

Verifying storyline attribution of extreme weather and climate events

Noah Diffenbaugh

diffenbaugh@stanford.edu

Stanford University <https://orcid.org/0000-0002-8856-4964>

Jared Trok

Stanford University <https://orcid.org/0000-0003-3020-8753>

Article

Keywords:

Posted Date: September 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11168403/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

1
2
3
4
5 **Verifying storyline attribution of extreme weather and climate events**
6

7
8
9 Noah S. Diffenbaugh^{1,*} and Jared T. Trok¹
10

11
12
13 ¹ Doerr School of Sustainability, Stanford University, Stanford, CA, USA
14

15
16 * corresponding author:
17

18 Noah S. Diffenbaugh
19

20 diffenbaugh@stanford.edu
21

22 473 Via Ortega
23 Stanford, CA, 94305-4216
24 USA
25
26

Abstract

Although the underlying methodological elements of extreme event attribution (EEA) have been thoroughly evaluated, verifying EEA results remains challenging, particularly for storyline attribution of extreme weather and climate events. We propose a methodology that provides an objective, falsifiable test of storyline EEA based on the difference in intensity of recent events (2000-2024) and analogues with similar meteorological patterns earlier in the satellite era (1979-1999). We apply this framework to recently published machine learning-based storyline attribution results for three regions spanning Europe and the Mediterranean. For all three regions, there is high correlation ($r^2 = 0.851\text{--}0.892$) between predicted and observed event-analogue temperature differences, and the biases that do exist suggest the attribution estimates are conservative. After removing the effect of meteorological differences between the hottest events of the year, the correlation remains high for the Southern ($r^2 = 0.733$) and Central ($r^2 = 0.617$) regions, suggesting that the EEA results accurately capture the influence of global temperature on historical extreme heat events over these regions. For the Northern region, the accuracy remains high for many individual events but the overall correlation is substantially reduced ($r^2 = 0.259$), suggesting that the EEA results do not fully capture the influence of global temperature over this region. While it is not possible to directly observe what would happen if today's meteorological conditions occurred in the absence of anthropogenic forcing, our approach provides a pathway to verify storyline EEA results within a framework of independent falsifiability.

1. Introduction

Extreme event attribution (EEA), in which objective analytical methods are used to systematically test the influence of global warming on weather and climate events, has expanded rapidly over the past two decades^{1–8}. Not only has the number of studies increased exponentially⁵, but methodologies have matured sufficiently that threshold-crossing events can be compared with pre-computed probabilities^{9–12}, rapid attribution assessments are routinely provided for high-profile events¹³, and real-time attribution is becoming operationalized^{5,12,14–18}. Further, EEA approaches are being extended to extreme event impacts^{5,7,19–25}, and contributing to the evidentiary basis for claims of financial responsibility^{5,26–33}.

The underlying methodological elements of EEA have been thoroughly evaluated. This includes the fidelity between climate model simulations and historical observations, uncertainty in probability calculations given the short non-stationary observational record, and the implications of underlying statistical paradigms for EEA conclusions^{4,5,40–42,10,13,34–39}. Methodological uncertainty has also been quantified by comparing the results of different EEA approaches for the same event^{5,34,43–45}. However, while these validations increase confidence in the EEA endeavor, they do not independently verify specific attribution results, which requires framing the analysis as a falsifiable prediction that can be tested using independent out-of-sample observations. The fact that extreme events occur under a particular set of meteorological conditions influenced by both unforced variability and global warming⁴¹ poses fundamental challenges for quantitative partitioning of causal contributions^{5,17,39–41,46,47}, and raises the question of whether it is possible to independently falsify EEA results⁴⁸.

Most EEA analyses fall into two broad categories⁵: “probabilistic event attribution” (in which the probability of exceeding an event threshold is calculated in actual and counterfactual

climates) and “storyline event attribution” (in which thermodynamic contributions to event intensity are quantified conditional on that event’s specific meteorological ingredients). Because probabilistic attribution is based on an implicit prediction of threshold exceedance, results can be evaluated using out-of-sample occurrence frequencies⁴⁸. However, because this evaluation relies on collective verification of a class of events, it does not address the challenges specific to storyline attribution⁶. This distinction is important given increasing emphasis on impact attribution⁵, for which storyline approaches are better suited than probabilistic approaches^{26,49}.

While it is not possible to directly observe identical meteorological conditions in the presence and absence of anthropogenic forcings, a sufficiently large set of storyline EEA results could be framed as a set of predictions that can be tested using independent observations of specific events. In this study, we propose a methodology based on the predicted and observed intensities of pairs of events with similar atmospheric circulation patterns (i.e., “analogues”) but different background climate conditions (Fig. 1). This framework shares some features with analogue-based EEA approaches^{5,6,50,51}. However, rather than using circulation analogues to quantify the anthropogenic contribution in the context of natural variability, our approach uses circulation analogues to identify pairs of historical events with which to test storyline EEA predictions using out-of-sample observations.

Our analysis leverages the results of Callahan et al.²¹, which extended the machine learning (ML)-based EEA methods of Trok et al.³⁴ to produce storyline analyses for multiple historical heatwaves over three IPCC regions encompassing Europe and the Mediterranean. For each day in 1979-2024, this dataset includes a prediction of the regional temperature at the actual annual global temperature and the pre-industrial baseline, conditional on daily meteorological inputs from the ERA5 reanalysis. Because this dataset includes predictions for all summer days

over almost half a century, it can be used to create an event-analogue pair for each of the annual hottest events of the recent period (2000-2024), in which the analogue has a similar meteorological pattern but occurred under different background conditions earlier in the satellite era (1979-1999). The predicted event-analogue differences can then be compared with the observed differences, after controlling for variations in meteorological conditions between the hottest events of the year.

2. Methods

2.1 Analysis code

The analysis code was written by the Anthropic “Claude” AI model and reviewed by the OpenAI “Codex” AI model, under the direction of N.S.D. (Note that AI was not used for any other task such as writing or editing text, identifying relevant literature, or scientific design.)

In the first phase, a Claude “sonnet” chatbot, running in the Stanford AI Playground without access to the source directories or datasets, produced initial analysis code. In the second phase, a Claude “sonnet” agent, running on Stanford’s Sherlock computing cluster with direct access to the directories and datasets, reviewed and organized the chatbot code, replicated the results, and made further refinements as requested by N.S.D. In the third phase, a different Claude agent replicated those results, identified inconsistencies, and made further refinements as requested by N.S.D. In the fourth phase, a Codex agent, running on the Sherlock cluster with direct access to the directories and datasets, conducted adversarial read-only reviews, which the Claude agent validated and corrected. As documented in the Supplemental Materials, this adversarial process required numerous rounds of review/correction.

The prompt and findings from the final round of Codex adversarial review, along with a summary of the corrections that the Claude agent made in response, are included in the Supplementary Materials and in the GitHub repository (https://github.com/sureshdiff/storyline_EEA_verification).

2.2 Datasets

We analyze the EEA results published in Callahan et al.²¹. These analyses used the ML-based EEA framework of Trok et al.³⁴, in which convolutional neural networks (CNNs) are trained, tested and validated on an ensemble of global climate model simulations across a broad range of forcings, with daily meteorological fields, the calendar day, and the annual global temperature anomaly (GMT_{annual}) used as inputs to predict daily regional-mean temperature. Once the CNNs are trained, tested and validated on GCM data, reanalysis data are used as out-of-sample inputs to predict daily regional-mean temperature across a range of GMT_{annual} , including both the GMT_{annual} at which the actual meteorological conditions occurred and counterfactual GMT_{annual} such as 0°C or 1.5°C. To account for uncertainty in the CNN parameters, multiple CNNs are trained, each with a different random seed. To account for differences between the training GCMs, the GCM label is used as an additional input.

Callahan et al. repeated this process for three different regions designated by the IPCC⁵²: NEU/Northern Europe (which we refer to as “Northern”), WCE/Western & Central Europe (which we refer to as “Central”), and MED/Mediterranean (which we refer to as “Southern”). For each region, the dataset consists of 25 predictions (5 GCMs x 5 random seeds) of the regional temperature for each summer day from 1979 through 2024 at both the actual GMT_{annual} and each counterfactual GMT_{annual} .

In the current study, we use the ensemble of 25 CNN predictions at the actual GMT_{annual} and at $GMT_{annual} = 0^{\circ}\text{C}$ for each of the three regions. In addition, we analyze daily temperatures and 500 hPa geopotential heights from the ERA5 reanalysis dataset⁵³, using the domain boundaries from Callahan et al.²¹ (see below).

2.3 Analyses

Our analogue-based evaluation of storyline attribution results includes four primary steps. First, we separate the reanalysis into a recent period (2000-2024) for extreme event identification and an early period (1979-1999) for analogue comparison. Second, for each year in 2000-2024, we identify the hottest 3-day event of the year. Third, for each annual hottest 3-day event in 2000-2024, we identify the 3-day summer period in 1979-1999 with the closest 500 hPa geopotential height anomaly pattern. Fourth, we compare the predicted and observed temperature differences between the 2000-2024 annual hottest 3-day events and their respective 1979-1999 analogues, both for the raw temperature differences and after isolating the influence of global temperature by controlling for meteorological differences between the annual hottest events. We emphasize that the goal of this approach is not to identify a perfect analogue for each event, but to pose predictions that can independently test whether the attribution framework accurately represents the difference in event intensity when similar meteorological conditions occur under different background conditions.

While there are various approaches for identifying meteorological analogues, we use the pattern correlation of 500 hPa geopotential heights as a first-order criterion⁵⁴. For this analogue search, we use the atmospheric domains from Callahan et al.²¹, which are centered over the respective IPCC regions and span $\sim 34^{\circ}\text{W}$ – 53°E and $\sim 21^{\circ}\text{N}$ – 55°N (Southern), $\sim 32^{\circ}\text{N}$ – 66°N

(Central) and $\sim 41^{\circ}\text{N}$ – 75°N (Northern). To identify analogue events during the hottest part of the year, we restrict this search to June, July and August. (Empirically, for 2000-2024, the hottest 3-day event of the year falls in June through August for the Northern and Central regions, and in July through August in the Southern region.)

Based on previous methods that search for analogues within the same calendar month as the event⁵⁴, for our primary analysis we calculate daily geopotential height anomalies relative to a 28-day running-mean baseline. This baseline is centered on the calendar day and calculated over the 1981-2010 period at each grid point. Thus, for each calendar day at a given grid point, the baseline climatology is the mean of the raw daily 500 hPa geopotential height for days within ± 14 calendar days during 1981-2010. For each day of each year in the 1979-2024 ERA5 dataset, we subtract the smoothed calendar-day baseline value from the raw daily 500 hPa geopotential height.

We use this 1979-2024 daily 500 hPa geopotential height anomaly dataset to identify the closest 1979-1999 analogue for each annual hottest 3-day event in 2000-2024. For a given year in 2000-2024, we calculate the mean of the three daily 500 hPa geopotential height anomaly fields during the hottest 3-day event of the year. We then identify each 3-day mean 500 hPa geopotential height anomaly field during 1979-1999 that (1) occurs during June-July-August and (2) has a calendar-day start date within ± 28 days of the calendar-day start date of the target event. We then calculate the pattern correlation and RMSE between each of these 500 hPa geopotential height anomaly fields from 1979-1999 and the 500 hPa geopotential height anomaly field of the target event. The anomaly field from 1979-1999 with the highest pattern correlation is selected as the event's analogue, with RMSE used as a secondary criterion in case of equivalent pattern correlation.

To confirm that the results are not sensitive to arbitrary choices in the analogue selection, we test the sensitivity to other treatments of the 500 hPa geopotential height fields, including using the raw 500 hPa geopotential heights, removing the 1981-2010 JJA mean field from each daily field, and using the normalized CNN inputs from Callahan et al. (which have been detrended with respect to GMT_{annual} and standardized with respect to the interannual calendar-day variability at each grid point). For these sensitivity tests, we search for analogues in the window from June 1st through August 31st of each year in 1979-1999.

Having identified the 500 hPa geopotential height analogues, we calculate the difference in event intensity between each target event in 2000-2024 and its respective analogue in 1979-1999. We refer to this event-analogue difference as τ . For each of the hottest 3-day events in 2000-2024, we calculate both the predicted and observed τ . For the predicted τ , we use the ensemble of 25 CNN predictions of daily regional temperature generated using ERA5 inputs at the actual GMT_{annual} . For the observed τ , we use the daily ERA5 reanalysis temperatures averaged over the region. For our main analysis, we use the absolute daily regional-mean temperature to calculate τ . In addition, we test the sensitivity of our results to other treatments of the regional temperature, including the daily temperature anomaly relative to the ERA5 calendar-day baseline, and bias-correcting the CNN predictions via quantile mapping of the daily temperature distribution.

Finally, we quantify the influence of annual global temperature on τ after controlling for meteorological differences between the hottest events of the year:

$$\tau = \tau_{meteorology} + \tau_{GMT} \quad (1)$$

where τ is the event-analogue temperature difference at the actual GMT_{annual} , $\tau_{meteorology}$ is the event-analogue temperature difference predicted at $GMT_{annual} = 0^\circ\text{C}$ that represents the

contribution of meteorological differences to τ , and τ_{GMT} is the residual that represents the contribution of GMT to τ after normalizing for meteorological differences (Fig. 2). We remove the ensemble of event-analogue temperature differences predicted at GMT = 0°C ($\tau_{meteorology}$) from both the ensemble of τ predicted at the actual GMT_{annual} and from the single observed τ , yielding ensembles of normalized predicted and observed event-analogue differences (τ_{GMT}).

This approach could in principle be used to evaluate storyline EEA results for individual events (e.g., Fig. 1D). In addition, we calculate the correlation between predicted and observed τ_{GMT} to evaluate the skill of the ML-based EEA framework in predicting the influence of global temperature across multiple events (Fig. 3).

3. Results

The predicted and observed temperature differences between the annual hottest events and their respective analogues (τ) are highly correlated (e.g., $r^2 = 0.877$ for the pooled data; Fig. ED1). In addition, the accuracy of the CNNs is robust to different treatments of circulation indicators and associated regional temperatures (Fig. ED2). This includes raw and normalized geopotential heights, and both bias-corrected temperature and calendar-day anomalies. The predicted and observed τ both exhibit a substantial range, including temperature differences that exceed 6°C (Fig. 1, ED1, ED2). This range reflects the aggregate effects of multiple differences between the events and their analogues, including the nuances of the daily circulation patterns (e.g., Fig. 1A, 1B), the subseasonal-to-seasonal antecedent conditions (e.g., soil moisture⁵⁵), and the annual global temperature anomaly²¹. The accuracy in predicting τ across a range of events and analogues reflects the capability of the CNNs to predict the regional temperature response to these various factors. Further, although there is heterogeneity in the similarity of atmospheric

conditions across the event-analogue pairs, the predicted τ is generally not sensitive to the correlation, RMSE, or calendar-day difference between the event and the analogue (Fig. ED3, ED4, ED5).

As our motivation is to evaluate EEA results, it is critical to isolate the contribution of global temperature (τ_{GMT}) from the influence of meteorological differences ($\tau_{meteorology}$). To do so, for each event we calculate the predicted τ if both the event and its analogue occurred at $GMT_{annual} = 0^{\circ}C$ ($\tau_{meteorology}$; Fig. 2, ED6). There is no time trend in $\tau_{meteorology}$ (Fig. 2, top row). However, given the strong influence of meteorological conditions on event intensity (Fig. ED6), normalizing for meteorological conditions reduces both the uncertainty in predicted τ for individual events and the correlation between the predicted and observed τ (Fig. 1D, 3).

In contrast, there is a clear time trend in the contribution of global temperature (τ_{GMT}): Based on the CNN predictions, we find very high likelihood ($p < 0.005$) that τ_{GMT} has increased over the 2000-2024 period in all three regions (Fig. 2, top row). The median OLS slope suggests an increase of $1.32^{\circ}C$ since the start of the 21st century over the Central region, along with increases of $0.74^{\circ}C$ and $0.65^{\circ}C$ over the Southern and Northern regions, respectively.

Comparing the predicted and observed event-analogue differences after normalizing for meteorological conditions isolates the skill of the ML-based EEA framework in capturing the influence of global temperature, and provides the key test of the EEA results. In the case of our published results²¹, before normalization there is high correlation between predicted and observed τ for all three regions ($r^2 = 0.851$ – 0.892 ; Fig. 3). After normalization, the skill remains high for both the Southern ($r^2 = 0.733$) and Central ($r^2 = 0.617$) regions. For the Northern region, after normalization there is strong agreement between predicted and observed values for many individual events, but the overall skill is substantially reduced ($r^2 = 0.259$).

4. Discussion

At present, our EEA verification is only tractable because ML-based approaches enable factual and counterfactual predictions for a large sample of days in the historical record. While it has previously been possible to generate large EEA datasets using pre-computed approaches^{9–12,56}, by definition these do not incorporate the physical response to anthropogenic forcing for the specific meteorological conditions of an actual event. Doing so has typically required computationally-intensive numerical experiments tailored to the conditions of each event^{14,57}. More recent approaches that leverage large archives of numerical experiments and advances in machine learning have made large storyline EEA datasets possible^{21,34,58–61}. Apart from facilitating greater access to EEA analyses⁵⁸, enabling storyline EEA results to be tested as falsifiable predictions is a major benefit of these approaches.

Our ML-based storyline EEA framework has already been shown to accurately predict historical trends in the hottest week of the year^{21,59}. The verification approach proposed here builds on this work to more explicitly test the predicted influence of global temperature. Pairing recent extreme heat events with analogues earlier in the historical record is a key element of this verification, as it enables quantitative evaluation of the EEA framework's ability to predict differences in the intensity of events that have similar circulation patterns but different background climate conditions. Controlling for meteorological differences between extreme events is also a key element, since the strong influence of meteorology on daily variability in surface conditions could create high correlation between predicted and observed event-analogue differences even if the EEA framework does not accurately capture the influence of global temperature.

In the case of our published results ²¹, after isolating the influence of global temperature, the agreement between predicted and observed event-analogue temperature differences remains high for both the Southern and Central regions. This durability suggests that the ML-based EEA results accurately capture the response of historical heat events to global warming over these regions. For the Northern region, although high agreement is maintained for many individual events, the overall agreement is substantially reduced, suggesting that over this region the skill in predicting raw event-analogue differences is strongly influenced by meteorological variability. While the Northern region exhibits the largest bias in predicted hottest week of the year ²¹, the disparity in capturing the influence of global temperature on event-analogue differences would not be apparent without controlling for the influence of meteorology (Fig. 3). Our results thus emphasize the importance both of rigorously evaluating the falsifiability of EEA results, and of carefully designing those evaluations to independently test the influence of global warming using out-of-sample observations.

The most prominent deficiency of the raw ML-based predictions is a persistent negative bias in τ (Fig. 3, ED1, ED2). This negative bias appears to be related to the fact that the CNNs tend to slightly underpredict the severity of the hottest events and slightly overpredict the severity of the analogues, a bias that is at least partially inherited from the GCM training data (especially in the Northern region ²¹). While correcting the CNN temperature predictions via quantile mapping addresses much of the bias in predicted τ , a slight negative bias persists even after quantile mapping (Fig. ED2). The sign of this bias suggests that the ML-based attribution results are conservative. Similarly, the relationship (i.e., slope and p-value) between τ and global temperature is weaker in the CNN predictions than the observations for all three regions (Fig.

ED7), likewise suggesting that biases in the CNN predictions make the EEA results conservative.

One possible explanation for the negative bias in τ is regional aerosol forcings, which are not explicitly captured by the ML-based predictions. These forcings have been shown to affect heatwaves over Europe, including a trend towards heatwave intensification since the start of the analogue period ^{62–64}. Incorporating these and other regional forcings is a key focus for future development, and will help to identify whether these biases are systematic or particular to the region and/or ML network configuration.

A number of other potential future improvements to ML-based EEA approaches are also in progress. These include configurations that enable spatially-explicit rather than regional-mean predictions ^{58–60,65}, as well as improvements in uncertainty quantification as the capacity to efficiently generate very large ensemble predictions continues to increase. In addition, the EEA predictions analyzed here isolate the influence of annual global temperature rather than the influence of different climate forcing factors. Our ML-based attribution framework can be extended to isolate the influence of shares of historical emissions ⁵⁹, and related AI-based approaches have recently been trained on global climate model simulations with and without anthropogenic forcing ⁵⁸. These and other innovations will further enhance the capacity to generate large EEA datasets that can be leveraged to systematically evaluate EEA results as falsifiable predictions.

5. Conclusions

While EEA methods have advanced substantially, verifying EEA results based on independent evaluation of falsifiable predictions remains challenging. We present a framework

that provides an objective, falsifiable test of storyline EEA results by pairing recent extreme events with circulation analogues earlier in the historical record, and isolating the influence of global temperature by controlling for meteorological differences between the extreme events. Although we focus on evaluating our own published results, our verification approach is not specific to any EEA storyline methodology or event type, and could be used to independently compare different methodologies. In evaluating our results for heatwaves over Europe and the Mediterranean ²¹, we find that the influence of global temperature on the event-analogue temperature difference is accurately predicted across years of events for two of the three regions, and the estimates are conservative for all three regions. These results suggest both that the ML-based EEA approach is able to accurately predict the difference in event intensity when similar meteorological conditions occur in different background climates, and that verification of EEA storyline results requires careful design to isolate the influence of global temperature. While our out-of-sample test is not a definitive confirmation of the contribution of historical global warming to each individual event, it does enable verification of storyline EEA results within a framework of independent falsifiability.

Acknowledgements

We thank Obi Nnorom Jr for assistance configuring Claude Code on the Sherlock computing cluster at Stanford University, and the Stanford Research Computing Center and the Stanford Doerr School of Sustainability Center for Computation for computing access and support. We acknowledge support from Stanford University.

Authorship Contributions

N.S.D. and J.T.T. conceived the study and designed the analyses and visualizations.
N.S.D. supervised the AI agents in writing and reviewing the custom computer code, and is
responsible for the outcome. N.S.D. wrote the initial draft of the manuscript, and J.T.T. provided
feedback on the text.

Data Availability

Custom computer code is available in the following GitHub repository:
https://github.com/sureshdiff/storyline_EEA_verification/. The CNN predictions are from
Callahan et al.²¹, whose replication materials are available at
<https://zenodo.org/records/17495872>. ERA5 reanalysis data are available from the EU
Copernicus Climate Data Store at [https://cds.climate.copernicus.eu/datasets/reanalysis-era5-](https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels)
[single-levels](https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels).

Competing Interests

The authors declare no competing interests.

References

1. Stott, P. A., Stone, D. A. & Allen, M. R. Human contribution to the European heatwave of 2003. *Nature* **432**, 610–614 (2004).
2. IPCC. *Managing the risks of extreme events and disasters to advance climate change adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2012).
3. NASEM. *Attribution of Extreme Weather Events in the Context of Climate Change*. (The National Academies Press, 2016).
4. Seneviratne, S. I. *et al.* Weather and climate extreme events in a changing climate. in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* 1513–1766 (Cambridge University Press, 2021).
5. NASEM. *Attribution of Extreme Weather and Climate Events and Their Impacts*. (National Academies Press, 2026). doi:10.17226/28590.
6. Shepherd, T. G. A Common Framework for Approaches to Extreme Event Attribution. *Curr. Clim. Chang. Reports* **2**, 28–38 (2016).
7. Clarke, B., Otto, F., Stuart-Smith, R. & Harrington, L. Extreme weather impacts of climate change: an attribution perspective. *Environ. Res. Clim.* **1**, 12001 (2022).
8. Otto, F. E. L. Attribution of Extreme Events to Climate Change. *Annu. Rev. Environ. Resour.* **48**, 813–828 (2023).
9. Christidis, N., Stott, P. & Zwiers, F. Fast-track attribution assessments based on pre-computed estimates of changes in the odds of warm extremes. *Clim. Dyn.* **45**, 1547–1564 (2015).
10. Diffenbaugh, N. S. *et al.* Quantifying the influence of global warming on unprecedented extreme climate events. *Proc. Natl. Acad. Sci.* **114**, 4881–4886 (2017).
11. Fischer, E. M. & Knutti, R. Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. *Nat. Clim. Chang.* **5**, 560–564 (2015).
12. Gilford, D. M., Pershing, A., Strauss, B. H., Haustein, K. & Otto, F. E. L. A multi-method framework for global real-time climate attribution. *Adv. Stat. Climatol. Meteorol. Oceanogr.* **8**, 135–154 (2022).
13. Otto, F. E. L. *et al.* Formally combining different lines of evidence in extreme-event attribution. *Adv. Stat. Climatol. Meteorol. Oceanogr.* **10**, 159–171 (2024).
14. Leach, N. J. *et al.* Heatwave attribution based on reliable operational weather forecasts. *Nat. Commun.* **15**, 4530 (2024).

- 397 15. Stott, P. A. & Christidis, N. Operational attribution of weather and climate extremes: what
398 next? *Environ. Res. Clim.* **2**, 13001 (2023).
- 399 16. Otto, F. E. L., Kew, S., Philip, S., Stott, P. & Van Oldenborgh, G. J. How to provide
400 useful attribution statements: lessons learned from operationalizing event attribution in
401 Europe. (2022).
- 402 17. Jézéquel, A. *et al.* Singular Extreme Events and Their Attribution to Climate Change: A
403 Climate Service–Centered Analysis. *Weather. Clim. Soc.* **12**, 89–101 (2020).
- 404 18. Haustein, K. *et al.* Real-time extreme weather event attribution with forecast seasonal
405 SSTs. *Environ. Res. Lett.* **11**, 64006 (2016).
- 406 19. Davenport, F. V., Burke, M. & Diffenbaugh, N. S. Contribution of historical precipitation
407 change to US flood damages. *Proc. Natl. Acad. Sci.* **118**, e2017524118 (2021).
- 408 20. Diffenbaugh, N. S., Davenport, F. V. & Burke, M. Historical warming has increased U.S.
409 crop insurance losses. *Environ. Res. Lett.* **16**, 84025 (2021).
- 410 21. Callahan, C. W. *et al.* Increasing risk of mass human heat mortality if historical weather
411 patterns recur. *Nat. Clim. Chang.* **16**, 26–32 (2026).
- 412 22. Strauss, B. H. *et al.* Economic damages from Hurricane Sandy attributable to sea level rise
413 caused by anthropogenic climate change. *Nat. Commun.* **12**, 2720 (2021).
- 414 23. Jézéquel, A. *et al.* Broadening the scope of anthropogenic influence in extreme event
415 attribution. *Environ. Res. Clim.* **3**, 42003 (2024).
- 416 24. Noy, I., Stone, D. & Uher, T. Extreme events impact attribution: A state of the art. *Cell*
417 *Reports Sustain.* **1**, (2024).
- 418 25. Perkins-Kirkpatrick, S. E. *et al.* Attributing heatwave-related mortality to climate change:
419 a case study of the 2009 Victorian heatwave in Australia. *Environ. Res. Clim.* **4**, 15004
420 (2025).
- 421 26. Lloyd, E. A. & Shepherd, T. G. Climate change attribution and legal contexts: evidence
422 and the role of storylines. *Clim. Change* **167**, 28 (2021).
- 423 27. Allen, M. Liability for climate change. *Nature* **421**, 891–892 (2003).
- 424 28. Wehner, M. F. & Reed, K. A. Operational extreme weather event attribution can quantify
425 climate change loss and damages. *PloS Clim.* **1**, e0000013 (2022).
- 426 29. Coumou, D. *et al.* How can event attribution science underpin financial decisions on Loss
427 and Damage? *PNAS Nexus* **3**, pgae277 (2024).
- 428 30. Marjanac, S. & Patton, L. Extreme weather event attribution science and climate change
429 litigation: an essential step in the causal chain? *J. Energy Nat. Resour. Law* **36**, 265–298

- 430 (2018).
- 431 31. Stuart-Smith, R. F., Otto, F. E. L. & Wetzer, T. Liability for climate change impacts: the
432 role of climate attribution science. in *Corporate Accountability and Liability for Climate*
433 *Change* 206–235 (Edward Elgar Publishing, 2024).
- 434 32. Burger, M., Wentz, J. A. & Horton, R. The law and science of climate change attribution.
435 *Env't L. Rep.* **51**, 10646 (2021).
- 436 33. Wentz, J., Merner, D., Franta, B., Lehmen, A. & Frumhoff, P. C. Research Priorities for
437 Climate Litigation. *Earth's Futur.* **11**, e2022EF002928 (2023).
- 438 34. Trok, J. T., Barnes, E. A., Davenport, F. V & Diffenbaugh, N. S. Machine learning–based
439 extreme event attribution. *Sci. Adv.* **10**, ead13242 (2024).
- 440 35. Paciorek, C. J., Stone, D. A. & Wehner, M. F. Quantifying statistical uncertainty in the
441 attribution of human influence on severe weather. *Weather Clim. Extrem.* **20**, 69–80
442 (2018).
- 443 36. Bellprat, O. & Doblas-Reyes, F. Attribution of extreme weather and climate events
444 overestimated by unreliable climate simulations. *Geophys. Res. Lett.* **43**, 2158–2164
445 (2016).
- 446 37. Bellprat, O., Guemas, V., Doblas-Reyes, F. & Donat, M. G. Towards reliable extreme
447 weather and climate event attribution. *Nat. Commun.* **10**, 1732 (2019).
- 448 38. Cheng, L., AghaKouchak, A., Gilleland, E. & Katz, R. W. Non-stationary extreme value
449 analysis in a changing climate. *Clim. Change* **127**, 353–369 (2014).
- 450 39. Mann, M. E., Lloyd, E. A. & Oreskes, N. Assessing climate change impacts on extreme
451 weather events: the case for an alternative (Bayesian) approach. *Clim. Change* **144**, 131–
452 142 (2017).
- 453 40. Stott, P. A., Karoly, D. J. & Zwiers, F. W. Is the choice of statistical paradigm critical in
454 extreme event attribution studies? *Clim. Change* **144**, 143–150 (2017).
- 455 41. Shepherd, T. G. Bringing physical reasoning into statistical practice in climate-change
456 science. *Clim. Change* **169**, 2 (2021).
- 457 42. Swain, D. L., Singh, D., Touma, D. & Diffenbaugh, N. S. Attributing Extreme Events to
458 Climate Change: A New Frontier in a Warming World. *One Earth* **2**, 522–527 (2020).
- 459 43. Angélil, O. *et al.* An independent assessment of anthropogenic attribution statements for
460 recent extreme temperature and rainfall events. *J. Clim.* **30**, 5–16, doi: 10.1175/JCLI-D-
461 16-0077.1 (2017).
- 462 44. Thompson, V., Ermis, S. & Athanase, M. The need for multi-method extreme event
463 attribution. *Weather* **81**, 40–45 (2026).

- 464 45. Uhe, P. *et al.* Comparison of Methods: Attributing the 2014 record European temperatures
465 to human influences. *Geophys. Res. Lett.* **43**, 8685–8693 (2016).
- 466 46. Jézéquel, A. *et al.* Behind the veil of extreme event attribution. *Clim. Change* **149**, 367–
467 383 (2018).
- 468 47. Lloyd, E. A. & Shepherd, T. G. Foundations of attribution in climate-change science.
469 *Environ. Res. Clim.* **2**, 35014 (2023).
- 470 48. Diffenbaugh, N. S. Verification of extreme event attribution: Using out-of-sample
471 observations to assess changes in probabilities of unprecedented events. *Sci. Adv.* **6**,
472 eaay2368 (2020).
- 473 49. Lloyd, E. A. & Shepherd, T. G. Environmental catastrophes, climate change, and
474 attribution. *Ann. N. Y. Acad. Sci.* **1469**, 105–124 (2020).
- 475 50. Cattiaux, J. *et al.* Winter 2010 in Europe: A cold extreme in a warming climate. *Geophys.*
476 *Res. Lett.* **37**, (2010).
- 477 51. Zhuang, Y., Fu, R., Santer, B. D., Dickinson, R. E. & Hall, A. Quantifying contributions
478 of natural variability and anthropogenic forcings on increased fire weather risk over
479 western United States. *Proc. Natl. Acad. Sci.* (2021).
- 480 52. Iturbide, M. *et al.* An update of IPCC climate reference regions for subcontinental analysis
481 of climate model data: definition and aggregated datasets. *Earth Syst. Sci. Data* **12**, 2959–
482 2970 (2020).
- 483 53. Hersbach, H. *et al.* The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049
484 (2020).
- 485 54. Yiou, P., Vautard, R., Naveau, P. & Cassou, C. Inconsistency between atmospheric
486 dynamics and temperatures during the exceptional 2006/2007 fall/winter and recent
487 warming in Europe. *Geophys. Res. Lett.* **34**, (2007).
- 488 55. Quesada, B., Vautard, R., Yiou, P., Hirschi, M. & Seneviratne, S. I. Asymmetric European
489 summer heat predictability from wet and dry southern winters and springs. *Nat. Clim.*
490 *Chang.* doi:10.1038/nclimate1536 (2012).
- 491 56. Christidis, N. & Stott, P. A. Attribution analyses of temperature extremes using a set of 16
492 indices. *Weather Clim. Extrem.* **14**, 24–35 (2016).
- 493 57. Reed, K. A., Stansfield, A. M., Wehner, M. F. & Zarzycki, C. M. Forecasted attribution of
494 the human influence on Hurricane Florence. *Sci. Adv.* **6**, eaaw9253 (2020).
- 495 58. Zaninelli, P. G., Barriopedro, D., Pérez-Aracil, J. & Drouard, M. AI-Based Anomaly
496 Detection for Extreme Event Attribution: An Analysis of European Heatwaves. *Earth*
497 *Syst. Environ.* <https://doi.org/10.1007/s41748-026-01122-6> (2026) doi:10.1007/s41748-
498 026-01122-6.

- 499 59. Trok, J. T., Barnes, E. A., Gordon, E. M., Davenport, F. V & Diffenbaugh, N. S. Global
500 emissions since the Paris Agreement have intensified Europe’s recent heatwaves.
501 *Geophys. Res. Lett.* **53**, e2026GL125079 (2026).
- 502 60. Loer, F., Schillinger, M. & Sippel, S. Probabilistic storyline attribution using machine
503 learning. *Mach. Learn. Earth* (2026).
- 504 61. Callahan, C. W. *et al.* Quantifying the contributions of climate change and adaptation to
505 mortality from unprecedented extreme heat events. *Proc. Natl. Acad. Sci.* **122**,
506 e2503577122 (2025).
- 507 62. Persad, G. G., Cummins, C. & Baldwin, J. W. Anthropogenic aerosol changes
508 disproportionately impact the evolution of global heatwave hazard and exposure. *Environ.*
509 *Res. Lett.* **20**, 84023 (2025).
- 510 63. Huntingford, T., Ye, K. & Osprey, S. Increase in European summer heatwaves driven by
511 greenhouse gases and amplified by aerosol emission reductions. *Environ. Res. Lett.* **21**,
512 114008 (2026).
- 513 64. Persad, G. G. *et al.* Rapidly evolving aerosol emissions are a dangerous omission from
514 near-term climate risk assessments. *Environ. Res. Clim.* **2**, 32001 (2023).
- 515 65. Lopez-Gomez, I. *et al.* Dynamical-generative downscaling of climate model ensembles.
516 *Proc. Natl. Acad. Sci.* **122**, e2420288122 (2025).
- 517
- 518

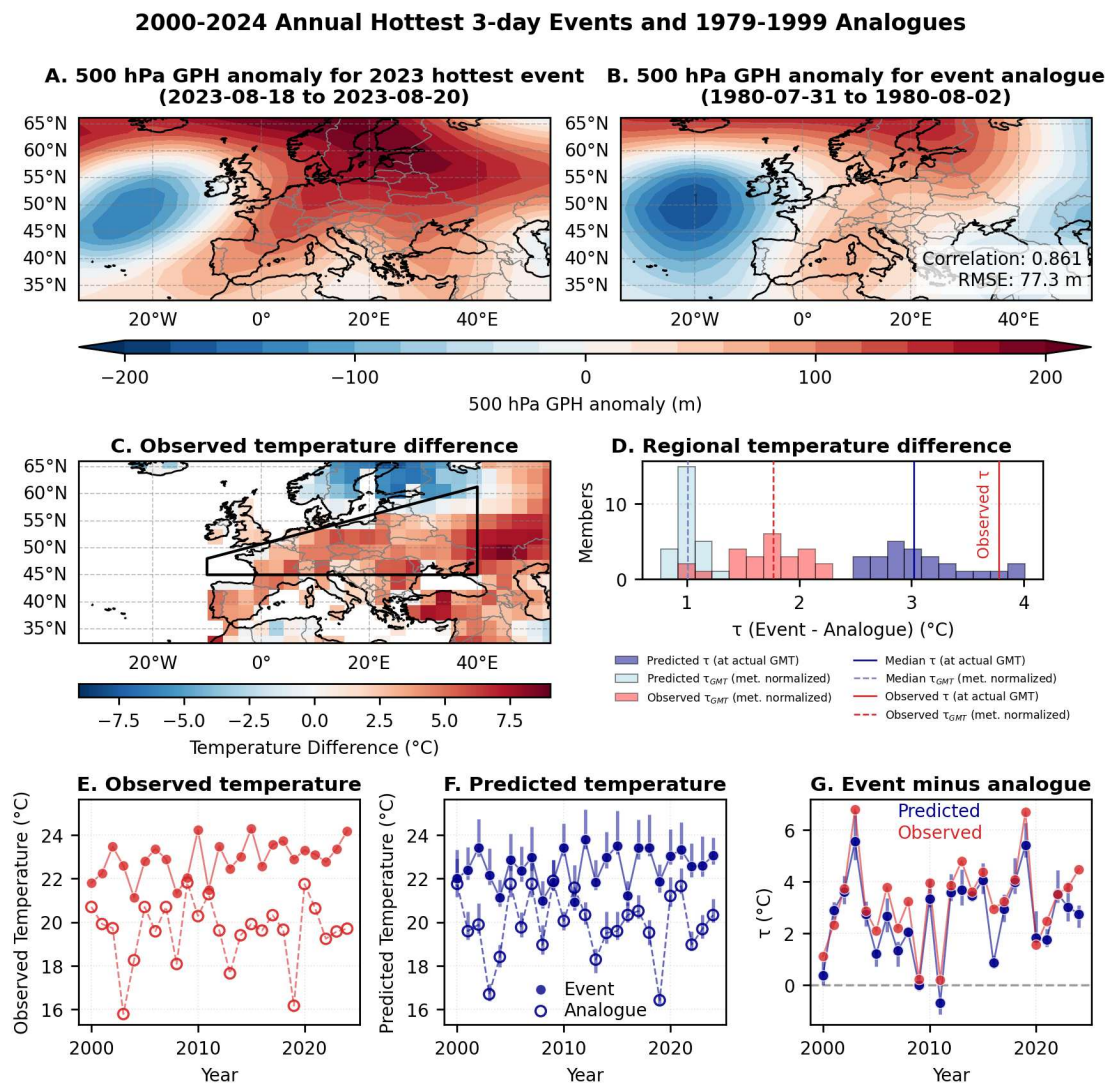


Figure 1. Annual hottest 3-day events over the IPCC’s “WCE” region. (A) 500 hPa geopotential height anomalies for the hottest 3-day event of the year in 2023. (B) 500 hPa geopotential height anomalies for the closest analogue of the 2023 event in the 1979-1999 period. (C) The difference in ERA5 reanalysis temperatures between the 2023 event and its analogue; the black polygon shows the boundaries of the WCE region. (D) Regional-mean temperature difference between the 2023 event and its analogue in the ERA reanalysis (red vertical line) and the ensemble of machine learning-based predictions from Callahan et al. (2026) (dark blue histogram). The light blue histogram shows the predictions after correcting for meteorological differences between the event and the analogue, and the red histogram shows the uncertainty in the observed value after correcting for those meteorological differences. (E) The timeseries of annual hottest 3-day events over the WCE region in the ERA5 reanalysis (solid markers), along with the respective 1979-1999 analogues (open markers). (F) As in E, but for the machine learning-based predictions; marker shows the median prediction and the whisker shows the 95% C.I. (G) The difference between event and analogue temperatures in the ERA5 reanalysis (red) and the machine learning-based predictions (black).

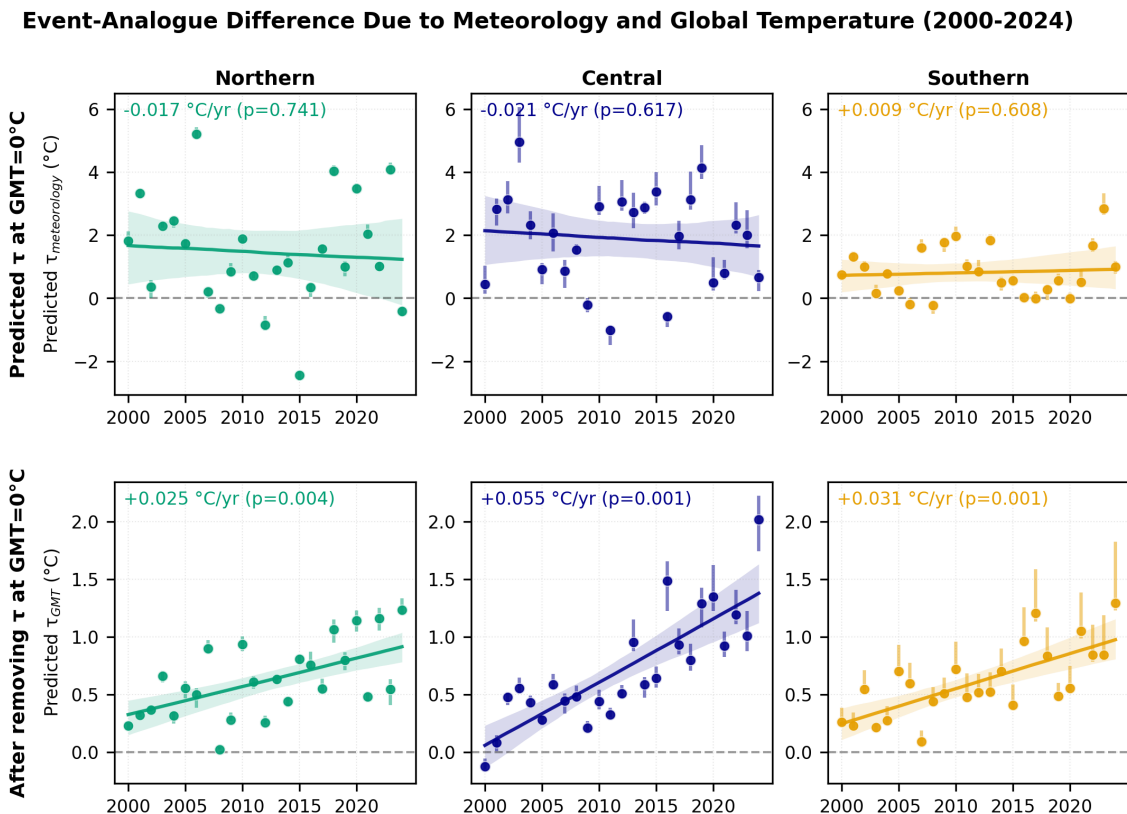


Figure 2. Decomposing the contribution of meteorology and global temperature to event-analogue temperature differences. (Top row) The contribution of meteorology to the difference in temperature between the annual hottest 3-day event of the year in 2000-2024 and its respective analogue in 1979-1999 for the IPCC’s NEU (“Northern”), WCE (“Central”) and MED (“Southern”) regions, calculated by taking the difference in predictions of both the event and the analogue at an annual GMT anomaly of 0°C above the pre-industrial. Markers show the median prediction and whiskers show the 95% C.I. (Bottom row) The contribution of GMT to the difference in temperature between the annual hottest 3-day event of the year in 2000-2024 and its respective analogue in 1979-1999, calculated as the residual difference in predicted temperature after removing the differences due to meteorological conditions shown in the top row. In all panels, the confidence interval for the OLS regression is calculated for using 1000 block bootstraps with replacement, with the block size determined by the length of statistically significant autocorrelation.

Isolating the Influence of Global Temperature on Observed vs. Predicted Event-Analogue Differences (2000-2024)

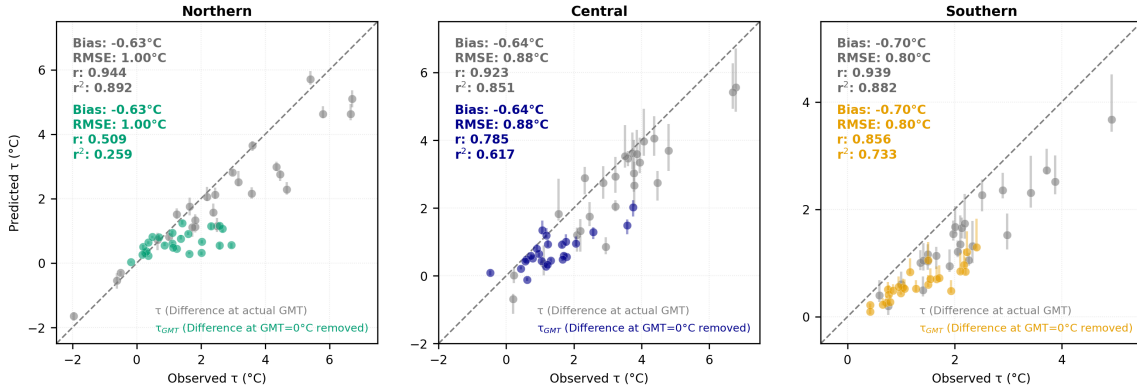


Figure 3. Isolating the influence of global temperature on observed vs. predicted event-analogue temperature differences. Gray markers show the predicted and observed difference in temperature between the hottest events of the year in 2000-2024 and their respective analogues in 1979-1999 (i.e., the values shown in Fig. 1G). Placement of gray markers on the y-axis shows the median prediction, and whiskers show the 95% C.I. Placement of gray markers on the x-axis shows the observed difference. Colored markers show the event-analogue difference in temperature after correcting for meteorological differences between the event and the analogue by removing the difference in predictions of both the event and the analogue at an annual GMT anomaly of 0°C above the pre-industrial. Placement of colored markers on the y-axis shows the median residual value from the predictions, and whiskers show the 95% C.I. (i.e., the values on the bottom row of Fig. 2). Placement of colored markers on the x-axis shows the median residual value after removing the predicted differences at 0°C from the observed value (i.e., the value shown in the red dashed line in Fig. 1D for 2023).

525

526

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [MLattributionverificationNCCSISUBMITTED.pdf](#)
- [MLattributionverificationNCCEDSUBMITTED.pdf](#)