CLIMATOLOGY

Verification of extreme event attribution: Using out-of-sample observations to assess changes in probabilities of unprecedented events

Noah S. Diffenbaugh^{1,2}*

Independent verification of anthropogenic influence on specific extreme climate events remains elusive. This study presents a framework for such verification. This framework reveals that previously published results based on a 1961–2005 attribution period frequently underestimate the influence of global warming on the probability of unprecedented extremes during the 2006–2017 period. This underestimation is particularly pronounced for hot and wet events, with greater uncertainty for dry events. The underestimation is reflected in discrepancies between probabilities predicted during the attribution period and frequencies observed during the out-of-sample verification period. These discrepancies are most explained by increases in climate forcing between the attribution and verification periods, suggesting that 21st-century global warming has substantially increased the probability of unprecedented hot and wet events. Hence, the use of temporally lagged periods for attribution—and, more broadly, for extreme event probability quantification—can cause underestimation of historical impacts, and current and future risks.

INTRODUCTION

The field of extreme event attribution has burgeoned since the seminal work of Stott *et al.* (1). In that time, numerous event attribution frameworks have been developed (2). Although there is heterogeneity in the design of these frameworks, most use a combination of instrumental observations and climate model simulations to quantify the influence of historical anthropogenic climate forcing on the probability and/or severity of individual events. The purpose of this study is to examine whether independent "out-of-sample" observations can be used to assess the accuracy of changes in extreme event return intervals that are either explicitly or implicitly predicted by attribution frameworks.

Since Stott et al. (1), attribution analyses have been published for many types of events (2), including heatwaves [e.g., (3–8)], cold snaps [e.g., (3, 5, 9)], heavy rainfall [e.g., (3–6, 10)], floods [e.g., (11)], droughts [e.g., (12)], tropical cyclone precipitation [e.g., (13, 14)], storm surge flooding [e.g., (15)], and extremely low Arctic sea ice [e.g., (4, 16)]. In addition, event attribution frameworks have been applied to the underlying physical causes of extremes (2, 17, 18), including atmospheric circulation patterns [e.g., (4, 19–22)], atmospheric water vapor (4), ocean heat content (23), and wildfire risk factors [e.g., (24)]. In recent years, attribution analyses have been applied increasingly quickly following an event [e.g., (10, 25)], with some techniques using forecasts generated before the event [e.g., (26, 27)]. "Precomputed" approaches (7) have likewise been used to quantify the influence of global warming on a particular type of event at each area of the globe, using observational data (4, 28), climate model simulations (6, 7), or a combination of the two (4, 5).

Independent verification of event attribution poses a particular challenge. In addition to the reliability of observational data and climate model simulations [e.g., (29, 30)], there are fundamental questions about the appropriate scientific framing through which causation

Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

can be measured [e.g., (2, 31–34)]. One inherent challenge is that single event attribution is conducted for conditions at one specific place and time; the event only occurs once, and by construction, the attribution quantification pertains only to that event. Further, because extreme events are by definition rare, the available population of events with which to independently verify attribution results is limited, a challenge that is exacerbated for events that are unprecedented in the observational record. One approach to resolving these challenges is to frame the attri-

One approach to resolving these challenges is to frame the attribution result as a falsifiable prediction, and then test that prediction using independent observations. Such an approach draws on the many aspects of climate and weather research that routinely use independent verification. For example, daily- and seasonal-scale forecasts are verified after the forecast period has passed [e.g., (35)]. This forecast verification includes daily fields such as temperature, precipitation, and winds, as well as extreme event phenomena such as tropical cyclones, severe thunderstorms, and river and storm surge flooding. Further, scientists have been making long-term climate projections for decades (36, 37). Older projections can be verified using current observations [e.g., (38)], and such comparisons are now made for global temperature anomalies in quasi-real time.

It is important to emphasize the distinction between verification of falsifiable predictions and evaluation of methodological uncertainty. Researchers have for years taken great care to thoroughly evaluate various aspects of uncertainty within climate attribution systems (2). This includes (i) assessing the robustness of the observational record and the fidelity of climate model simulations for different types of events [e.g., (2-4, 30)]; (ii) quantifying uncertainty in the climate model simulations, including the sensitivity to historical emissions (4), the ability to simulate the statistical properties of the historical observations [e.g., (4, 21, 39, 40)], and the ability to simulate the underlying physical processes that cause different types of events [e.g., (21, 41)]; (iii) quantifying uncertainty in the statistical analysis, including the appropriateness of the underlying statistical assumptions (4, 42–45); and (iv) applying different attribution methodologies to the same event (4, 16, 46, 47), including systematic

¹Department of Earth System Science, Stanford University, Stanford, CA, USA. ²Woods Institute for the Environment, Stanford University, Stanford, CA, USA. *Corresponding author. Email: diffenbaugh@stanford.edu

reanalysis of multiple published results (3, 33). However, despite this emphasis on uncertainty quantification, independent observational verification of specific, quantitative attribution results remains elusive.

Central to the analysis presented in this study is the idea that attribution results that are generated from estimates of return intervals in previous historical time periods can be verified using the frequency of extreme events that occur over large geographic domains during subsequent, multi-year, out-of-sample time periods (see Materials and Methods). For example, many attribution analyses have used global climate model simulations from the Coupled Model Intercomparison Project (CMIP5) [e.g., (4-8, 20, 38)]. Because the CMIP5 Historical and Natural simulations were only run through 2005 (48, 49), simulations using the actual climate forcings do not cover the most recent period of observations. Attribution analyses that use CMIP5 can thus either restrict the historical analyses to this pre-2006 period [e.g., (4-6, 12, 20)] or use the early period of the CMIP5 future projections to extend the historical simulations (in which case the anthropogenic and non-anthropogenic simulations cover different time periods) [e.g., (8, 38)]. In the case of previously published global attribution analyses, which used the CMIP5 Historical and Natural simulations to quantify the influence of historical forcing on the probability of unprecedented hot, wet, and dry extremes at each area of the globe, the attribution analysis was limited to the pre-2006 period (5). However, this limitation also presents an opportunity, because the frequency of record-setting events during 2006-2017 can now be used to independently verify the published results that used data from 1961 to 2005.

Previous global attribution analyses (4) examined four different attribution metrics: (i) the contribution of the observed trend to the event magnitude, (ii) the contribution of the observed trend to the event probability, (iii) the probability of the observed trend in the historical forcing, and (iv) the contribution of the historical forcing to the event probability. This work was recently extended (5), using CMIP5 data to quantify the fourth metric for natural and anthropogenic forcing during the historical period, and for future levels of forcing consistent with the United Nations Paris Agreement goals and commitments.

The current study focuses on verifying the second and fourth metrics using out-of-sample observations. The contribution of historical climate change to the event probability is measured using an "attribution ratio" (*AR*), which is calculated as the ratio between the return interval in a counterfactual world without climate change and the return interval in the actual observed world with climate change (4, 5). For the contribution of the observed trend to the event probability, observational data are used to estimate the return intervals of extreme events, with the attribution ratio (*AR*_{Obs-dt}) calculated from the return interval in the actual time series (*RI*_{Obs}) and the return interval in the detrended time series (*RI*_{Obs-dt})

$$AR_{Obs-dt} = (RI_{Obs-dt}) \div (RI_{Obs})$$

For the contribution of the historical forcing to the event probability, observational data are used to correct systematic biases in the climate model simulations, which are then used to estimate the change in return intervals under historical (HIST) and natural (NAT) climate forcing

$$AR_{\text{Forcing}} = (RI_{\text{Obs-dt}}) \div (RI_{(\text{HIST-NAT}) + \text{Obs-dt}})$$

An attribution ratio of 1 indicates equal probability with and without global warming. Because return intervals are the inverse of event probabilities, larger ratios indicate greater influence of global warming (e.g., a ratio of 2 indicates that the probability of an event is twice as large with global warming). Block bootstrapping of the time series at each location is used to quantify a distribution describing the uncertainty in the event probabilities at each location (4, 5).

The present study is focused on two objectives. The first phase of the analysis uses specific, previously published predictions to demonstrate the framework for verifying extreme event attribution results. Independent data (i.e., observations over the 2006–2017 time period) are used to derive the return intervals of unprecedented events over different regions, based on the regional frequency of record-setting events. These out-of-sample return intervals are then compared with the regional-mean distributions of return intervals (e.g., 5th, 25th, 50th, 75th, and 95th percentiles) that were predicted from the detrended 1961–2005 observational data at each grid point in the region. The ratio is referred to as a "verification ratio" (VR)

 $VR_{\text{Obs:}2006-2017} = RI_{\text{Obs-dt:}1961-2005} \div RI_{\text{Obs:}2006-2017}$

where $RI_{Obs-dt:1961-2005}$ is the regional-mean of the return intervals in the detrended 1961–2005 time series at each grid point, and $RI_{Obs:2006-2017}$ is the regional-mean return interval implied by the frequency of record-setting events in the region during the out-ofsample 2006–2017 verification period. These verification ratios are compared with attribution ratios that quantify the contribution of historical climate change during the 1961–2005 attribution period, calculated from both the observational record (AR_{Obs-dt}) and the CMIP5 global climate model ensemble ($AR_{Forcing}$). Thus, by construction, the out-of-sample comparison tests the stability of the attribution results over time, within the context of a nonstationary climate.

The second phase of the analysis attempts to understand discrepancies between the verification and attribution ratios. This analysis tests whether any such discrepancies are due to structural mismatches between the attribution and verification methods. It also tests whether there have been changes in the frequency of record-setting events between the attribution and verification periods, and whether any changes are due primarily to external climate forcing or to internal climate variability. Understanding discrepancies in the predicted probabilities of record-setting events and the actual out-of-sample occurrence is important not only for verifying extreme event attribution but also for evaluating the durability of design and planning guidelines that use similar return interval quantification when conducting risk analysis (such as for infrastructure design, land use planning, and disaster management).

In principle, this verification framework could be applied to any type of extreme event. The focus of this initial application is on events that are unprecedented in the baseline historical period (1961–2005). Unprecedented events pose important challenges for event attribution (4). First, statistical uncertainty increases as values reach further into the tails of the distribution. Events that fall outside of the historical range are, by definition, in the extreme tail, amplifying the challenges posed by small samples. Second, climate change is increasing the probability of unprecedented events (4). Quantifying the effects of this nonstationarity is a general challenge for risk assessment [e.g., (50, 51)] and poses specific challenges for event attribution (4). Third, climate models are the only available tool for systematically testing the influence of global warming on the physical processes that

shape extremes, making climate models a necessary component of event attribution frameworks (2). However, because historically unprecedented events often arise from rare combinations of physical ingredients, they generally pose the greatest challenge for accurate climate model simulation (2, 17, 18, 30).

Despite these potential barriers, events that fall outside of the historical experience are critical for a suite of design and management decisions [e.g., (50, 52-54)], as well as climate change mitigation and adaptation considerations [e.g., (4, 5, 54, 55)]. Given both the societal relevance and methodological challenges, this initial verification study focuses on the attribution of events that are unprecedented in the historical observations.

RESULTS

The regional verification ratios for 2006–2017 frequently exceed the published attribution ratios calculated from the 1961–2005 data (Fig. 1), suggesting that the attribution framework underestimates the influence of historical global warming. For example, for the influence of anthropogenic forcing, the median attribution ratio is less than 2.0 for all three extreme indices (hottest days, wettest days, and longest dry spell) over the United States, Europe, and East Asia. In contrast, the median verification ratio for the hottest days exceeds 4.0 over Europe and 2.5 over East Asia, with >95% of the verification ratio distribution exceeding the median attribution ratio. Likewise, the median verification ratio for the wettest days exceeds 3.0 over the United States and Europe, with >95% of the verification ratio distribution again exceeding the median attribution ratio.

Although the trend-based attribution ratio is generally larger than the forcing-based attribution ratio (Fig. 1), the verification ratio for 2006–2017 still frequently exceeds the trend-based attribution ratio (Fig. 1). For example, for the hottest days, >95% of the verification ratio distribution exceeds the median trend-based attribution ratio over Europe, and ~75% exceeds the median trend-based attribution ratio over East Asia. Similarly, for the wettest days, >95% of the verification ratio distribution exceeds the median trend-based attribution ratio over the United States and Europe.

In a number of cases, the median values of both the attribution and verification ratios are close to 1.0 (Fig. 1). For the hottest days, both the forcing- and trend-based attribution ratios exhibit median values just above 1.0 over the United States, while the median verification ratio is just below 1.0. Likewise, for the longest dry spells, the attribution and verification ratios are near 1.0 over the United States, Europe, and East Asia. In these cases, the range of values is larger for the attribution ratios than for the verification ratios, including greater likelihood of large increases in extreme event probability. However, the attribution and verification distributions largely overlap.

The discrepancies between the attribution and verification ratios for record-setting events (Fig. 1) are reflected in discrepancies between the probabilities predicted from the 1961–2005 observations and the frequencies observed in 2006–2017. For example, the 2006–2017 frequency of record-setting hottest days exceeds the 99th percentile of predicted probabilities over both Europe and East Asia (Fig. 2). Similarly, the 2006–2017 frequency of record-setting wettest days exceeds the 99th percentile of predicted probabilities over both the United States and Europe (Fig. 3). Further, in cases where the discrepancies between the verification and attribution ratios are less pronounced, such as the hottest days over the United States and wettest



Α



Downloaded from http://advances.sciencemag.org/ on March 22, 2020

Fig. 1. Verification of the anthropogenic influence on unprecedented hot, wet, and dry events. The verification framework is based on the probability, during the out-of-sample verification period (2006–2017), of exceeding the most extreme value found in the period for which the attribution metrics were calculated (1961–2005). The framework is used to verify the attribution metrics published in (4) and (5), for (A) hottest day of the year (TXx), (B) percentage of annual precipitation falling in days that are wetter than the 95th percentile of the 1961–1990 period (R95p), and (C) longest consecutive dry spell of the year (CDD). Maps show the median attribution ratio calculated from the 1961–2005 trend at each northern hemisphere grid point for which there are continuous data in the CLIMDEX dataset (see Materials and Methods). The blue distribution shows the uncertainty in the attribution ratio calculated from the 1961–2005 trend (i.e., the metric shown in the map) over the United States, Europe, and East Asia. The purple distribution shows the uncertainty in the regional attribution ratio calculated from anthropogenic climate forcing. The red distribution shows the uncertainty in the regional verification ratio calculated from the 2006–2017 observations. Uncertainty in each ratio is depicted by the 5th, 25th, 50th, 75th, and 95th percentile values of the bootstrapping described in (4) and (5).

days over East Asia (Fig. 1), the 2006–2017 frequency still falls in the tail of predicted probabilities (Figs. 2 and 3).

There are at least two possible explanations for these discrepancies between the probabilities predicted during the attribution period (1961–2005) and the frequencies observed during the verification period (2006–2017). The first possibility is a structural discrepancy in the comparison, such as if the regional-mean of the probabilities calculated from the 1961–2005 grid-point time series did not accurately predict the regional frequencies during an overlapping time period. A second possibility is that there have been changes in the probabilities of record-setting events between the attribution and verification periods.



Fig. 2. Observed and simulated regional extreme event frequencies for the hottest day of the year (TXx). (A) The map shows the difference in the mean value between the out-of-sample verification period (2006-2017) and the period for which the attribution metrics were calculated (1961–2005). (B) The red line shows, for each year of the 2006–2017 verification period, the observed northern hemisphere frequency of events in which the grid-point value exceeded the maximum gridpoint value during the period for which the attribution metrics were calculated (1961–2005). The blue distribution shows the uncertainty in the hemispheric mean probability of exceeding the most extreme value found in the period for which the attribution metrics were calculated (1961–2005). The probability of the record-setting event is calculated by fitting an extreme value distribution to the 1961-2005 time series at each grid point, as described in (4); uncertainty is depicted by the percentile values of the bootstrapping described in (4). The blue circles show the regional frequency simulated by the CMIP5 climate model ensemble during the IPCC's baseline period (1986-2005). The red circles show the regional frequency simulated by the CMIP5 climate model ensemble during the verification period (2006-2017). (C) The blue distribution shows the uncertainty in the regional-mean probability of exceeding the most extreme value found in the period for which the attribution metrics were calculated (1961-2005). The blue horizontal line shows the observed regional frequency during the IPCC's baseline period (1986-2005); blue circles show the regional frequency simulated by the CMIP5 climate model ensemble during the IPCC's baseline period. The red horizontal line shows the observed regional frequency during the out-of-sample verification period (2006-2017); red circles show the regional frequency simulated by the CMIP5 climate model ensemble during the verification period.

The results favor the second possibility. For example, the actual regional frequencies that occurred during the Intergovernmental Panel on Climate Change (IPCC's) baseline period (1986-2005) all fall within the 5th to 95th percentile uncertainty range predicted from the 1961-2005 observations, and the majority fall within the 25th to 75th percentile uncertainty range (Figs. 2 to 4). Further, the CMIP5 climate model ensemble, which is an independent dataset with which to predict the frequency of record-setting events at a given level of climate forcing, exhibits close overlap with the predicted probabilities and the observed 1986-2005 regional frequencies (Figs. 2 to 4). Even in the cases where the 1986-2005 CMIP5 ensemble spread is furthest from the median of the predicted probabilities (such as the longest dry spells over the United States, Europe, and East Asia), the ensemble range still falls within the distribution of predicted probabilities (Fig. 4). The fact that the observed and simulated 1986-2005 frequencies fall well within the distributions of probabilities predicted from the 1961-2005 observations (Figs. 2 to 4) suggests that discrepancies between the attribution and verification ratios (Fig. 1) are not caused by structural discrepancies between the underlying metrics.

In contrast, there are substantial differences in the observed frequency of record-setting events between 1986–2005 and 2006–2017. For example, the observed frequency is at least ~50% higher in 2006–2017 for hottest days over Europe and East Asia (Fig. 2), wettest days over the United States and Europe (Fig. 3), and longest dry spells over East Asia (Fig. 4). Likewise, with the exception of the longest dry spells over the United States and East Asia (Fig. 4), the frequency observed during 2006–2017 falls further from the median predicted probability, while the frequency observed during 1986–2005 falls closer to the median (Figs. 2 to 4). These comparisons quantify a substantial increase in the risk of unprecedented events between the attribution and verification periods, particularly for hot and wet events.



Fig. 3. Observed and simulated regional extreme event frequencies for the wettest days. As in Fig. 2, but for the percentage of annual precipitation falling in days that are wetter than the 95th percentile of the 1961–1990 baseline period (R95p).



Fig. 4. Observed and simulated regional extreme event frequencies for the longest dry spell. As in Fig. 2, but for the longest consecutive dry spell of the year (CDD).

One concern about this analysis is that the verification period is relatively short (12 years) compared to a standard climatological baseline period (nominally 30 years). To test the robustness of the results to a longer period, the verification period can be extended to include the period from the beginning of the IPCC baseline (1986) to the end of the out-of-sample verification period (2017). As would be expected, mixing the out-of-sample verification period (2006–2017) with the end of the attribution period (1961–2005) to form an extended verification period (1986–2017) yields verification results that generally fall between the original attribution results and the out-ofsample verification results (tables S1 to S3). However, in a number of cases, the verification results for this modified period still exceed the original attribution results, including hot events over Europe (table S1) and wet events over the United States and Europe (table S3).

By generating multiple realizations of the climate system within a given level of forcing, the CMIP5 simulations can also provide an independent evaluation of whether the change in frequency of recordsetting events is due primarily to climate variability, or has instead been influenced by the increase in climate forcing between the attribution and verification periods. For the hottest days over Europe and East Asia (Fig. 2) and the wettest days over the United States, Europe, and East Asia (Fig. 3), both the observations and the CMIP5 ensemble exhibit higher probability of record-breaking events in 2006-2017 than in 1986-2005. Likewise, for the hottest days over Europe and East Asia (Fig. 2), wettest days over the United States, Europe, and East Asia (Fig. 3), and longest dry spells over the United States and East Asia (Fig. 4), the frequency of record-breaking events observed in 2006-2017 has a higher likelihood of occurring in 2006-2017 of CMIP5 than in 1986–2005 of CMIP5. These patterns are also true at the scale of the northern hemisphere for both the hottest and wettest days, where the CMIP5 ensemble exhibits higher frequency

of record-setting events in 2006–2017 than in 1986–2005, and the observed 2006–2017 frequencies have a higher likelihood of occurring in 2006–2017 of CMIP5 than in 1986–2005 of CMIP5 (Figs. 2 and 3). The fact that the frequency of record-setting hot and wet events observed during the 2006–2017 verification period generally falls within the CMIP5 ensemble spread for 2006–2017 and generally outside the CMIP5 ensemble spread for 1986–2005 suggests that the observed increase in occurrence was likely influenced by the increase in forcing between the attribution and verification periods.

In contrast, the verification of record-setting longest dry spells suggests that, at both the regional and hemispheric scales, global warming has not had a clear influence on the probability of record-setting events. This lack of attribution was already suggested by the high fraction of attribution ratios near 1.0 (Fig. 1) (5). The fact that the verification ratios are also clustered near 1.0 (Fig. 1) strengthens that conclusion. Further, the close overlap between the observed and simulated frequencies for 1986–2005 and 2006–2017 (Fig. 4) suggests that, in contrast to hot and wet events (Figs. 2 and 3), the recent increase in climate forcing has not altered the probability of record-setting longest dry spells over the analysis regions. However, it is important to note that other areas of the globe may have experienced verifiable increases in the probability and/or intensity of dry spells [e.g., (4)].

DISCUSSION

The fact that the verification framework reveals the published global attribution results to be overly conservative for hot and wet events carries a number of implications. For example, those attribution results suggested that global warming had already influenced the magnitude and probability of unprecedented events at large fractions of the globe, including >80% for hot events and >50% for wet events (4). This includes 71% of North America, 77% of Europe, and 56% of East Asia for the record hottest day of the year, and 80% of North America, 89% of Europe, and 70% of East Asia for the record percentage of annual precipitation falling in the wettest days (5)). The verification results presented here suggest that the influence of global warming on these events has been even more pervasive than suggested by those original attribution results.

Likewise, because many of the impacts of global warming are felt through extremes (54), attribution of the influence of global warming on record-setting events is highly relevant for quantifying the impacts of historical anthropogenic climate forcing on natural and human systems. In revealing previously published attribution results to be largely conservative, the verification results suggest that the impacts of global warming have been even larger than originally implied (4, 5). Further, attribution quantification is now being used to assign specific responsibility for the damages resulting from individual events (55). The results presented here highlight the importance of independent verification of the attribution frameworks that are used to assign responsibility for damages.

The underestimation of the probability of record hot and wet events during the verification period implies a rapid intensification of extreme event probability—and therefore risk—resulting from relatively small increases in climate forcing. This intensification has important implications both for extreme event attribution and for accurately quantifying probabilities of extreme values in the current and near-term climate. Although the calculation of record-setting probabilities attempts to account for nonstationarity in the observational time series (4), the verification results suggest that even one to two additional decades of global-scale climate forcing can lead to substantial underestimation of the probability of record-setting hot and wet events (Figs. 2 and 3).

The fact that the observed and simulated frequencies of recordsetting events exhibit such large nonstationarities between the baseline period (ending in 2005) and the verification period (2006–2017) suggests that extreme event attribution assessments—as well as other risk assessments-should take particular care to use techniques that capture conditions in the current time period. Researchers have used a number of approaches to extend the period of the attribution analysis. For metrics that rely only on observational data, researchers have used the period of available data at the time of the event [e.g., (4, 10, 28, 46)]. Other researchers have calculated statistical relationships between the event probability and the global mean temperature (10, 13, 14). For metrics that rely on climate model simulations (including coordinated archived experiments such as CMIP5), researchers have used climate model projections to extend the period of analysis up to the time of the event [e.g., (38)] or to generate attribution results for different levels of global warming (including projected future levels) [e.g., (5, 6, 8, 13)]. For the attribution results evaluated here, the original study (5) included projections of return interval ratios for 2016-2035 and 2036-2055 in the CMIP5 RCP8.5 experiment, enabling comparisons with 1° to 2°C and 2° to 3°C of global warming.

Extending probability predictions under higher levels of global warming has been less common in other applications that rely on extreme event probability quantification, such as infrastructure design and risk assessment [e.g., (52)]. The verification results suggest that those applications could benefit from such approaches, particularly given that those planning decisions are more explicitly futureoriented than attribution analysis. For example, the underprediction of occurrence of record-setting events during the out-of-sample verification period provides evidence in support of dynamic design guidelines that can be updated as new observational data become available [e.g., (50, 52–54)]. Likewise, the fact that the CMIP5 projections for 2006–2017 most accurately capture the actual 2006–2017 frequency of record-setting hot and wet events (Figs. 2 and 3) suggests that ensemble climate model projections could be used to improve probability quantification for applications that have traditionally relied solely on historical observations.

In addition to capturing the response of extreme events to increasing climate forcing, ensemble climate model projections can also help to quantify the influence of variability on future extreme event probabilities. For example, the 1961–2005 attribution metrics suggest >50% likelihood that global warming has increased the probability of record-setting hottest days over the United States (Fig. 1). Further, comparison of the CMIP5 simulations for 2006–2017 and 1986–2005 predicts very high likelihood of a substantial increase in the frequency of record-setting hot events in the later period (Fig. 2). However, 75% of the verification ratio distribution is less than 1.0 over the United States (Fig. 1), driven by a 2006–2017 frequency that is in the lowest quartile predicted from the 1961–2005 observations (Fig. 2).

This relatively low frequency of record-setting hottest days over the United States is consistent with the well-documented "warming hole," a pattern of reduced warming over the central and southeastern United States that has been attributed alternatively to atmospheresoil moisture feedbacks (56), the aerosol-indirect effect (57), and

global warming are projected to cause substantial warming throughout North America, the lower rates of warming associated with the warming hole are projected to persist over the near-term decades, with relatively high summer temperature variability over the central and southeastern United States persisting throughout the 21st century (59). Although there is some indication that the mechanisms causing the warming hole may have reversed early in the 21st century (58), the pattern of reduced warming over the central and southeastern United States is present in the mean hottest day of the year for 2006-2017 relative to 1961-2005 (including negative anomalies over the central United States; Fig. 2). Notably, although the observed frequency of record-setting hottest days is lower over the United States in 2006-2017 compared to 1986-2005 (Fig. 2), the 2006-2017 frequency does overlap with the lowest CMIP5 value, highlighting the importance of climate variability within the context of increasing forcing.

internal ocean-atmosphere variability (58). Although high levels of

CONCLUSIONS

The motivation for this study is to introduce and demonstrate a framework for independent verification of extreme event attribution results. The field of extreme event attribution has expanded rapidly in the past two decades. Results are now the subject of frequent public interest (2). This interest has extended into various public decision-making processes, both as motivation for incorporating climate change into decisions [e.g., (52)] and as a basis for assigning responsibility for damages (55). The use of attribution results raises the burden for scientists to independently verify those results, particularly for events that are unprecedented in the historical experience (and therefore pose the most acute risks).

Numerous methods for event attribution have been developed (2). Although different dimensions of methodological uncertainty have been thoroughly evaluated, and in some cases the results of different methods have been systematically intercompared, extreme event attribution results have not yet been independently verified within a framework of scientific falsifiability. To fulfill that need, this study presents a framework for using the attribution calculation to create falsifiable predictions of the frequency of record-setting events and then uses out-of-sample observations to test those predictions. As an initial proof of concept, the verification framework is applied to previously published attribution results for record-setting hot, wet, and dry events at different areas of the globe (4, 5).

Independent verification suggests that those published attribution results frequently underestimate the influence of global warming on the probability of unprecedented hot and wet extremes, with greater uncertainty for dry extremes. The discrepancy between the attribution and verification ratios can be most explained by the increase in climate forcing since the end of the period in which the attribution ratios were generated. This is particularly true for hot events and wet events, for which the discrepancies between the attribution and verification ratios are greatest. Overall, the verification results suggest not only that historical global warming has increased the probability of unprecedented hot and wet events over the northern hemisphere but also that the magnitude of this effect has increased during the 21st century.

Although this study focuses on record-setting hot, wet, and dry events over land areas of the northern hemisphere, the verification framework could also be applied to a suite of other extreme climate variables [e.g., (49)] and physical ingredients [e.g., (4)], with different data sources providing coverage for different areas of the globe. Further development and application of this and other frameworks will provide a more comprehensive verification of the magnitude of anthropogenic influence on different types of extreme events in different regions of the world.

The verification of previously published results from one attribution method does offer some generalizable lessons. The first is that although many attribution analyses have leveraged the unique insights available from multi-institution climate model archives such as CMIP5 [e.g., (4-8, 20, 38)], such "ensembles of opportunity" also present limitations. For example, because the coordinated experiments require multiple years to plan and run, the simulations that use historical forcings do not extend to the present at the time that a new event occurs (48). This means that analyses must either cover historical periods that do not extend to the present [e.g., (4-6, 12, 20)] (which, as this study shows, results in an underestimation of the influence of global warming on hot and wet events) or use approaches to extend the calculation past the period of the historical simulations [e.g., (8, 10, 14, 38)]. The commonly implemented approach of using the early period of climate model projections to extend the calculation still presents limitations, both because researchers must compare the extended simulations with counterfactual simulations that do not reach up to the present [e.g., (8, 38)] and because the early period of the climate model projections does not include the actual forcings that occurred, which can hamper accurate attribution (60).

Another generalizable conclusion is that although precomputed approaches remove bias in the selection of events that are studied and enable unified analysis of multiple types of events across multiple regions of the world, the fact that the precalculation necessarily limits the analysis to an earlier baseline period likely leads to an underestimation of current probabilities. As a result, other precomputed calculations [e.g., (6, 7)] are likely also subject to a similar underestimation of the influence of historical forcing on the probability of events in the current climate. The verification results presented in this study highlight the importance for precomputed event attribution analyses to include calculations for higher levels of forcing [e.g., (5, 6, 8)] and to update the precomputed results as new observations become available. These results also suggest that "rapid" attribution approaches [which produce analyses soon after a specific event has occurred; e.g., (10, 14, 25)] should likewise continue to use methods that align the climate forcing in the attribution analysis with the forcing at the time of the event. Efforts to develop and deploy "operational" attribution systems [e.g., (27)] that update observations and simulations in real time will also help to address this limitation.

Last, the verification results have general implications beyond extreme event attribution. Historical climate observations are widely used as the basis for risk management decisions in areas as diverse as land use, infrastructure, water resources, supply chain management, disaster relief, finance, insurance, and liability. In many of these cases, decisions must be robust to both current and future probabilities of extreme events. Although decision-makers have been aware of the challenges posed by climate nonstationarity for a number of years [e.g., (50, 51)], many of these decisions still rely primarily on historical observations for calculating extreme event probability [e.g., (52)]. The methods for calculating those probabilities from historical data are closely linked to the methods used in the attribution framework evaluated here (4, 5). The out-of-sample verification results presented in this study thus highlight the importance of incorporating present and future nonstationarity into the extreme event probability quantification that underlies a broad suite of climatesensitive risk management decisions.

MATERIALS AND METHODS Data

The analysis uses data from the CLIMDEX project, which has archived observational and climate model values for multiple extreme climate indices (49). The observational values are calculated from station observations and gridded to a global grid, based on data continuity criteria. The climate model values are calculated from the CMIP5 climate model experiments (48).

The current study uses the observational data, along with the Historical and Natural climate model simulations. The Historical simulations include both natural forcings (such as volcanic aerosols and variations in solar output) and anthropogenic forcings (such as greenhouse gases and aerosols); the Natural simulations include only the natural forcings. The Historical and Natural simulations were run through the year 2005 (48). Comparison of the Historical and Natural simulations thus quantifies the influence of anthropogenic forcings during the historical climate period through 2005.

Attribution metrics

This study evaluates the extreme event attribution analyses that were published by Diffenbaugh *et al.* (5). The study focuses on three of the CLIMDEX indices included in that analysis, which together measure hot, wet, and dry events: the hottest day of the year (TXx; "hottest day"), the percentage of annual precipitation falling in days that are wetter than the 95th percentile of the 1961–1990 period (R95p; "wettest days"), and the longest consecutive dry spell of the year (CDD; "longest dry spell").

Diffenbaugh *et al.* (5) calculated the attribution ratio described in (4), using the CMIP5 Historical and Natural simulations over the 1961–2005 period. This attribution ratio ($AR_{\text{Forcing}:1961-2005}$) quantifies the influence of anthropogenic forcing on the probability of exceeding the most extreme value observed at each grid point during the 1961–2005 period. The metric is calculated as the ratio between the return interval of the observed record value in the lower level of forcing ($RI_{\text{NAT}:1961-2005}$) and the return interval of the observed record value in the higher level of forcing ($RI_{\text{HIST}:1961-2005$). For example, if the most extreme observed value has a return interval of 100 years in the Natural forcing (probability = 0.01) and a return interval of 50 years in the Historical forcing (probability = 0.02), then the attribution ratio ($AR_{\text{Forcing}:1961-2005}$) is 2, suggesting that anthropogenic forcing has doubled the probability of exceeding the most extreme observed value.

Diffenbaugh *et al.* (4) also calculated the contribution of the historical trend at each grid point to the probability of exceeding that grid point's most extreme observed value. This metric ($AR_{Obs-dt:1961-2005}$) is calculated as the ratio of the return interval of the observed record value in the detrended historical time series ($RI_{Obs-dt:1961-2005}$) and the return interval of the observed record value in the actual historical time series ($RI_{Obs:1961-2005}$)

```
AR_{\text{Obs-dt}:1961-2005} = (RI_{\text{Obs-dt}:1961-2005}) \div (RI_{\text{Obs}:1961-2005})
```

This second metric ($AR_{Obs-dt:1961-2005}$) thus relies only on observational data (without any climate model simulations) and is agnostic about the cause of the historical trend.

The current study evaluates both the attribution ratio due to anthropogenic forcing ($AR_{\text{Forcing:1961-2005}}$) and the attribution ratio due to the observed trend ($AR_{\text{Obs-dt:1961-2005}}$). Both attribution metrics report an uncertainty distribution of attribution ratios. These distributions are based on the uncertainty distribution of return intervals for the record setting event ($RI_{\text{Obs:1961-2005}}$), which are calculated from the observational time series using a block bootstrapping approach.

Verification framework

To verify the previously published attribution ratios, the uncertainty distributions calculated for 1961–2005 are compared with the frequency of occurrence of record-setting events observed during 2006–2017. This verification approach is conceptually similar to the attribution calculation of Coumou *et al.* (28), except here the verification data are kept out of sample (i.e., the verification data are not used in the calculation of the counterfactual time series from which the counterfactual probabilities are quantified).

First, the maximum value of each climate index is calculated at each grid point during the 1961–2005 period of the CLIMDEX observations. Then, for each grid point, all events during 2006–2017 that exceed the respective 1961–2005 grid-point maximum are identified. The frequency of occurrence of record-setting events in 2006–2017 ($F_{Obs:2006-2017}$) is then calculated over the Northern Hemisphere, the United States (30–50°N, 120–60°W), Europe (30–60°N, 0–50°E), and East Asia (20–45°N, 90–135°E), where

 $F_{\rm Obs:2006-2017} =$

[the total number of exceedances in the region in 2006-2017] \div [(the number of grid points in the region) \times (the number of years in 2006-2017)]

This regional frequency of occurrence ($F_{Obs:2006-2017}$) is then converted to a regional verification ratio ($VR_{Obs:2006-2017}$) that can be compared with the attribution ratios described in (5) and (4). First, the regional frequency of occurrence is converted to a "regional return interval" ($RI_{Obs:2006-2017}$) using the formula for the return interval

$$RI = 1 \div (1 - P)$$

but using the regional frequency of occurrence ($F_{\text{Obs:}2006-2017}$) as the measure of probability

$$RI_{\text{Obs}:2006-2017} = 1 \div (F_{\text{Obs}:2006-2017})$$

The regional-mean return interval of the observed record value in the detrended historical time series ($RI_{Obs-dt:1961-2005}$) is then computed by first calculating the mean of the grid-point probabilities in the detrended time series ($P_{Obs-dt:1961-2005}[i,j]$) and then calculating the regional-mean return interval from that regional-mean probability. (Note that the order of operations matters: It is important to first calculate the regional-mean of the grid-point probabilities to avoid the regional-mean return interval being dominated by any single grid-point return interval value.) The uncertainty in the regionalmean return interval ($RI_{Obs-dt:1961-2005}$) is quantified by calculating

Diffenbaugh, Sci. Adv. 2020; 6 : eaay2368 18 March 2020

the regional-mean at each quantile of the uncertainty distribution of grid-point probabilities ($P_{\text{Obs-dt:1961-2005}}[i,j]$).

Last, the uncertainty distribution of regional-mean return intervals in the detrended 1961–2005 time series ($RI_{Obs-dt:1961-2005}$) is divided by the regional-mean 2006–2017 return interval ($RI_{Obs:2006-2017}$), generating an uncertainty distribution of verification ratios ($VR_{Obs:2006-2017}$) for each region

$$VR_{\text{Obs:}2006-2017} = RI_{\text{Obs-dt:}1961-2005} \div RI_{\text{Obs:}2006-2017}$$

This distribution of verification ratios ($VR_{Obs:2006-2017}$) is compared with the regional-means of the grid-point distributions of attribution ratios from anthropogenic forcing ($AR_{Forcing:1961-2005}$) and attribution ratios from the observed trend ($AR_{Obs-dt:1961-2005}$).

To understand the comparisons between the published attribution ratios and the regional verification ratios, a number of regional extreme event frequencies are calculated using the IPCC's baseline period (1986–2005). These include the regional frequency of events that exceed the observed 1961-2005 maximum during the 1986–2005 period of the observations ($F_{\text{Obs:}1986-2005}$), the regional frequency of events that exceed the simulated 1961-2005 maximum during the 1986-2005 period of the CMIP5 Historical simulations $(F_{\text{CMIP5:1986-2005}})$, and the regional frequency of events that exceed the simulated 1961-2005 maximum during the 2006-2017 period of the CMIP5 RCP8.5 simulations (F_{CMIP5:2006-2017}). For each observed or simulated climate realization, the regional frequency is calculated as the number of times during the evaluation period (1986-2005 or 2006-2017) that a grid-point value within the region exceeds the respective 1961–2005 grid-point maximum, divided by the number of grid points in the region, divided by the number of years in the evaluation period.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/ content/full/6/12/eaay2368/DC1

Table S1. Verification metrics for the hottest day of the year (TXx), calculated for different time periods.

Table S2. Verification metrics for the percent of precipitation from wettest days (R95p), calculated for different time periods.

Table S3. Verification metrics for the longest dry spell of the year (CDD), calculated for different time periods.

REFERENCES AND NOTES

- 1. P. A. Stott, D. A. Stone, M. R. Allen, Human contribution to the European heatwave of 2003. *Nature* **432**, 610–614 (2004).
- 2. The National Academies Press, Attribution of Extreme Weather Events in the Context of Climate Change (The National Academies Press, 2016).
- O. Angélil, D. Stone, M. Wehner, C. J. Paciorek, H. Krishnan, W. Collins, An independent assessment of anthropogenic attribution statements for recent extreme temperature and rainfall events. J. Climate 30, 5–16 (2017).
- N. S. Diffenbaugh, D. Singh, J. S. Mankin, D. E. Horton, D. L. Swain, D. Touma, A. Charland, Y. Liu, M. Haugen, M. Tsiang, B. Rajaratnam, Quantifying the influence of global warming on unprecedented extreme climate events. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4881–4886 (2017).
- N. S. Diffenbaugh, D. Singh, J. S. Mankin, Unprecedented climate events: Historical changes, aspirational targets, and national commitments. *Sci. Adv.* 4, eaao3354 (2018).
- E. M. Fischer, R. Knutti, Anthropogenic contribution to global occurrence of heavyprecipitation and high-temperature extremes. *Nat. Clim. Chang.* 5, 560–564 (2015).
- N. Christidis, P. A. Stott, F. W. Zwiers, Fast-track attribution assessments based on pre-computed estimates of changes in the odds of warm extremes. *Climate Dynam.* 45, 1547–1564 (2015).
- M. M. Vogel, J. Zscheischler, R. Wartenburger, D. Dee, S. I. Seneviratne, Concurrent 2018 hot extremes across northern hemisphere due to human-induced climate change. *Earth's Futur.* 7, 692–703 (2019).

- D. Singh, D. L. Swain, J. S. Mankin, D. E. Horton, L. N. Thomas, B. Rajaratnam, N. S. Diffenbaugh, Recent amplification of the North American winter temperature dipole. *J. Geophys. Res. Atmos.* **121**, 9911–9928 (2016).
- K. van der Wiel, S. B. Kapnick, G. J. van Oldenborgh, K. Whan, S. Philip, G. A. Vecchi, R. K. Singh, J. Arrighi, H. Cullen, Rapid attribution of the August 2016 flood-inducing extreme precipitation in south Louisiana to climate change. *Hydrol. Earth Syst. Sci.* 21, 897–921 (2017).
- P. Pall, T. Aina, D. A. Stone, P. A. Stott, T. Nozawa, A. G. J. Hilberts, D. Lohmann, M. R. Allen, Anthropogenic greenhouse gas contribution to flood risk in England and Wales in autumn 2000. *Nature* 470, 382–385 (2011).
- N. S. Diffenbaugh, D. L. Swain, D. Touma, Anthropogenic warming has increased drought risk in California. Proc. Natl. Acad. Sci. U.S.A. 112, 3931–3936 (2015).
- 13. K. Emanuel, Assessing the present and future probability of Hurricane Harvey's rainfall. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12681–12684 (2017).
- G. J. van Oldenborgh, K. van der Wiel, A. Sebastian, R. Singh, J. Arrighi, F. Otto, K. Haustein, S. Li, G. Vecchi, H. Cullen, Attribution of extreme rainfall from Hurricane Harvey, August 2017. *Environ. Res. Lett.* 12, 124009 (2017).
- N. Lin, R. E. Kopp, B. P. Horton, J. P. Donnelly, Hurricane Sandy's flood frequency increasing from year 1800 to 2100. Proc. Natl. Acad. Sci. U.S.A. 113, 12071–12075 (2016).
- M. C. Kirchmeier-Young, F. W. Zwiers, N. P. Gillett, Attribution of extreme events in Arctic sea ice extent. J. Climate 30, 553–571 (2017).
- 17. K. E. Trenberth, J. T. Fasullo, T. G. Shepherd, Attribution of climate extreme events. *Nat. Clim. Chang.* **5**, 725–730 (2015).
- T. G. Shepherd, A common framework for approaches to extreme event attribution. *Curr. Clim. Chang. Rep.* 2, 28–38 (2016).
- D. Coumou, V. Petoukhov, S. Rahmstorf, S. Petri, H. J. Schellnhuber, Quasi-resonant circulation regimes and hemispheric synchronization of extreme weather in boreal summer. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12331–12336 (2014).
- M. E. Mann, S. Rahmstorf, K. Kornhuber, B. A. Steinman, S. K. Miller, D. Coumou, Influence of anthropogenic climate change on planetary wave resonance and extreme weather events. *Sci. Rep.* 7, 45242 (2017).
- D. L. Swain, M. Tsiang, M. Haugen, D. Singh, A. Charland, B. Rajaratnam, N. S. Diffenbaugh, The extraordinary California drought of 2013-2014: Character, context, and the role of climate change [in "Explaining Extremes of 2013 from a Climate Perspective"]. *Bull. Am. Meteorol. Soc.* 95, S3–S7 (2014).
- S. Wang, L. Hipps, R. R. Gillies, J.-H. Yoon, Probable causes of the abnormal ridge accompanying the 2013–2014 California drought: ENSO precursor and anthropogenic warming footprint. *Geophys. Res. Lett.* 41, 3220–3226 (2014).
- K. E. Trenberth, L. Cheng, P. Jacobs, Y. Zhang, J. Fasullo, Hurricane Harvey links to ocean heat content and climate change adaptation. *Earth's Futur.* 6, 730–744 (2018).
- M. C. Kirchmeier-Young, N. P. Gillett, F. W. Zwiers, A. J. Cannon, F. S. Anslow, Attribution of the influence of human-induced climate change on an extreme fire season. *Earth's Futur.* 7, 2–10 (2019).
- F. E. L. Otto, K. van der Wiel, G. J. van Oldenborgh, S. Philip, S. F. Kew, P. Uhe, H. Cullen, Climate change increases the probability of heavy rains in Northern England/Southern Scotland like those of storm Desmond—A real-time event attribution revisited. *Environ. Res. Lett.* 13, 24006 (2018).
- K. Haustein, F. E. L. Otto, P. Uhe, N. Schaller, M. R. Allen, L. Hermanson, N. Christidis, P. McLean, H. Cullen, Real-time extreme weather event attribution with forecast seasonal SSTs. *Environ. Res. Lett.* **11**, 64006 (2016).
- A. Hannart, A. Carrassi, M. Bocquet, M. Ghil, P. Naveau, M. Pulido, J. Ruiz, P. Tandeo, DADA: Data assimilation for the detection and attribution of weather and climate-related events. *Clim. Change* **136**, 155–174 (2016).
- D. Coumou, A. Robinson, S. Rahmstorf, Global increase in record-breaking monthly-mean temperatures. *Clim. Change* 118, 771–782 (2013).
- F. C. Lott, M. Gordon, R. J. Graham, A. A. Scaife, M. Vellinga, Reliability of African climate prediction and attribution across timescales. *Environ. Res. Lett.* 9, 104017 (2014).
- T. C. Peterson, R. R. Heim Jr., R. Hirsch, D. P. Kaiser, H. Brooks, N. S. Diffenbaugh, R. M. Dole, J. P. Giovannettone, K. Guirguis, T. R. Karl, R. W. Katz, K. Kunkel, D. Lettenmaier, G. J. McCabe, C. J. Paciorek, K. R. Ryberg, S. Schubert, V. B. S. Silva, B. C. Stewart, A. V. Vecchia, G. Villarini, R. S. Vose, J. Walsh, M. Wehner, D. Wolock, K. Wolter, C. A. Woodhouse, D. Wuebbles, Monitoring and understanding changes in heat waves, cold waves, floods and droughts in the United States: State of knowledge. *Bull. Am. Meteorol. Soc.* **94**, 821–834 (2013).
- M. E. Mann, E. A. Lloyd, N. Oreskes, Assessing climate change impacts on extreme weather events: The case for an alternative (Bayesian) approach. *Clim. Change* 144, 131–142 (2017).
- A. Hannart, J. Pearl, F. E. L. Otto, P. Naveau, M. Ghil, Causal counterfactual theory for the attribution of weather and climate-related events. *Bull. Am. Meteorol. Soc.* 97, 99–110 (2015).
- A. Jézéquel, V. Dépoues, H. Guillemot, M. Trolliet, J.-P. Vanderlinden, P. Yiou, Behind the veil of extreme event attribution. *Clim. Change* 149, 367–383 (2018).

- P. A. Stott, D. J. Karoly, F. W. Zwiers, Is the choice of statistical paradigm critical in extreme event attribution studies? *Clim. Change* 144, 143–150 (2017).
- R. B. Alley, K. A. Emanuel, F. Zhang, Advances in weather prediction. *Science* 363, 342–344 (2019).
- W. S. Broecker, Climatic change: Are we on the brink of a pronounced global warming? Science 189, 460–463 (1975).
- J. Hansen, I. Fung, A. Lacis, D. Rind, S. Lebedeff, R. Ruedy, G. Russell, P. Stone, Global climate changes as forecast by Goddard Institute for Space Studies three-dimensional model. J. Geophys. Res. Atmos. 93, 9341–9364 (1988).
- T. R. Knutson, J. Kam, F. Zeng, A. T. Wittenberg, CMIP5 model-based assessment of anthropogenic influence on record global warmth during 2016. *Bull. Am. Meteorol. Soc.* 99, S11–S15 (2018).
- O. Bellprat, F. Doblas-Reyes, Attribution of extreme weather and climate events overestimated by unreliable climate simulations. *Geophys. Res. Lett.* 43, 2158–2164 (2016).
- O. Bellprat, V. Guemas, F. Doblas-Reyes, M. G. Donat, Towards reliable extreme weather and climate event attribution. *Nat. Commun.* 10, 1732 (2019).
- A. Weisheimer, N. Schaller, C. O'Reilly, D. A. MacLeod, T. Palmer, Atmospheric seasonal forecasts of the twentieth century: Multi-decadal variability in predictive skill of the winter North Atlantic Oscillation (NAO) and their potential value for extreme event attribution. Q. J. Roy. Meteorol. Soc. 143, 917–926 (2017).
- S. Jeon, C. J. Paciorek, M. F. Wehner, Quantile-based bias correction and uncertainty quantification of extreme event attribution statements. *Weather Clim. Extrem.* 12, 24–32 (2016).
- S. Sippel, D. Mitchell, M. T. Black, A. J. Dittus, L. Harrington, N. Schaller, F. E. L. Otto, Combining large model ensembles with extreme value statistics to improve attribution statements of rare events. *Weather Clim. Extrem.* 9, 25–35 (2015).
- B. Christiansen, The role of the selection problem and non-Gaussianity in attribution of single events to climate change. J. Climate 28, 9873–9891 (2015).
- C. J. Paciorek, D. A. Stone, M. F. Wehner, Quantifying statistical uncertainty in the attribution of human influence on severe weather. *Weather Clim. Extrem.* 20, 69–80 (2018).
- L. J. Wilcox, P. Yiou, M. Hauser, F. C. Lott, G. J. van Oldenborgh, I. Colfescu, B. Dong, G. Hegerl, L. Shaffrey, R. Sutton, Multiple perspectives on the attribution of the extreme European summer of 2012 to climate change. *Climate Dynam.* 50, 3537–3555 (2018).
- M. Hauser, L. Gudmundsson, R. Orth, A. Jézéquel, K. Haustein, R. Vautard, G. J. van Oldenborgh, L. Wilcox, S. I. Seneviratne, Methods and model dependency of extreme event attribution: The 2015 european drought. *Earth's Futur.* 5, 1034–1043 (2017).
- K. E. Taylor, R. J. Stouffer, G. A. Meehl, An overview of CMIP5 and the experiment design. Bull. Am. Meteorol. Soc. 93, 485–498 (2012).
- J. Sillmann, V. V. Kharin, X. Zhang, F. W. Zwiers, D. Bronaugh, Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res. Atmos.* **118**, 1716–1733 (2013).
- P. C. D. Milly, J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, R. J. Stouffer, Stationarity is dead: Whither water management? *Science* 319, 573–574 (2008).
- L. Cheng, A. AghaKouchak, E. Gilleland, R. W. Katz, Non-stationary extreme value analysis in a changing climate. *Clim. Change* 127, 353–369 (2014).
- Climate-Safe Infrastructure Working Group, Paying It forward: The Path Toward Climate-Safe Infrastructure in California (California Natural Resources Agency, 2018).
- 53. E. Mills, Insurance in a climate of change. Science **309**, 1040–1044 (2005).
- 54. IPCC, Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (Cambridge Univ. Press, 2012).
- C. Huggel, D. Stone, H. Eicken, G. Hansen, Potential and limitations of the attribution of climate change impacts for informing loss and damage discussions and policies. *Clim. Change* 133, 453–467 (2015).
- Z. T. Pan, R. W. Arritt, E. S. Takle, W. J. Gutowski Jr., C. J. Anderson, M. Segal, Altered hydrologic feedback in a warming climate introduces a "warming hole". *Geophys. Res. Lett.* 31, L17109 (2004).
- S. Yu, K. Alapaty, R. Mathur, J. Pleim, Y. Zhang, C. Nolte, B. Eder, K. Foley, T. Nagashima, Attribution of the United States "warming hole": Aerosol indirect effect and precipitable water vapor. *Sci. Rep.* 4, 6929 (2014).
- G. A. Meehl, J. M. Arblaster, C. T. Y. Chung, Disappearance of the southeast U.S. "warming hole" with the late 1990s transition of the Interdecadal Pacific Oscillation. *Geophys. Res. Lett.* 42, 5564–5570 (2015).
- N. S. Diffenbaugh, M. Ashfaq, M. Scherer, Transient regional climate change: Analysis of the summer climate response in a high-resolution, century-scale ensemble experiment over the continental United States. J. Geophys. Res. Atmos. 116, D24111 (2011).
- B. D. Santer, C. Bonfils, J. F. Painter, M. D. Zelinka, C. Mears, S. Solomon, G. A. Schmidt, J. C. Fyfe, J. N. S. Cole, L. Nazarenko, K. E. Taylor, F. J. Wentz, Volcanic contribution to decadal changes in tropospheric temperature. *Nat. Geosci.* 7, 185–189 (2014).

Acknowledgments: I thank the editor and three anonymous reviewers for insightful and constructive feedback. I thank the CLIMDEX project for access to the observational and climate model data, as well as DDE's PCMDI and the participating climate modeling groups (which made the climate model data available to CLIMDEX and the broader community). I also thank the CEES and SRCC at Stanford University for access to computational resources. Funding: I acknowledge funding support from Stanford University. Author contributions: N.S.D. is the sole author. Competing interests: The author declares no competing interests. Data and materials availability: All data are available from the CLIMDEX archive.

Submitted 31 May 2019 Accepted 20 December 2019 Published 18 March 2020 10.1126/sciadv.aay2368

Citation: N. S. Diffenbaugh, Verification of extreme event attribution: Using out-of-sample observations to assess changes in probabilities of unprecedented events. *Sci. Adv.* **6**, eaay2368 (2020).

Verification of extreme event attribution: Using out-of-sample observations to assess changes in probabilities of unprecedented events

Noah S. Diffenbaugh

Sci Adv **6** (12), eaay2368. DOI: 10.1126/sciadv.aay2368

ARTICLE TOOLS	http://advances.sciencemag.org/content/6/12/eaay2368
SUPPLEMENTARY MATERIALS	http://advances.sciencemag.org/content/suppl/2020/03/16/6.12.eaay2368.DC1
REFERENCES	This article cites 57 articles, 10 of which you can access for free http://advances.sciencemag.org/content/6/12/eaay2368#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science Advances (ISSN 2375-2548) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title Science Advances is a registered trademark of AAAS.

Copyright © 2020 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).