A common null hypothesis in one approach to attribution—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 (a rejection of a null hypothesis that is true) occurs when an extreme event is produced by natural variability but is 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. We must weigh the likelihood and seriousness of type I and type II errors in formulating our assessment. It has been suggested that the risk-based approach tends to

understate anthropogenic contributions to extreme events (a tendency toward type II errors).

At present, the American public's perception of the risks from anthropogenic climate change 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 when—as is typical in attribution studies—the scientists express their conclusions in probabilistic terms, the takeaway message conveyed to the public by the media is often simplified to “yes” or “no.”

We will therefore focus on considerations related to binary yes–no decision making with the goal of highlighting connections between two problems in alerting the public about dramatic impacts on life and properly that are nominally very different: issuing hazardous-weather warnings and attributing extreme weather to climate change.

Scoring the Warnings Dilemma

Weather forecasters have long dealt with the tradeoffs between over- and under-warning. Consider three forecasters, Jane, Jennifer, and Jaclyn, who are deciding whether to issue tornado warnings during a period when 80 tornadoes actually occur. 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?

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

Two key measures of their forecast quality are the probability of detection (POD) and false alarm rate (FAR). POD is the ratio of the number of correctly warned events to all the actual hazardous events. FAR is the ratio of the number of warnings issued for events that did not occur to the total number of warnings.

Jane has the highest POD of 0.65, but she also has the highest FAR (0.74). 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 (0.69) relative to Jane by 0.05; but she also reduced her POD (0.50) 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? In the context of extreme-event attribution and climate change, 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, sounding the alarm at essentially every opportunity.

As a point of reference, the National Weather Service (NWS) FAR for tornado warnings remained near 0.75 from 1990–2011 before dipping slightly to 0.70 in the period 2012–2017. Roughly 3 out of every 4 tornado warnings are false alarms! In contrast, POD has risen from

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

▲ * 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. The criterion for an extreme remains fixed at 2σ above the control-climate mean, independent of the threshold set for attribution.

▲ * 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}/σ . The mean in the current climate is $\sigma/2$ warmer than μ_c . Areas *a* (giving the probability of correctly attributing the event to global warming) and *b* (giving the probability of incorrectly attributing the extreme temperature event to global warming) 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.

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.

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. Why did POD get worse after 2011? A high FAR was identified as a possible contributor to the high tornado death toll in 2011, and an emphasis on reducing the FAR gave rise to the degradation of POD. The large changes in POD relative to FAR in the NWS tornado

forecasts have been modeled using signal detection theory. Reducing FAR without significantly impacting POD would require major improvements in forecasting capabilities.

Jane's performance metrics are roughly similar to the NWS's performance over the period 2002–2011, whereas Jacyln's numbers are similar to those of the NWS between 2012 and 2017. 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.

Extreme-temperature attribution

Let us see how considerations about acceptable POD and FAR in tornado warnings might translate to climate-change attribution. We will consider attribution rubrics for which we evaluate the expected average values of POD and FAR by determining statistical distributions for the probability with which the event would occur, both in a world without anthropogenic warming and in the current climate. This is the same pair of statistical distributions that would be used in traditional

As in the previous figure, but that the rubric illustrated attributes all normalized temperatures greater than $\mu_c + 2.5\sigma$ to global warming. Under this new rubric, the probability that a 2σ extreme event is produced by climate change, but not correctly attributed to that change, is given by area *c*. *** ▼**

risk-based climate attribution studies.

Assume, for example, we have an accurate representation 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 two times in a 3-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.” The false alarm ratio associated with this rubric would be 0.34—far less than the FARs associated with NWS tornado warnings, while the POD would be a perfect 1.0. If we attempt to reduce the FAR by switching to a more restrictive rubric with an attribution threshold 2.5σ above the control climate’s mean, we miss some of the 2σ extremes and produce a decrease in the POD that is 9 times the decrease in the FAR!

Increasing the attribution threshold to 3σ above the control-climate mean produces a

Figure 2. POD as a function of FAR for maximum temperature attribution metrics, with thresholds at μ_c (the control climate normal) + $a\sigma$ 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 for tornado warnings from 2003 to 2011, and 2014 to 2017, respectively. The (POD, FAR) values for the cases illustrated in the previous figure are indicated by blue dots.

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 decision thresholds produces reductions in both the FAR and POD, with substantially larger reductions in the POD. Of course, we can shift the POD back to 1.0 if we move the goalposts by redefining the class of extreme events as those 3σ above μ_c (corresponding to one event in a given month every 25 years). One difficulty in focusing on such very rare events is that their expected frequency is highly sensitive to the specification of the tails in the probability distributions, and it can be difficult to determine the tails with confidence.

Tornadoes pose the most difficult severe-weather-warning challenge. FARs are lower, and the PODs are higher, for other hazards, yet FAR values for severe thunderstorms, flash floods, high winds, and winter storms, which range from 0.44 to 0.56, are still much higher than those associated with the climate-attribution rubrics considered here.

Discussion and conclusions

Using methods adapted from epidemiology, risk-based climate attribution studies have made effective use of metrics that are closely related to the false alarm ratio. 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 that extend beyond the false alarm ratio to include the probability of detection.

As we have seen, 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.

Climate scientists, who are faced with trade-offs reminiscent of those faced by forecasters issuing warnings for hazardous weather, may need to follow the practice of operational meteorologists and tolerate a significant false alarm ratio to properly alert the public. In contrast, an approach analogous to Jennifer's overly cautious tornado forecasting strategy would be 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 an 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. ♡

METADATA

BAMS: What would you like readers to learn from this article?

Dale Durran (University of Washington): *Whether forecasting severe weather or attributing extreme events to climate change, we strive to meet two goals. We want to be sure to alert for actual events and we want to minimize false alarms. There is inevitably a tradeoff between these two goals, and it can be very lopsided. Efforts to reduce false alarms are often accompanied by a much larger reduction in the fraction of events for which alerts are properly issued.*

BAMS: How did you become interested in the topic of this article?

DD: *I was reading a paper about climate-change attribution by Elisabeth Lloyd and Naomi Oreskes in which they commented that meteorologists have a history of dealing with the problem of false alarms. Digging into this, I found a 2004 paper in BAMS by Harold Brooks, applying signal detection theory to tornado warnings; that in turn served as a key departure point for my research.*

BAMS: What surprises/surprised you the most about what you found?

DD: *I was surprised by the simple mathematical relationship between the "false alarm ratio" in weather forecasting and the "fraction of attributable risk" in climate-change attribution. They are even both abbreviated FAR. Yet, at the same time, I am not aware of any common climate-change attribution metric related to weather forecasting's other key metric, the "probability of detection."*

BAMS: What was the biggest challenge you encountered while doing this work?

DD: *This was a new area of study for me, so it required a lot of reading and communication with scientists working on both the severe weather forecasting and climate-change problems.*

BAMS: How will you follow up?

DD: *I am interested in applying the rubric approach, illustrated in my paper, using idealized maximum temperature distributions, to real-world data.*

300 miles

>> How a silver lining forms

>> It starts at sea.
>> Tropical waters heat up.
>> Warm air soars skyward.
>> Cold air rushes to the void.
>> Cold air warms up.
>> Cycle repeats.
>> Faster and faster—a 50,000 foot engine of air.
>> At seventy four miles per hour it earns a name.
>> Harvey, Irma, Katrina.
>> Then landfall.
>> Roads rendered useless.
>> Buildings destroyed.
>> Families stranded.
>> But for a brief moment,
>> A silver lining appears.
>> People see neighbors instead of strangers.
>> And labels that divide are forgotten.

>> But when rains ease,
>> when clouds part,
>> silver linings need not fade.

>> Let's embrace our shared humanity.
>> Let's connect with one another.
>> Let's find our love for each other.
>> Every single day.

>> Come together at lovehasnolabels.com

29.7604° N, 95.3698°, W SFC 134