CLIMATOLOGY

Machine learning–based extreme event attribution

Jared T. Trok[1]*, Elizabeth A. Barnes[2], Frances V. Davenport[3], Noah S. Diffenbaugh[1,4]

The observed increase in extreme weather has prompted recent methodological advances in extreme event attribution. We propose a machine learning–based approach that uses convolutional neural networks to create dynamically consistent counterfactual versions of historical extreme events under different levels of global mean temperature (GMT). We apply this technique to one recent extreme heat event (southcentral North America 2023) and several historical events that have been previously analyzed using established attribution methods. We estimate that temperatures during the southcentral North America event were 1.18° to 1.42°C warmer because of global warming and that similar events will occur 0.14 to 0.60 times per year at 2.0°C above preindustrial levels of GMT. Additionally, we find that the learned relationships between daily temperature and GMT are influenced by the seasonality of the forced temperature response and the daily meteorological conditions. Our results broadly agree with other attribution techniques, suggesting that machine learning can be used to perform rapid, low-cost attribution of extreme events.

INTRODUCTION

Since the 1800s, human emissions of greenhouse gasses have triggered a rapid period of global warming that is unprecedented in at least the past 2000 years (*1*). Many of the most destructive consequences are felt through extreme weather events such as heat waves, heavy precipitation, and droughts, which have increased in frequency and intensity in many parts of the world (*2*–*5*). Some of these trends have been formally attributed to anthropogenic climate change. For example, a recent assessment by the Intergovernmental Panel on Climate Change (IPCC) has concluded that it is extremely likely that human activity has contributed to the observed global increases in the frequency and intensity of daily temperature extremes, and more changes are expected in the future (*4*). However, since all extreme events result from the complex interaction of dynamic and thermodynamic processes, the precise contribution of anthropogenic forcing to any individual event remains difficult to quantify (*6*).

In recent years, there have been many advances in the subfield of climate science known as extreme event attribution (*3*, *7*–*10*), which is focused on understanding whether—and, if so, how—individual extreme events are influenced by human-induced climate change. Event attribution studies attempt to quantify the extent to which climate change has affected the frequency and/or intensity of individual extreme weather events by comparing the characteristics of extremes between the historical climate and a "counterfactual" climate scenario. Most event attribution studies fall under the category of "probability-based" (or "risk-based") assessments. These approaches often involve estimating changes in the probability of extremes by analyzing trends in the observational record [e.g., (*3*, *11*)], calculating return intervals of extreme events in large climate model ensembles [e.g., (*3*, *12*)], or comparing the characteristics of extreme events in multiple climate model simulations initialized with different levels of global warming [e.g., (*6*, *7*, *13*–*15*)]. Another line of probability-based attribution studies estimates the contribution of historical climate change to the magnitude of individual events of a given return interval (*6*). These studies typically use observational data [e.g., (*16*)] and/or climate model simulations [e.g., (*17*, *18*)] to estimate the change in event magnitude. Since these probability-based approaches often analyze existing climate model simulations or observational datasets, they are quite computationally efficient. These techniques can be used to make rapid attribution assessments closely following an extreme event—such as those released by the World Weather Attribution initiative [e.g., (*16*, *19*)]—or used to create "precomputed" attribution estimates [e.g., (*20*)].

Other extreme event attribution studies use a "storyline" approach (*21*) to compare event magnitude between several dynamically consistent realizations of an extreme event across different mean climate states. These approaches typically involve simulating multiple counterfactual realizations of an extreme event [or an entire historical period, (*22*)] using a climate model initialized with and without anthropogenic forcings [e.g., (*23*–*25*)] or, similarly, using counterfactuals in the historical record by comparing event intensity between several nearly identical atmospheric "flow analogs" (*26*) of an extreme event in the observational record (*27*, *28*). By ensuring that each counterfactual realization has the same atmospheric circulation patterns, these storyline-based approaches attempt to isolate the influence of anthropogenic forcing on thermodynamic drivers of a particular event (*29*). Although there is some evidence that anthropogenic forcing may alter the large-scale atmospheric circulation leading to extreme events [e.g., (*30*–*33*)], the influence of anthropogenic forcing on thermodynamic drivers is far more well-documented and well-understood [e.g., (*23*, *34*)].

Despite these recent advances, current approaches to extreme event attribution still have numerous limitations. The main disadvantage of probability-based techniques is that they rely on climate model simulations that can have large biases in how the atmospheric circulation (and other processes that influence extremes) responds to anthropogenic forcing (*35*, *36*). Therefore, it is difficult to discern whether simulated changes in extreme events are the result of robust thermodynamic changes or highly uncertain dynamic changes (*29*). Another shortcoming of using climate model ensembles to make probability-based attribution assessments is that these simulations do not include the actual meteorological conditions that occurred

[1]Department of Earth System Science, Stanford University, Stanford, CA, USA. [2]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA. [3]Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO, USA. [4]Doerr School of Sustainability, Stanford University, Stanford, CA, USA.
*Corresponding author. Email: trok@stanford.edu

during the historical extreme events (*3*). Instead, these attribution techniques are based on comparing statistical distributions within a large population of simulated weather events at different levels of climate forcing. To avoid this issue, alternative attribution approaches can be used to analyze trends in observational datasets rather than climate model simulations, but these approaches have added uncertainties in their attribution estimates due to the limited length of the observational record (*3*). This results in a trade-off in the choice of dataset used for probability-based attribution assessments, whereby climate model ensembles have large sample sizes but potentially large physical biases, while observational datasets do not have these biases but typically have larger uncertainties in the statistical fit to the climate data. To manage this trade-off, recent probability-based attribution studies have synthesized results from both observational and climate model datasets into a single comprehensive attribution statement [e.g., (*17*)].

Storyline approaches for extreme event attribution can overcome some of these issues by forcing the counterfactual events to be dynamically consistent with the observed extreme event. However, these storyline techniques provide an incomplete assessment of attribution since they do not account for anthropogenically forced changes in the probability of the meteorological conditions contributing to an event (*21*). Unfortunately, generating dynamically consistent counterfactual simulations is also quite computationally expensive and difficult to automate (*19*), which can make these storyline approaches infeasible for rapid attribution assessments of recent events unless a large amount of computing resources are available [e.g., (*25*)]. Storyline techniques are also limited in their ability to quantify future changes in event frequency (*22*), which is a vital attribution metric for informing future adaptation measures (*21*).

Given these limitations, there is an opening for additional attribution techniques that can create dynamically consistent counterfactual events to assess changes in both the magnitude and frequency of individual extreme events without requiring expensive additional climate model simulations. Recently, machine learning (and statistical learning) models have been used to detect the large-scale forced response (i.e., "fingerprint") of climate change and attribute these changes to anthropogenic forcing even amid large internal variability [e.g., (*37*–*41*)]. Moreover, although the use of deep neural networks has grown rapidly in climate research [e.g., (*40*, *42*)], they have not yet been extensively used to perform attribution analyses for extreme weather events. However, given the ability for neural networks to learn complex relationships within large climate datasets [e.g., (*38*)], these machine learning models are a promising tool for extreme event attribution that may reveal additional insights into the historical and future influence of climate change on extreme events.

Convolutional neural networks (CNNs) constitute a particular type of deep neural network specifically designed to learn relationships between two-dimensional gridded input maps and the desired output variable (*43*). Geoscience applications of CNNs are becoming increasingly commonplace, with applications in weather forecasting [e.g., (*44*, *45*)], identification of extreme weather events [e.g., (*34*, *46*–*48*)], detection of changes in extreme event frequency [e.g., (*34*)], statistical downscaling [e.g., (*49*, *50*)], and climate model parameterization [e.g., (*51*, *52*)]. Recently, explainable artificial intelligence techniques [e.g., layer-wise relevance propagation, which can be used to determine which input pixels are most important for a given output prediction (*53*)] have been used to interpret the predictions of trained CNNs (*34*, *47*, *54*) and gain insights into the physical climate processes simulated by neural networks (*38*, *42*, *55*). Likewise, partial dependence analysis [which can be used to visualize the average relationship between input and output variables (*56*)] is an explainable artificial intelligence technique that has recently been applied to CNNs to visualize the complex, nonlinear relationships between input and output variables in the geoscience context (*57*, *58*).

Building on the work of recent attribution studies, we present a framework for using machine learning to evaluate the contribution of human-caused climate change to individual extreme weather events (Fig. 1). Our storyline-based attribution framework uses CNNs and partial dependence analysis to create dynamically consistent counterfactuals for historical extreme events without requiring expensive additional climate model simulations. To perform this analysis, we first design a CNN that predicts daily maximum 2-m air temperature (TMAX) using the following input variables: the calendar day, the annual global mean surface temperature (GMT), and daily maps of sea-level pressure (SLP), soil moisture (SM), and geopotential height (GPH). Then, we train multiple CNNs to predict daily TMAX across a range of past and future climates using climate model simulations of the 1850 to 2100 period as training data (Fig. 1A). Since there are various factors that can influence the predictions of trained CNNs, we train multiple different CNNs to explore the sensitivity of our results to randomness in the CNN training process and differences between the climate model simulations used as training data. Next, to understand how a historical extreme event is influenced by anthropogenic climate forcing, we use these trained CNNs to create dynamically consistent counterfactual versions of the event by using the SLP, SM, and GPH fields from a historical reanalysis dataset as inputs to the CNNs and predicting TMAX at various different levels of GMT (ranging between 0.0° and 4.0°C above the 1850 to 1900 mean GMT) (Fig. 1B). Lastly, by comparing these counterfactual TMAX predictions across different levels of GMT, we estimate the sensitivity of the frequency and intensity of the event to changes in the GMT anomaly (relative to the 1850 to 1900 baseline levels of GMT) (Fig. 1C). In the following sections, we evaluate this approach for machine learning–based extreme event attribution by applying this technique to one recent extreme event (southcentral North America 2023) and multiple historical extreme events for which attribution assessments have been published using other methods.

RESULTS

Event attribution for the June 2023 heat wave in southcentral North America

We train CNNs (Fig. 1) to predict daily TMAX over a region in southcentral North America using data from two different general circulation models (GCMs)—CanESM5 and UKESM1-0-LL—from the Coupled Model Intercomparison Project phase 6 (CMIP6) database. For each GCM, we construct separate training datasets using data from five GCM simulations of the 1850 to 2100 period, with three GCM simulations used for CNN training, one GCM simulation used for CNN validation, and one GCM simulation used for CNN testing (see Materials and Methods). On each of these training datasets, we then train three separate CNNs using different random seeds to explore the sensitivity of our results to randomness

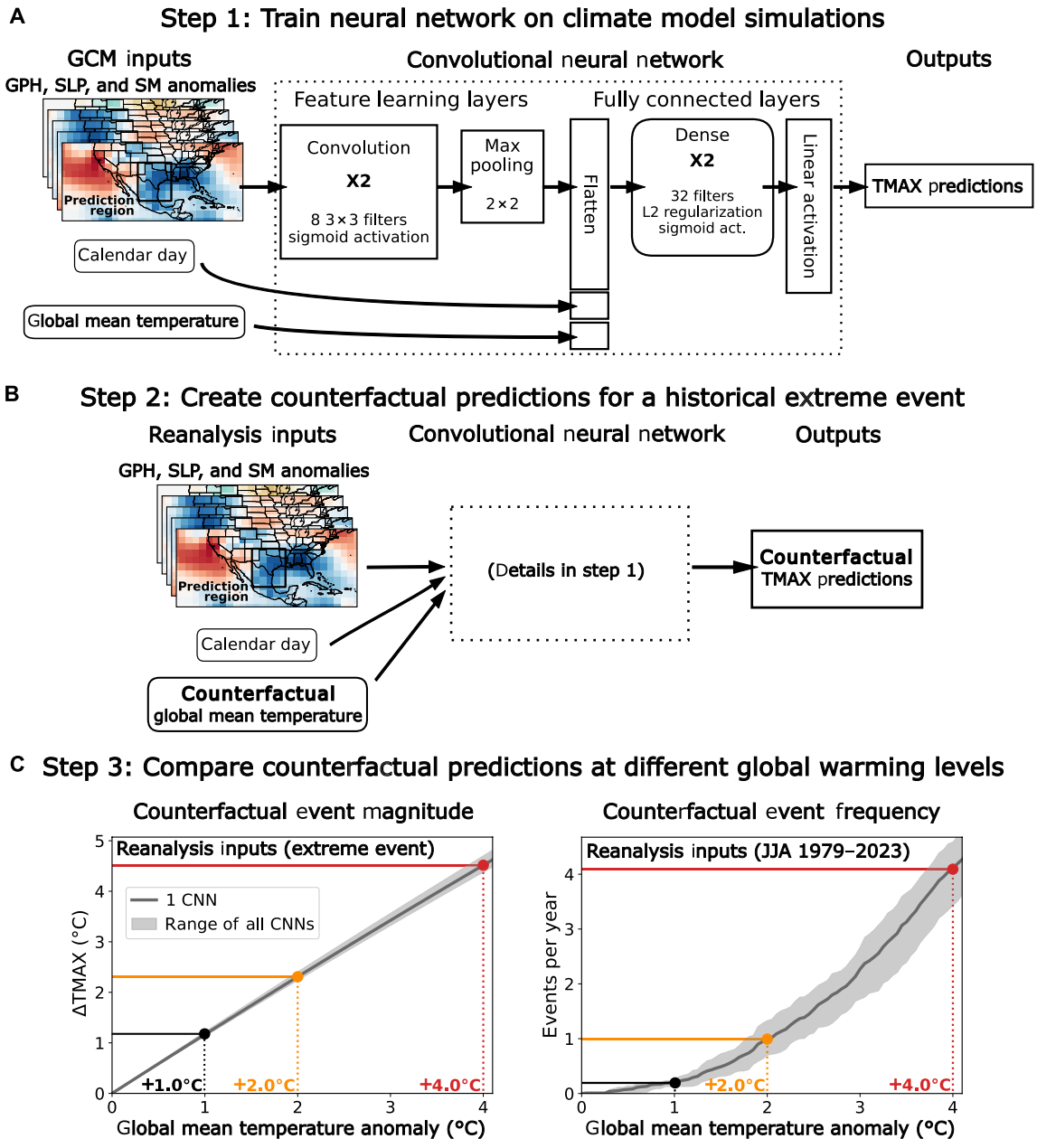

Fig. 1. Schematic of machine learning–based approach for extreme event attribution. (**A**) A CNN is trained to predict the daily TMAX over the prediction region (black box) using the following inputs: the calendar day, the annual GMT, and two-dimensional calendar-day anomaly maps of SLP, SM, and GPH at 700, 500, and 250 mbar. Three CNNs are trained on data from GCM simulations for the 1850 to 2100 period. See Materials and Methods for details of the CNN architecture and training process. (**B**) The trained CNNs are used to predict TMAX for a historical extreme event using the input variables from a historical reanalysis. Then, counterfactual TMAX predictions for this extreme event are obtained by letting the GMT input vary between +0.0° and +4.0°C relative to the 1850 to 1900 baseline period (i.e., the GMT anomaly). (**C**) (Left) Change in the counterfactual CNN TMAX prediction as a function of the GMT anomaly calculated using the reanalysis input maps from the historical extreme event. (Right) The number of extreme events [hotter than the event from (B)] as a function of the GMT anomaly calculated using the counterfactual CNN predictions from the reanalysis input maps during the June to August 1979 to 2023 period. Shown are the results for one individual CNN (dark gray line) and the range of results from three CNNs (gray shading) each trained with a different random seed.

in the CNN training process. We find that the CNNs are able to predict daily TMAX with an average R^2 value of 0.99 and an average root-mean-squared error (RMSE) of 0.75°C on the unseen GCM test datasets (averaged across six total CNNs from the two GCM training datasets) (fig. S1). We then evaluate the performance of these trained CNNs on unseen data from the European Center for Medium-Range Weather Forecasting Reanalysis fifth-generation historical reanalysis dataset (ERA5) and find that the CNNs trained on CanESM5 data achieve an average R^2 value of 0.96 and an average RMSE value of 1.31°C, while the CNNs trained on UKESM1-0-LL data achieve an average R^2 value of 0.94 and an average RMSE value of 1.62°C (fig. S1).

Since the specific meteorological conditions during an event are an essential component of the extreme event (*29*), we carefully evaluate whether the daily ERA5 input maps (SLP, GPH, and SM) improve the CNN's TMAX predictions (fig. S2). To isolate the contribution of the daily ERA5 input maps to the overall CNN skill, we compare the CNN performance with the original daily input maps against a baseline threshold for CNN performance obtained by setting all pixel values in the daily input maps to the long-term mean value for each grid cell (i.e., zero), which ensures that there is no daily variation in the meteorological conditions. We find a 52.8% reduction in the mean RMSE when we include the original meteorological input maps rather than the maps that have been replaced with the grid-cell mean (fig. S2). This suggests that, rather than ignoring the daily input maps and only using information from the GMT and calendar-day input variables to predict TMAX, the CNNs use the daily ERA5 meteorological input maps to explain daily-scale variations in ERA5 TMAX that are caused by short-term variations in weather patterns.

After training and evaluating CNN performance, we use the trained CNNs to analyze the June 2023 heat wave over southcentral North America (Fig. 2). This event reached its peak intensity on 19 to 28 June 2023 (fig. S3), during which the 10-day mean TMAX (36.13°C) was the hottest 10-day mean TMAX over this region in the ERA5 dataset (1979 to 2023). Our CNNs closely reproduce the temporal evolution and magnitude of this event, predicting a mean TMAX of 36.19°C averaged across all six CNNs (with individual CNNs ranging from 35.30° to 37.29°C). Moreover, for all six of the CNNs, this event contains the first or second hottest 10-day mean TMAX in June to August (JJA) within the entire time series of CNN predictions (1979 to 2023) (Fig. 2A). This confirms that our CNNs closely reproduce the magnitude and historical ranking of this event, conveying confidence that these CNNs can be used to quantify the sensitivity to changes in GMT.

Using partial dependence analysis [see Materials and Methods; (*56*)], we calculate the sensitivity of the CNN's TMAX prediction to the GMT input value. We find that the magnitude of this extreme

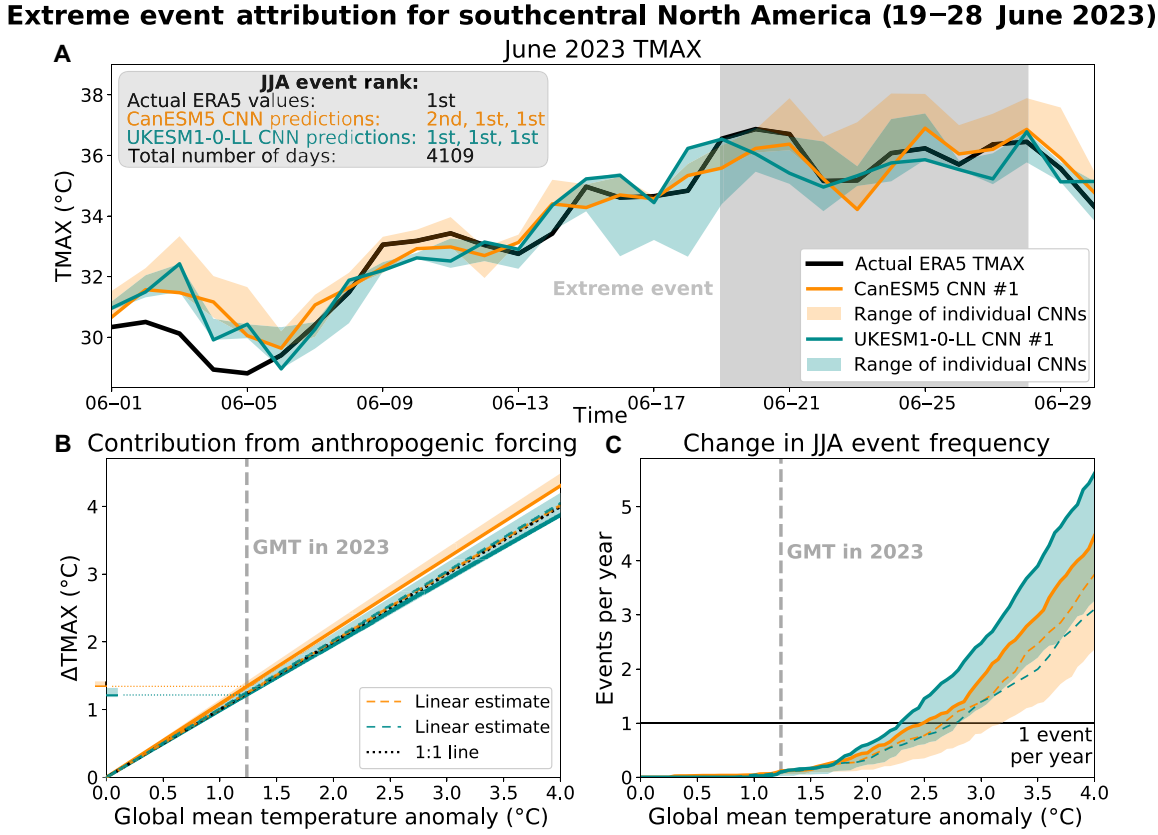

Fig. 2. Machine learning–based extreme event attribution for southcentral North America. (**A**) Daily TMAX predicted by CNNs during June 2023 (gold/green). Results for CNNs trained on CanESM5 simulations are shown in gold, and results for CNNs trained on UKESM1-0-LL simulations are shown in green. Solid lines show results for one individual CNN, and shading reveals the range of CNN predictions across three CNNs each trained with a different random seed. Actual TMAX values from the European Center for Medium-Range Weather Forecasting Reanalysis fifth generation historical reanalysis (ERA5) are also shown (black). (Gray box) Comparison between the event ranking in the distribution of actual ERA5 TMAX data and the event ranking in the distribution of CNN predictions (for each CNN used in this analysis). Event rank is calculated by counting the number of distinct 10-day periods in June to August (JJA) 1979 to 2023 in which the 10-day average TMAX surpasses the average TMAX during the 19 to 28 June 2023 event. (**B**) Change in the mean counterfactual TMAX prediction during the 19 to 28 June 2023 event as a function of the GMT input value, for values ranging between +0.0° and +4.0°C relative to the 1850 to 1900 baseline period (i.e., the GMT anomaly). (**C**) The number of 10-day periods hotter than the June 2023 event as a function of the GMT anomaly calculated using the counterfactual CNN predictions from the ERA5 input maps during the JJA 1979 to 2023 period. Dashed lines in (B) and (C) show the estimated change in event magnitude and frequency predicted by a linear extrapolation of the ERA5 TMAX values to different levels of GMT according to the calendar-day linear trend in TMAX (calculated with respect to GMT) from the CanESM5 and UKESM1-0-LL realizations.

heat event increases roughly linearly as a function of the GMT anomaly from the 1850 to 1900 baseline period (Fig. 2B), which agrees with results from the most recent IPCC report (*4*). Our results suggest that anthropogenic forcing since 1850 increased the magnitude of the 2023 event by 1.18° to 1.42°C (where the range indicates the spread across all six CNNs used in this analysis). Likewise, these results also suggest that the same meteorological conditions would produce an event 2.65° to 3.07°C hotter than the 2023 event if they were to occur in a climate with a GMT anomaly of 4.0°C (Fig. 2B).

We also use the daily JJA reanalysis inputs from 1979 to 2023 to calculate the frequency of 10-day periods in which the counterfactual CNN predictions are hotter than the 19 to 28 June 2023 period. We find that the frequency increases nonlinearly as a function of the GMT anomaly (Fig. 2C). These results suggest that the same daily JJA meteorological conditions from 1979 to 2023 would produce 0.0 events per year hotter than the June 2023 event at a GMT anomaly of 0.0°C, 0.14 to 0.60 events per year at 2.0°C, and 2.36 to 5.62 events per year at 4.0°C (Fig. 2C).

For comparison, we also estimate the change in event magnitude and frequency obtained using a linear extrapolation of the ERA5 TMAX values to different levels of GMT according to the calendar-day linear trend in TMAX with respect to GMT (calculated from the GCM training data). For the UKESM1-0-LL CNNs, we find that the results for our counterfactual CNN predictions are similar to the estimates obtained from the linear extrapolation method (Fig. 2, B and C). In contrast, all three of our CanESM5 CNNs predict a larger change in event magnitude compared to the linear extrapolation approach (Fig. 2B). However, this does not result in a larger change in event frequency for the CanESM5 CNNs compared to the linear estimates (as we would expect if the CanESM5 CNNs were to systematically predict a larger warming rate across all days) (Fig. 2C). The difference between these methods suggests the possibility that the CNNs predict different warming rates on different days, resulting in nonuniform changes in the CNN-predicted TMAX distributions with increasing GMT.

Evaluating patterns in CNN-based attribution for southcentral North America

We further analyze the CNNs trained over southcentral North America to test whether the CNN-predicted sensitivity to increasing GMT is based only on a basic linear regional scaling of TMAX with increasing GMT or if the daily meteorological input maps influence the CNN's daily attribution estimate. We find that the difference in counterfactual TMAX predictions between GMT anomaly values of 0.0° and 4.0°C (i.e., the daily contribution from anthropogenic forcing at 4.0°C) exhibits a strong annual cycle, with the calendar-day mean varying between 3.5° and 4.8°C throughout the year (Fig. 3A). This suggests that the CNNs have learned seasonal differences in the warming rate of regional TMAX with increasing GMT. We also find that southcentral North America CNNs predict day-to-day variations away from the calendar-day mean annual cycle of contribution at 4.0°C (i.e., "contribution anomalies"), ranging from −0.18° to +0.22°C across all JJA days and −0.12° to +0.18°C across the hottest 10% of JJA days (Fig. 3B). This suggests that the relationship between the GMT input variable and daily TMAX that the CNNs have learned from the GCM training dataset changes depending on time of year and the daily meteorological conditions.

To identify what might be causing these day-to-day variations in the CNN-predicted contributions at 4.0°C (Fig. 3B), we compare the average daily meteorological conditions on days with the 10% highest and 10% lowest contribution anomalies on JJA T90 days (Fig. 3, C to H). We find statistically significant differences ($P < 0.01$; based on a Student's t test) between the daily meteorological patterns (TMAX, 500-mbar GPH, and SM anomalies) associated with the highest and lowest daily contribution anomalies on the hottest 10% of JJA days (JJA "T90" days) in the ERA5 dataset (Fig. 3, C to H). (Results for SLP, 250-mbar GPH, and 700-mbar GPH are shown separately; fig. S4.) More specifically, we find that the days with the highest 10% of contribution anomalies tend to have a positive 500-mbar GPH anomaly situated over southwestern North America and a negative 500-mbar GPH anomaly situated over the southeastern US. This 500-mbar GPH pattern is conducive to north-northwesterly flow into southcentral North America originating from continental regions north of the prediction region (Fig. 3D). In contrast, days with the lowest 10% of contribution anomalies tend to have a negative 500-mbar GPH anomaly situated over the western US and a positive 500-mbar GPH anomaly situated over the Gulf of Mexico (and surrounding regions), which is conducive to south-southwesterly flow into southcentral North America originating from over the Pacific Ocean (Fig. 3G). This suggests that the differences in daily contribution anomalies predicted by the CNNs are influenced by consistent physical differences in the daily meteorological input maps.

Event attribution for historical extreme events

As a further evaluation of our machine learning–based attribution approach, we analyze a number of historical extreme events that have been extensively studied using other attribution methods [e.g., (*7*, *19*, *59*)]. These include the southern Europe heat wave in August 2003 (fig. S5), the western Russia heat wave in August 2010 (fig. S6), and the western India heat wave in March 2022 (fig. S7), all of which were either the hottest or second hottest event of the 1979 to 2023 period over the respective regions (Figs. 4 to 6). Before conducting our full attribution analysis, we compare the TMAX predicted by the CNNs using the ERA5 meteorological input fields with the actual ERA5 TMAX during each event (Figs. 4A, 5A, and 6A). The mean TMAX predictions for each of the six CNNs fall within 0.29° to 2.50°C of the actual TMAX in southern Europe, within 0.37° to 2.85°C in western Russia, and within 0.38° to 1.07°C in western India. We also find that 12 of the 18 total CNNs are able to exactly reproduce the historical event ranking and that all 18 CNNs rank these events as either the first, second, or third most extreme events over the analysis period (Fig. 4A, 5A, and 6A).

Like the 2023 event in southcentral North America, we find that the predicted magnitude of the 2003 southern Europe heat wave increases roughly linearly with increasing GMT, while the predicted event frequency increases nonlinearly (Fig. 4, B and C). These results suggest that anthropogenic forcing (since 1850) increased the magnitude of the 2003 southern Europe event by 1.43° to 1.84°C and that these same meteorological conditions would produce an event 5.38° to 6.33°C hotter than the 2003 event if they were to occur in a climate with a GMT anomaly of 4.0°C above the 1850–1900 baseline period (Fig. 4B). Furthermore, our counterfactual results suggest that the same daily JJA atmospheric and land-surface conditions from 1979 to 2023 would produce 0.0 events per year hotter than the August 2023 event at a GMT anomaly of 0.0°C, 0.14 to 0.86 events per year at 2.0°C, and 3.41 to 6.27 events per year at 4.0°C (Fig. 4C).

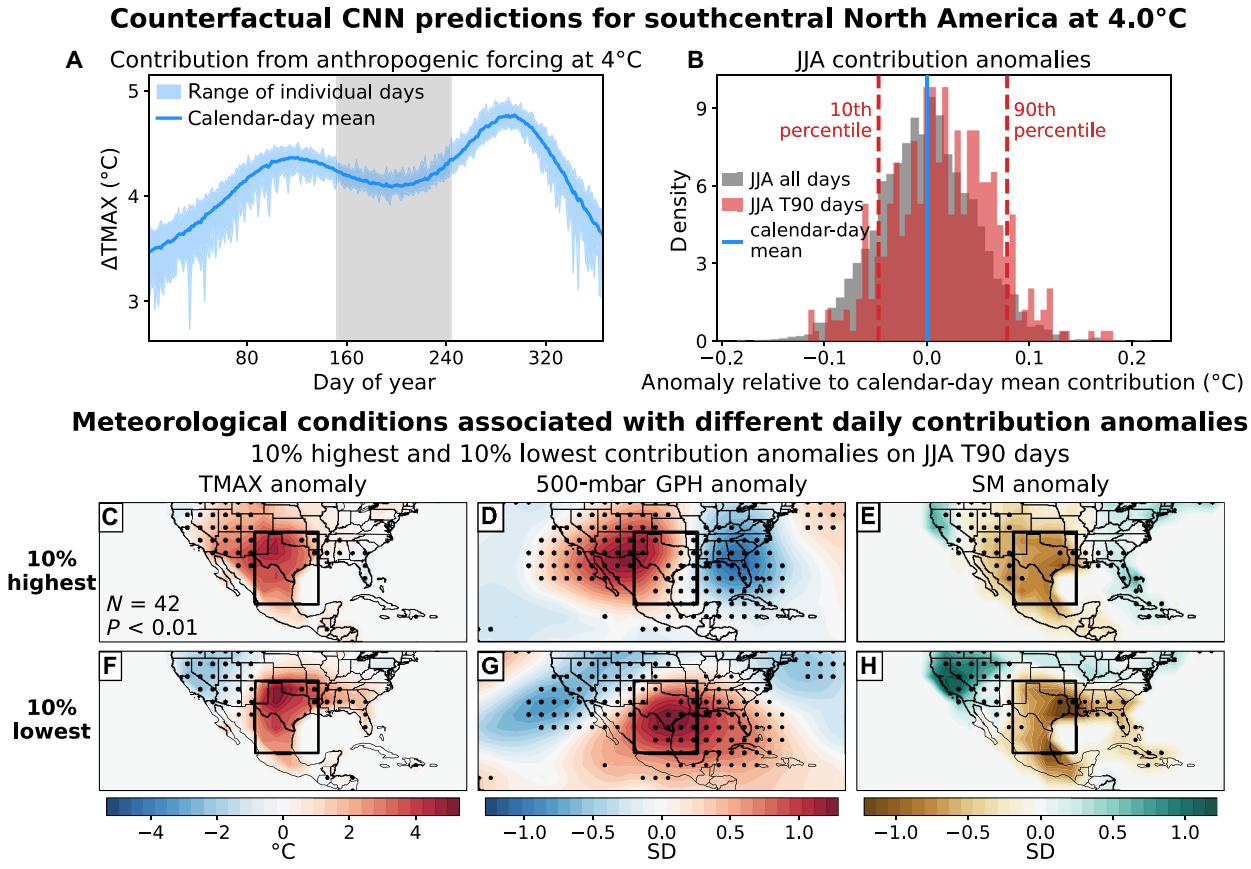

Fig. 3. Analysis of counterfactual predictions for CNNs trained over southcentral North America. (**A**) The calendar-day mean contribution from anthropogenic forcing at 4.0°C (calculated as the difference in counterfactual CNN predictions between GMT anomalies of 0.0° and 4.0°C) for all days in the period 1979 to 2023 (blue line), along with the range of contributions on individual days (shading). (**B**) Distribution of daily anomalies from the calendar-day mean contribution [from (A)] for all June to August (JJA) days (gray) and for the hottest 10% of JJA days (i.e., JJA T90 days, red). Also shown are the thresholds for the JJA T90 days with 10% highest and 10% lowest contribution anomalies (dashed). (**C** to **E**) Mean calendar-day anomaly maps for (C) daily TMAX, (D) 500-mbar GPH, and (E) SM on the JJA T90 days with the 10% highest contribution anomalies. (**F** to **H**) As in (C) to (E) but for the JJA T90 days with the 10% lowest contribution anomalies. Stippling indicates where there is a statistically significant difference (Student's *t* test with $P < 0.01$, $N = 42$) between the mean anomaly maps for the highest and lowest contribution anomalies. Results in (A) to (H) are averaged across all six trained CNNs.

Similarly, for the 2010 western Russia heat wave, the predicted event magnitude also increases linearly with increasing GMT, while the predicted event frequency increases nonlinearly (Fig. 5, B and C). Our counterfactual results suggest that anthropogenic forcing increased the magnitude of the 2010 western Russia event by 1.51° to 2.16°C, and these same meteorological conditions would produce an event 4.19° to 5.54°C hotter than the 2010 event if they were to occur in a climate with a GMT anomaly of 4.0°C above the 1850 to 1900 baseline period (Fig. 5B). These results also suggest that the same daily JJA atmospheric and land-surface conditions from 1979 to 2023 would produce 0.0 events per year hotter than the August 2010 event at a GMT anomaly of 0.0°C, 0.05 to 0.11 events per year at 2.0°C, and 0.41 to 0.87 events per year at 4.0°C (Fig. 5C).

Lastly, for the 2022 western India heat wave, the predicted event magnitude again increases linearly with increasing GMT, while the predicted event frequency increases nonlinearly (Fig. 6, B and C). Using counterfactual TMAX predictions, we estimate that anthropogenic forcing increased the magnitude of the 2022 western India event by 1.20° to 1.71°C, and these same meteorological conditions would produce temperatures 2.82° to 3.98°C hotter than the 2022 event if they were to occur in a climate with a GMT anomaly of 4.0°C above the 1850 to 1900 baseline period (Fig. 6B). This event was especially anomalous because it was a prolonged heat event that occurred in March (before the typical April to May pre-monsoon warm season) when daily TMAX typically falls in the range 28° to 34°C. Therefore, to assess the influence of anthropogenic forcing on the frequency of similar early-season heat events, we only consider 18-day events that occur before 01 April [i.e., January to March (JFM)]. Our counterfactual results suggest that the same daily JFM atmospheric and land-surface conditions from 1979 to 2023 would produce 0.0 events per year hotter than the March 2022 event at a GMT anomaly of 0.0°C, 0.06 to 0.15 events per year at 2.0°C, and 0.37 to 0.74 events per year at 4.0°C (Fig. 6C).

Similar to our analysis of the southcentral North America CNNs (Fig. 3), we also analyze which factors influence the relationship between TMAX and GMT for the CNNs trained on southern Europe (Fig. 4, D to K), western Russia (Fig. 5, D to K), and western India (Fig. 6, D to K). We conclude that the strength of the relationship between daily TMAX and GMT is influenced by the region, the time of year, and the daily meteorological conditions. Specifically, we find

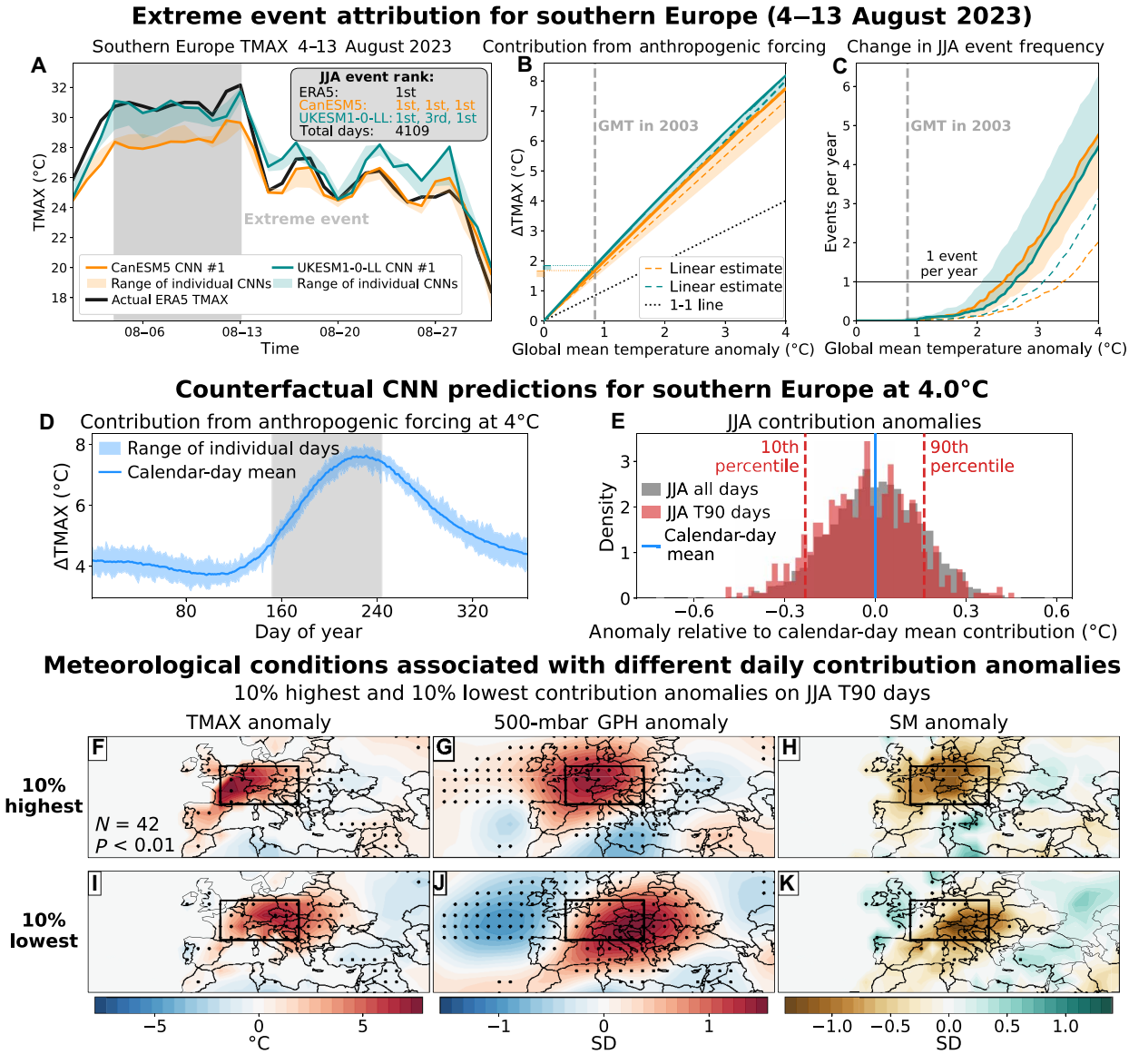

Fig. 4. Machine learning–based extreme event attribution for southern Europe. (**A** to **C**) As in Fig. 2 but for the extreme heat event that occurred in southern Europe on 4 to 13 August 2003. (**D** to **K**) As in Fig. 3 but for the CNNs trained to predict daily TMAX over southern Europe.

that the difference between CNN TMAX predictions at GMT = 4.0°C and GMT = 0.0°C (i.e., the anthropogenic contribution at 4.0°C) exhibits a strong annual cycle, varying between 3.7° and 7.6°C in southern Europe (Fig. 4D), 5.3° and 7.5°C in western Russia (Fig. 5D), and 3.3° and 4.9°C in western India (Fig. 6D). These regional differences in the magnitude of the annual cycle are consistent with the linear trends in seasonal TMAX estimated from the CanESM5 and UKESM1-0-LL simulations (fig. S8). Aside from the annual cycle, we also find that these regions experience day-to-day variations away from the calendar-day mean annual cycle of contribution at 4.0°C (i.e., contribution anomalies) that vary from −0.72° to +0.59°C in southern Europe JJA (Fig. 4E), −0.95° to +0.76°C in western Russia JJA (Fig. 5E), and −0.26° to +0.27°C in western India JFM (Fig. 6E).

To identify which daily meteorological patterns are associated with the highest and lowest daily contribution anomalies during a specific time of year, we compare the average daily meteorological conditions on days with the 10% highest and 10% lowest contribution anomalies for JJA T90 days in southern Europe (Fig. 4, F to K), for JJA T90 days in western Russia (Fig. 5, F to K), and for JFM T90 days in western India (Fig. 6, F to K). In each of these regions, we find statistically significant differences ($P < 0.01$; based on a Student's t test) between the daily meteorological patterns (TMAX, 500-mbar GPH, and SM anomalies) associated with the highest and lowest daily contribution anomalies. (Results for SLP, 250-mbar GPH, and 700-mbar GPH are shown separately; fig. S9). Similar to the results from southcentral North America (Fig. 3G), we find that the T90 days associated with lower daily warming rates have 500-mbar GPH patterns conducive to south-southwesterly flow originating from over the nearby oceans in southern Europe (Fig. 4J) and western India (Fig. 6J). In addition, for western India, we find that

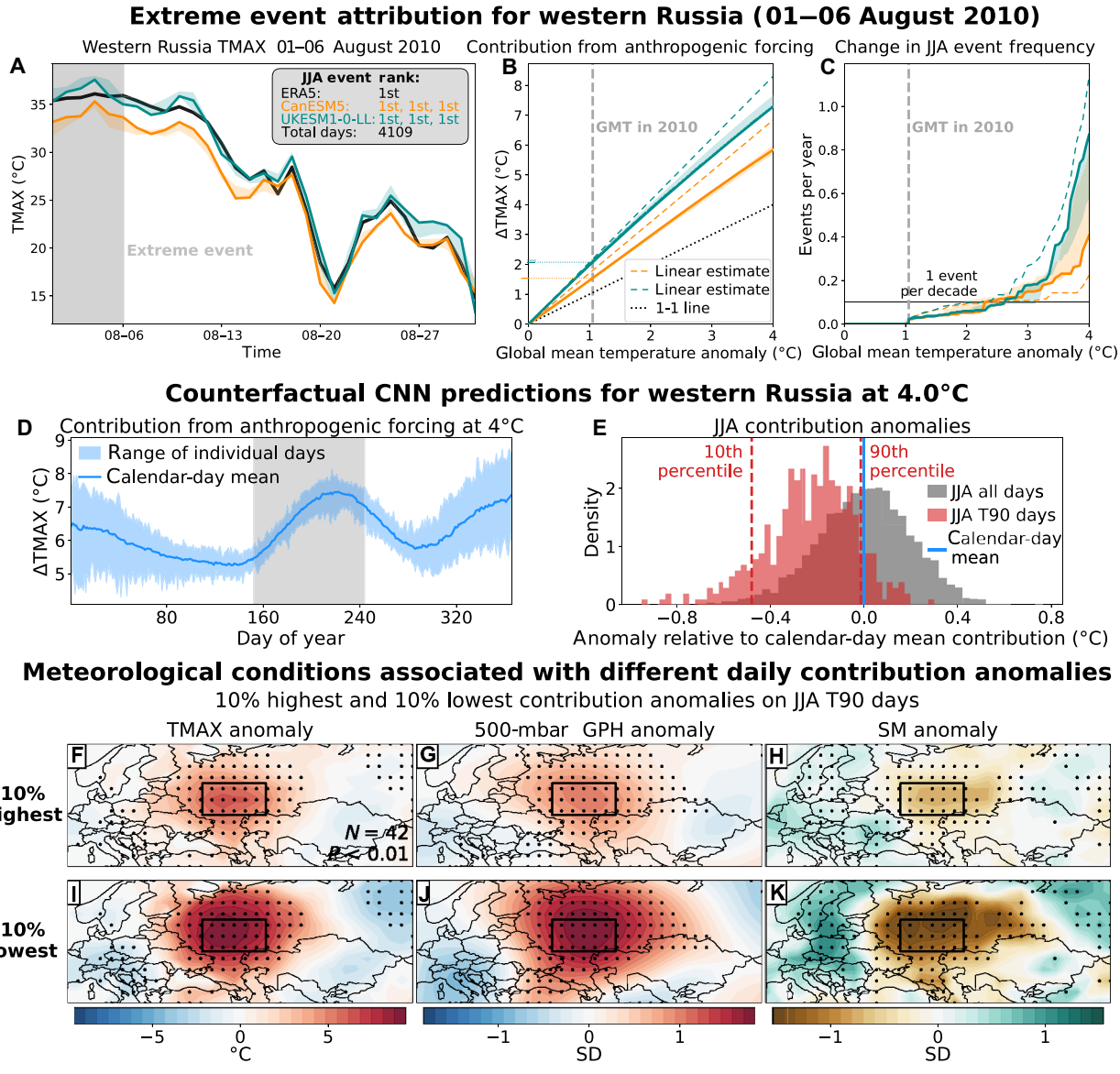

Fig. 5. Machine learning–based extreme event attribution for western Russia. As in Fig. 4 but for the extreme heat event that occurred in western Russia on 01 to 06 August 2010.

the T90 days associated with higher daily warming rates have 500-mbar GPH patterns conducive to north-northwesterly flow originating from the continental regions north of the prediction region (Fig. 6G), which is similar to the behavior observed in southcentral North America (Fig. 3D). In contrast, the meteorological conditions in western Russia (Fig. 5, F to K)—which is located in the interior of the Eurasian continent—do not exhibit the same patterns observed in the other three regions.

DISCUSSION

Our attribution results (Figs. 2 and 4 to 6) agree broadly with the "collective attribution" (*6*) findings in the most recent IPCC report, which indicate that annual maximum daily temperatures over the global land surface increase roughly linearly with increasing GMT (*4*), while the frequency of extreme heat events increases nonlinearly (*1*). This is clearly seen in our analysis of the recent June 2023 heat wave in southcentral North America, for which our counterfactual CNN predictions suggest a roughly linear relationship between TMAX and GMT (including an estimated contribution of 1.18° to 1.42°C at a GMT anomaly of ~1.2°C above the 1850 to 1900 baseline and an additional contribution of 2.65° to 3.07°C at 4.0°C; Fig. 2B) and a nonlinear relationship between event frequency and GMT (considering 10-day events hotter than the June 2023 heat wave; Fig. 2C). We also show that differences in daily meteorological conditions explain differences in the CNN-based attribution estimates on summer days over southcentral North America (Fig. 3), suggesting that the influence of anthropogenic forcing on daily TMAX differs on the basis of the underlying meteorological conditions. Specifically, we find that this region experiences higher rates of

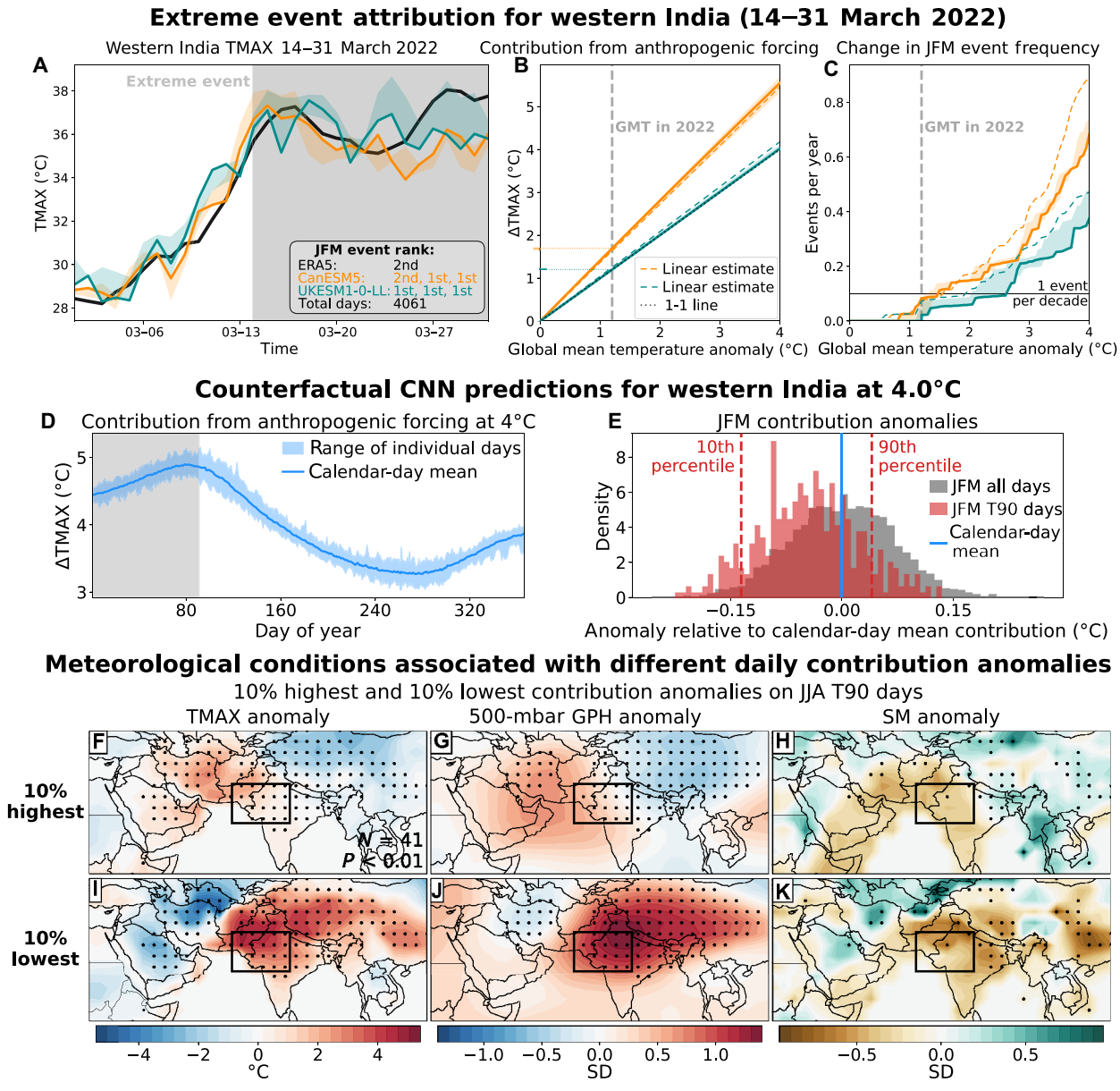

Fig. 6. Machine learning–based extreme event attribution for western India. As in Fig. 4 but for the extreme heat event that occurred in western India on 14 to 31 March 2022.

warming on days with 500-mbar GPH patterns conducive to north-northwesterly flow originating from central North America (Fig. 3D) and lower rates of warming on days with 500-mbar GPH patterns conducive to south-southwesterly flow originating from over the Pacific Ocean (Fig. 3G). These results thus demonstrate an additional potential benefit of our machine learning–based attribution technique, which allows for the attribution estimates to be influenced by the actual daily meteorological conditions that occur during an extreme event, leading to nonuniform changes in the CNN-predicted TMAX distribution with increasing GMT. Although there is some observational evidence suggesting that different groups of days experience different rates of warming [e.g., extreme cold days are warming faster than extreme hot days, (*4*)], more research is needed to better understand how the rate of warming may be different for different meteorological patterns (*31*).

Comparison with previously published results

As part of the evaluation of our method, we perform machine learning–based extreme event attribution analyses for several historical extreme heat events that have been previously studied using established attribution techniques (Fig. 4 to 6). Although different attribution studies often use different analysis periods, analysis regions, and attribution metrics, we find that our results broadly agree with previous studies. For the 2003 southern Europe event, we find that anthropogenic forcing contributed 1.43° to 1.84°C to the overall magnitude of the event (Fig. 4B). In comparison, by analyzing 2000

regional climate model simulations with and without anthropogenic forcing, Mitchell *et al.* (*60*) found that anthropogenic forcing contributed 1.0°C to the mean temperature across continental Europe during JJA 2003. For the 2010 western Russia event, our results suggest that anthropogenic climate change since 1850 contributed 1.51° to 2.16°C to the overall magnitude of the event (and that climate change since 1980 has contributed 0.77° to 1.09°C) (Fig. 5B). This result is consistent with Wehrli *et al.* (*61*), who used regional climate model experiments to show that recent climate change (since 1980) contributed about 1.2°C to the temperature anomaly in western Russia during the period 15 July to 14 August 2010. Similarly, for the 2022 western India event, our results suggest that anthropogenic forcing contributed 1.20° to 1.71°C to the overall magnitude of the event and that these same meteorological conditions would produce an event 0.80° to 1.15°C warmer than the 2022 event if they were to occur in a climate with a GMT anomaly of 2.0°C above the 1850 to 1900 baseline (Fig. 6B). These findings broadly agree with Zachariah *et al.* (*19*), which applied a probability-based framework to show that anthropogenic forcing contributed about 1.0°C (95% confidence interval of 0.2° to 2.1°C) to the March to April 2022 mean TMAX in this region, with an additional contribution of 1.0°C (95% confidence interval of 0.3° to 1.7°C) expected if this event were to occur under 2.0°C of global warming.

We also apply our technique to calculate the influence of anthropogenic forcing on the frequency of extreme events using the daily meteorological conditions from all seasonal days in the 1979 to 2023 reanalysis as input to the CNNs. For each of these three historical events, we find a nonlinear increase in event frequency with increasing GMT (Fig. 4C, 5C, and 6C). In particular, our counterfactual TMAX predictions suggest that these historical events (which are the first or second most extreme events in the 1979 to 2023 period) will increase nonlinearly in frequency as the GMT anomaly increases to 4.0°C above the 1850 to 1900 baseline, occurring 3.41 to 6.27 times per year for the southern Europe event, 0.41 to 0.87 times per year for the western Russia event, and 0.37 to 0.74 times per year for the western India event.

There are some challenges in quantitatively comparing our counterfactual frequency results at 0°C of global warming against previous extreme event attribution assessments that used different frequency metrics to attribute changes in event probability between the preindustrial and the present. For example, in what is widely considered to be the original attribution study, Stott *et al.* (*7*) used generalized Pareto distributions to estimate the influence of anthropogenic forcing on the probability of the 2003 Europe event, concluding with >90% confidence that human-caused climate change doubled the probability of extreme temperatures. Likewise, Rahmstorf and Coumou (*59*) estimated an ~80% probability that the record-breaking July 2010 monthly mean temperature in western Russia occurred as a result of climate change. Similarly, Zachariah *et al.* (*19*) used the World Weather Attribution framework (*17*) to determine that anthropogenic climate change made the March to April 2022 event in western India ~30 times more likely than it would have been under preindustrial levels of climate forcing.

In theory, we could quantitatively compare our frequency results with these previous attribution studies by calculating the ratio between our CNN-based predictions of event frequency at current levels of GMT and preindustrial levels of GMT. However, while our machine learning–based approach enables a more comprehensive quantification of the sensitivity of event frequency to changes in GMT compared with strict storyline approaches, the limited sample size available in the historical reanalysis data (i.e., 4109 JJA days and 4061 JFM days in the ERA5 dataset) poses a major limitation in estimating the frequency of extreme events under preindustrial levels of GMT. For example, our counterfactual results suggest a frequency of 0.0 events per year at a GMT anomaly of 0.0°C for each extreme event in this analysis, although the probability of these extreme events is likely nonzero at preindustrial levels of GMT (*62*).

We emphasize that our approach—which uses reanalysis data as out-of-sample input for CNNs trained on GCM simulations—is based on the actual daily meteorological conditions that occurred during the recent historical period. For the event intensity, this has the advantage of other storyline techniques in that our estimate for each individual event is based on the actual meteorological conditions during that event. In contrast, this means that our CNN-based counterfactuals can only provide a partial assessment of attribution (*21*) because we do not allow for possible climate-driven changes in atmospheric dynamics to influence our daily CNN TMAX predictions. For example, our CNN predictions of event frequency (which are based on the exact population of daily meteorological conditions that occurred in 1979 to 2023) provide an incomplete assessment of attribution (*21*) because they do not account for potential changes in the likelihood of extreme atmospheric circulation patterns over the analysis region [e.g., (*31*, *33*)]. We therefore urge caution in using our frequency results to calculate precise changes in probability ratios at very low GMT anomaly values. However, since climate-driven changes in the frequency of atmospheric circulation patterns are highly uncertain and poorly understood (*35*), we present this CNN-based prediction approach for estimating event frequency as an alternative to techniques that rely on raw GCM-simulated event frequencies.

Generalizability to other extreme events

Our results suggest that this machine learning–based attribution framework may be generalizable to other extreme heat events, other types of extreme weather, and other regions of the world. To further probe the limitations of our approach, we also examine two cases in which the trained CNNs exhibit substantial biases relative to the original ERA5 TMAX values. Given these biases, we conclude that these CNNs should not be used to quantify attribution metrics for the respective extreme events. As a first cautionary example, we consider the Pacific Northwest heat wave that occurred in June 2021 (fig. S10). This extreme event reached its peak intensity on 27 to 30 June (fig. S11), during which the average ERA5 TMAX was 35.80°C, which is the single hottest 4-day period over this region in the ERA5 reanalysis. Our six CNNs substantially underestimate the magnitude of this event, predicting TMAX to be 4.32° to 7.15°C lower than the actual ERA5 reanalysis TMAX. Moreover, the six CNNs rank this event as the 3rd, 12th, 14th, 38th, 50th, and 60th hottest 4-day events in the entire time series of CNN predictions for the 1979 to 2023 period (fig. S10A). Since our CNNs cannot accurately represent the magnitude and historical rank of this extreme event, we do not recommend using these CNNs to perform an attribution analysis of this event, as the CNN's underrepresentation of the event magnitude and rank may lead to unrealistic attribution results. For example, for the three CNNs trained on the UKESM1-0-LL simulations, our framework suggests the Pacific Northwest region should experience an average of 1.11 to 1.88 events per year at a GMT anomaly of 1.0°C above the 1850 to 1900 baseline (fig. S10C). In

reality, this event is far more rare in the current climate (*63*), suggesting that the biased CNNs do indeed generate attribution results that are unrealistic. We do not find a large bias in the mean climatology of TMAX in the GCM datasets used to train the CNNs for the Pacific Northwest analysis (fig. S12A). This suggests that the bias in the CNN predictions for the 2021 Pacific Northwest event likely results from the complexity of the physical processes driving this particular extreme event. Climate model experiments suggest that temperatures during the 2021 Pacific Northwest heat wave were enhanced by multi-day heat accumulation facilitated by the combination of a strong omega block, deep atmospheric boundary layers, and upwind latent heating in the days leading up to the event (*64*). The absence of detailed information about these processes (e.g., boundary layer height and latent heating) and their temporal evolution in our daily CNN input maps may lead to inaccurate predictions when the trained CNNs are confronted with the ERA5 input maps from the actual event (fig. S11).

As a second cautionary example, we evaluate the ability of the CNNs used to analyze the March 2022 event in western India (Fig. 6) to predict JJA TMAX over the same region. These CNNs are trained using input from all calendar days, and although they perform quite well when predicting daily TMAX in January to April of 2022, they exhibit a pronounced bias when predicting daily TMAX in June to August of 2022 (e.g., +6.9°C relative to the ERA5 reanalysis for CanESM5 CNNs), making them unreliable for attribution analyses during the summer season (fig. S13B). Furthermore, we find that a similar bias in the annual cycle of TMAX is also present in the GCM datasets used to train these CNNs (fig. S13A), suggesting that the bias was inherited from the GCMs during the CNN training process. These two cautionary examples suggest that CNN prediction biases can result from either insufficient information about complex physical processes in the daily input maps or from biases in the mean climate of the GCM training data. Since biases in CNN TMAX predictions may lead to unrealistic attribution assessments (fig. S10), we emphasize the need to evaluate the ability of the CNNs to predict the actual extreme event before moving forward with machine learning–based counterfactual analysis.

Further, generalizing beyond extreme heat events, this technique could also be used to make attribution assessments for other types of extreme events (e.g., heavy precipitation and droughts), provided that sufficient GCM training data are available and the trained CNNs are able to accurately predict the key characteristics of the event. As an example, we apply this machine learning–based attribution framework to analyze extreme precipitation in the US Pacific Northwest (fig. S14). Following the same methodology as the heat wave analysis (with a few minor changes; see the Supplementary Materials for details), we determine that anthropogenic climate change since 1850 has increased the magnitude of the extreme precipitation event that occurred on 03 December 2007, by 1.0 to 2.4% (fig. S14B). Our counterfactual precipitation predictions also suggest an increase in the frequency of similar extreme events with increasing GMT (fig. S14C), although our ability to estimate changes in the frequency of precipitation extremes is limited by the long, thin tail of the precipitation distribution and the limited sample size of historical weather conditions (fig. S14E). More generally, our CNNs predict that rainfall intensity on extreme wet days (above the 99th percentile; P99 days) will increase at an average rate of ~3.2% for each 1.0°C increase in GMT (with individual P99 days ranging between 0.6 and 4.8% per °C) (fig. S14D). We also find that the frequency of P99 days is expected to increase from 1.96 days per year at a GMT anomaly of 0.0°C to 2.96 days per year at 2.0°C, and 4.50 days per year at 4.0°C (fig. S14E). Although we have not shown that our framework is transferable to all types of extreme events, this analysis of extreme precipitation suggests that our approach may be used to conduct attribution analyses across a wider range of extremes than just temperature.

Last, the computational efficiency of our approach is an important feature that makes our framework generalizable to different types of extreme events, different regions of the world, and different attribution timescales (e.g., rapid attribution). First, the ability to efficiently train CNNs for different events and regions using large ensembles of existing climate model simulations greatly enhances the generalizability, provided that the networks are able to predict the key characteristics of the event with sufficient accuracy. In addition, the computational efficiency of this technique also raises the possibility that our approach could be used for rapid attribution, which aims to have the attribution analysis available as close as possible to the occurrence of an extreme event (*17*, *65*). While most storyline-based attribution approaches are not used for rapid attribution assessments unless a large amount of computing resources are available [e.g., (*25*)], our approach can efficiently create multiple counterfactual realizations of an individual extreme event without requiring expensive additional climate model simulations. The main temporal limitation of the analysis we present here is the availability of ERA5 input maps, which are released with a delay of about 5 days (*66*). Although a separate CNN needs to be trained for each analysis region, we find that different regional CNNs can be trained using the same model architecture. As a result, the CNNs used in this study do not require extensive computational resources or training time (~2 hours on 12 CPUs). To further speed up this analysis, these CNN models can be pretrained (using previously simulated GCM data) over any region of the globe and rapidly implemented as soon as the reanalysis input maps become available. These advances also raise the possibility that this technique could contribute to operational extreme event attribution (*65*, *67*) by using weather forecasting data as inputs for pretrained CNNs to perform attribution analyses before, during, and/or immediately following an extreme event.

Additional sources of uncertainty

In this study, we explore a number of the potential sources of uncertainty in our machine learning–based approach to extreme event attribution. For example, we compare the spread of results across three individual CNNs to assess the sensitivity of our results to randomness in the CNN training process. In addition, we have shown that our attribution results may be influenced by biases between the GCMs and the ERA5 reanalysis (fig. S13). These issues can be largely avoided by evaluating the ability of the CNNs to reproduce the magnitude and historical ranking of an extreme event before producing attribution results. Despite these precautions, biases in the GCM training datasets create an additional source of uncertainty in our CNN-based attribution results. Therefore, to probe the range of results caused by differences between the GCM training datasets, we compare attribution results between CNNs trained using data from two different GCMs (CanESM5 and UKESM1-0-LL). In addition, to reduce biases in the CNN predictions, CNNs may be trained using the GCMs that have the smallest bias for the particular variables and analysis region, or by using bias-corrected GCM datasets [e.g., (*68*),

although data availability for the full period from the preindustrial to present is limited]. Alternatively, different historical reanalysis datasets could also be used to ensure that GCM biases are not specific to one particular reanalysis product [e.g., (*69*)]. We also show that CNNs may under-predict the magnitude of heat waves in which the temperatures are enhanced by complex physical processes that are not well-represented in the CNN input data (e.g., the 2021 Pacific Northwest heat wave; fig. S10). This issue may be avoided by including additional input variables (such as latent heat release) from several days leading up to an extreme event to allow the CNN to learn more complex physical relationships that can enhance temperatures during an event (*64*). In addition, biases in the CNN predictions may be reduced by increasing the number of GCM realizations in the CNN training dataset to provide a larger sample of meteorological conditions from which the CNNs are able to learn, or by using ensemble boosting [which uses initial condition ensembles to generate additional simulations of individual extreme events in existing GCM simulations, (*65*)] to artificially increase the frequency of extreme circulation events in the GCM training datasets.

Future research

While the initial results from our machine learning–based attribution approach are promising, more analysis is needed before these results can be used for high-stakes applications such as improving adaptation decisions, attributing climate damages, and informing climate litigation. In particular, a thorough assessment of the other potential sources of uncertainty is needed. These include quantifying uncertainty using an even larger ensemble of CNNs (each with different model architectures, hyperparameters, and training outcomes), comparing CNNs trained on a larger number of GCMs (subject to availability of the necessary input variables at daily timescales), and incorporating GCM training data from different future emission scenarios (e.g., net-zero scenarios such as SSP1-2.6 and SSP1-1.9). Although the computational efficiency of this attribution framework makes our approach a promising tool for quantifying uncertainty in storyline attribution analyses, designing and testing appropriate architectures will require further work.

Our machine learning–based technique presents potential advances in the field of extreme event attribution. In addition to demonstrating the general potential for machine learning–based attribution techniques, we show specifically that CNNs trained on GCM simulations can be used to create counterfactual versions of extreme events using the actual daily meteorological conditions to make out-of-sample predictions of event magnitude under different levels of GMT. Using the actual meteorological conditions in a truly predictive framework increases the fidelity of the calculated counterfactual intensities and frequencies. In addition, the computational efficiency of this approach increases generalizability to different extreme events, different regions, and different attribution timescales (e.g., rapid attribution). Together, this initial study suggests that our machine learning–based extreme event attribution approach is a promising tool that can be used for rapid, low-cost attribution analysis of individual extreme events. By demonstrating the potential for machine learning–based extreme event attribution, we hope to open future research into additional applications of machine learning to better understand the influence of historical and future climate change on different types of extreme events in different regions of the world.

MATERIALS AND METHODS

Experimental design

We develop a framework for using machine learning techniques to evaluate the contribution of human-caused climate change to individual extreme events (Fig. 1) and apply that framework to multiple recent and historical extreme heat events. To perform this analysis, we first train multiple CNNs to predict daily TMAX across a range of past and future climates using training data from an individual GCM over the 1850 to 2100 period (Fig. 1A). Then, we use these trained CNNs to predict TMAX during an individual historical extreme event using unseen input data from the ERA5 historical reanalysis dataset. After confirming that these trained CNNs accurately reproduce the magnitude and historical ranking of the actual extreme event when given historical reanalysis data as inputs, we use partial dependence analysis to create counterfactual versions of the extreme event under different levels of annual GMT (Fig. 1B). By calculating the sensitivity of our counterfactual CNN predictions to the GMT input value, we quantify the contribution of anthropogenic forcing to the event magnitude and estimate the sensitivity of event frequency to changes in GMT (conditional on the daily meteorological conditions that occurred during the season of interest from 1979 to 2023) (Fig. 1C). We repeat this analysis using training data from two different GCMs to explore how structural differences between GCMs might affect the results.

CNN training datasets

We construct CNN training datasets that use several climate variables simulated by a GCM as predictors of regional average daily TMAX. Each of these training datasets consists of the following input variables: daily SLP, daily GPH on three pressure surfaces (700, 500, and 250 millibars), daily SM from the 0- to 10-cm layer, calendar day (normalized between 0 and 1), and the GMT averaged over the previous 12 months (normalized between 0 and 1). We selected this set of daily input maps (SLP, GPH, and SM) to provide daily information about the state of the large-scale atmospheric circulation and the top-layer of the land surface; however, we do not claim that this is the optimal set of input variables for this prediction task.

To create these training datasets, we use an ensemble of CMIP6 simulations under historical greenhouse gas forcing (from 1850 to 2014) and the IPCC's "very high" future emissions scenario (SSP5-8.5; from 2015 to 2100). We use the very high emissions scenario to ensure that our training data contain extreme events that occur across the broadest range of GMTs. Although there are 39 CMIP6 GCMs with simulations of the historical and SSP5-8.5 scenarios, our analysis requires three-dimensional atmospheric output available at daily time steps from multiple realizations of the 1850 to period. This requirement substantially limits the availability of GCMs for this analysis. Therefore, for this initial application of our technique, we use realizations from each of two CMIP6 GCMs for which large ensembles of daily data are available (CanESM5 and UKESM1-0-LL). From each of these GCM datasets, we select five realizations with perturbed initial conditions that were simulated with the same forcing and physics configurations. To explore the sensitivity of our results to differences between the GCMs, we construct a separate CNN training dataset for each of these two GCMs. For each GCM dataset, we download GPH (at 700, 500, and 250 mbar), SM (from the upper portion of the soil column), SLP, and TMAX at daily timescales from five different realizations of the 1850 to 2100 period. We use bilinear interpolation to convert the daily GPH, SM, and SLP

input maps to a common rectangular grid with 2.8125° × 2.8125° horizontal resolution (128 longitude points × 64 latitude points). We also calculate the area-weighted global average of monthly near-surface air temperature for each GCM realization (1850 to 2100) and compute the 12-month moving average to obtain each month's annual GMT value. Since this GMT calculation requires monthly temperature data from the preceding 12 months, we use the daily GCM data from 1851 to 2100 (and the monthly GCM temperature data from 1850 to 2100) to train our CNNs.

Before CNN training, we remove the grid-cell calendar-day linear trends in GPH, SLP, and SM calculated with respect to the GMT for each GCM realization (fig. S15). This detrending removes any linear correlation between the GMT input and the GPH/SLP/SM input maps caused by nonuniform thermal dilation of the troposphere [for GPH; (*31*)] and/or trends in precipitation and evapotranspiration [for SM; (*1*)] and ensures that the linear signal from anthropogenic climate change has been removed from all neural network input variables except the GMT input. Therefore, the CNN will rely on the daily GPH, SLP, and SM anomalies to explain daily-scale TMAX variability, while the GMT input variable is used to account for the effects of anthropogenic forcing on TMAX (driven by long-term trends in air temperature, humidity, thermal dilation, SM, etc.). To improve the training process, we normalize the GPH, SLP, and SM inputs by subtracting the grid-cell calendar-day mean and dividing by the grid-cell calendar-day SD for each GCM simulation. We also scale the GMT inputs into the range 0 to 1 by subtracting the minimum GMT and dividing by the range of GMT values from the entire 1850 to 2100 period. After processing this training data, we split the GCM datasets into training, validation, and testing subsets, using three GCM realizations (i.e., three simulations of 1850 to 2100 at daily timescales) for CNN training, one GCM realization for CNN validation, and one GCM realization for CNN testing (which is left out of the training process entirely).

To evaluate our machine learning–based attribution approach, we analyze one very recent extreme heat event (southcentral North America in June 2023) and several historical extreme heat events that have been studied using established attribution techniques: southern Europe in 2003 (*7*); western Russia in 2010 (*61*); western India in 2022 (*19*); and the US Pacific Northwest in 2021 (*63*). We create separate CNN training datasets for each extreme event in this study and then train CNNs to predict the daily TMAX averaged over all non-ocean grid cells in the analysis region.

The analysis regions for each heat event in this study are defined as follows: southcentral North America (21° to 37°N, −106° to −92°E), southern Europe (43° to 53°N, 0° to 20°E), western Russia (51° to 59°N, 37° to 53°E), western India (20° to 30°N, 65° to 80°E), and the US Pacific Northwest (41° to 51°N, −125° to −115°E). To give sufficient spatial context for each TMAX prediction, we use broad spatial input maps (roughly 35° latitudinally and 85° longitudinally) for the GPH/SLP/SM inputs centered on each analysis region (*34*, *58*).

CNN architecture

We train CNNs (Fig. 1A) to predict daily TMAX over the analysis region using training data from several GCM simulations over the 1850 to 2100 period (see details of training datasets in the CNN training datasets section above). For each CNN, the two-dimensional input maps (i.e., GPH, SLP, and SM) are passed through two convolutional layers (eight 3 by 3 filters with sigmoid activation) followed by a 2 by 2 max pooling layer. The resulting feature vector is then flattened into a one-dimensional vector, and the normalized calendar-day and GMT inputs are concatenated to the end. This vector is passed through two dense layers (32 filters with sigmoid activation), after which a final linear activation outputs the daily TMAX prediction. During the training process, we pass the daily GCM inputs into a CNN and compare the model's TMAX prediction against the actual daily TMAX (calculated from the true GCM output) and adjust the CNN's weights and biases to minimize the loss function (mean squared error) on the GCM validation dataset. Multiple model architectures and hyperparameter combinations were tested to identify an architecture that performs well on the GCM validation dataset (tuning results not shown). To avoid issues caused by the nonuniform TMAX distributions, we use the DenseWeight algorithm (*70*) to modify the loss function during the training process (using sample weights inversely proportional to the sample frequencies). We use Tensorflow with Keras 2.7.0 (*71*) to construct and train each CNN model.

Extreme event attribution using CNNs

After training the CNNs on GCM data (see details in sections CNN training datasets and CNN architecture above), we make out-of-sample TMAX predictions for individual extreme events in the historical record using the actual daily meteorological conditions from the ERA5 historical reanalysis dataset as input maps for the trained CNNs. We download the ERA5 dataset (*72*) over the period January 1979 through July 2023 and aggregate the hourly ERA5 data to obtain daily maps of SLP, SM content (from 0- to 7-cm soil layer), and GPH (at 700-, 500-, and 250-mbar levels). Similarly, we download the ERA5-Land dataset (*73*) over the same 1979 to 2023 period and aggregate the hourly ERA5 data to obtain daily maps of SM content (from 0- to 7-cm soil layer). Using bilinear interpolation, we convert the daily SLP, SM, and GPH input maps from the ERA5 dataset to a rectangular grid with 2.8125° × 2.8125° horizontal resolution (128 longitude points × 64 latitude points) to match the GCM training data. We also calculate the area-weighted global average of the monthly ERA5 2-m temperature data and compute the 12-month moving average to obtain each month's GMT value. Since this GMT calculation relies on monthly temperature data from the preceding 12 months, we include monthly ERA5 temperature data from 1978 to calculate the monthly GMT values for 1979. Similar to the GCM training datasets, we remove the grid-cell calendar-day linear trends in the ERA5 GPH, SLP, and SM fields (calculated with respect to the ERA5 GMT) and normalize the daily input maps by subtracting the ERA5 grid-cell calendar-day mean and dividing by the ERA5 grid-cell calendar-day SD. We also scale the ERA5 GMT inputs into the range 0 to 1 by subtracting the minimum GCM GMT and dividing by the range of GCM GMT values (calculated across all five GCM realizations of the 1850 to 2100 period). We use the area-weighted average values for daily TMAX over the analysis region as our target data to assess the performance of the neural network on the unseen ERA5 dataset.

To evaluate the performance of our CNNs on the ERA5 dataset, we compare the event magnitude and event rank for each extreme event between the CNN TMAX predictions and the actual ERA5 TMAX values. We calculate event rank by counting the number of distinct periods in the TMAX time series that are more extreme than the extreme event (in terms of magnitude and/or duration).

Then, to determine how anthropogenic forcing may influence the frequency and intensity of individual extreme events, we use

partial dependence analysis (*56*) to calculate counterfactual TMAX values for individual days in the historical record under different levels of GMT (Fig. 1B). To calculate these counterfactual TMAX predictions, we use the daily ERA5 input maps from each individual day in the extreme event and let the GMT input vary across a range of annual values observed in the GCM simulations (i.e., a discrete set of GMT values uniformly distributed—at 0.05°C increments—from +0.0° to +4.0°C relative to the 1850 to 1900 GCM mean across all realizations). Then, we pass these input combinations through the trained CNN to quantify how the daily TMAX predictions change as a function of GMT for the actual meteorological conditions that occurred during the extreme event (Fig. 1C). We repeat this process across all samples in the ERA5 dataset (1979 to 2023) from the 3-month season in which the event occurred (e.g., June to August) to generate a counterfactual time series of ERA5 TMAX values for each level of GMT. We then analyze these counterfactual TMAX time series to quantify the sensitivity of extreme event frequency to changes in the GMT (conditional on the actual daily meteorological conditions that occurred from 1979 to 2023). For example, given a 10-day event in boreal summer with a mean TMAX of 30°C, we obtain the event frequency in the counterfactual June to August time series by counting the number of periods at least 10 days in duration in which the mean TMAX is greater than or equal to 30°C, counting fractional events when necessary (i.e., an 11-day period will be counted as 11/10 = 1.1 events).

To examine some of the potential sources of uncertainty in this analysis, we train three separate CNNs for each analysis region (with each CNN trained using a different random seed). Comparing the spread of results across these three individual CNNs (Fig. 1C) allows us to quantify some of the possible variation in our results caused by randomness during the CNN training process. In addition, we repeat this analysis using two different GCM training datasets (described in section CNN training datasets above) to explore the sensitivity of our results to differences between GCMs. This gives us a total of six CNNs (three trained on each GCM dataset) to use for the attribution analysis of each extreme event.

Supplementary Materials

This PDF file includes:

Supplementary Text
Figs. S1 to S15

REFERENCES AND NOTES

1. V. P. Masson-Delmotte, P. Zhai, S. L. Pirani, C. Connors, S. Péan, N. Berger, Y. Caud, L. Chen, M. I. Goldfarb, P. M. Scheel Monteiro, IPCC, 2021: Summary for policymakers, 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* (2021); https://doi.org/10.1017/9781009157896.001.
2. D. R. Easterling, G. A. Meehl, C. Parmesan, S. A. Changnon, T. R. Karl, L. O. Mearns, Climate extremes: Observations, modeling, and impacts. *Science* **289**, 2068–2074 (2000).
3. 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).
4. S. I. Seneviratne, X. Zhang, M. Adnan, W. Badi, Chapter 11: Weather and climate extreme events in a changing climate in *IPCC Sixth Assessment Report Working Group 1: The Physical Science Basis* (IPCC, 2021).
5. E. M. Fischer, S. Sippel, R. Knutti, Increasing probability of record-shattering climate extremes. *Nat. Clim. Chang.* **11**, 689–695 (2021).
6. D. L. Swain, D. Singh, D. Touma, N. S. Diffenbaugh, Attributing extreme events to climate change: A new frontier in a warming world. *One Earth* **2**, 522–527 (2020).
7. P. A. Stott, D. A. Stone, M. R. Allen, Human contribution to the European heatwave of 2003. *Nature* **432**, 610–614 (2004).
8. P. A. Stott, N. Christidis, F. E. L. Otto, Y. Sun, J.-P. Vanderlinden, G. J. van Oldenborgh, R. Vautard, H. von Storch, P. Walton, P. Yiou, F. W. Zwiers, Attribution of extreme weather and climate-related events. *Wiley Interdiscip. Rev. Clim. Change* **7**, 23–41 (2016).
9. M. Hulme, Attributing weather extremes to "climate change". *Prog. Phys. Geogr.* **38**, 499–511 (2014).
10. National Academies of Sciences, Engineering, and Medicine, Division on Earth and Life Studies, Board on Atmospheric Sciences and Climate, Committee on Extreme Weather Events and Climate Change Attribution, *Attribution of Extreme Weather Events in the Context of Climate Change* (National Academies Press, 2016); https://play.google.com/store/books/details?id=WWEpDQAAQBAJ.
11. M. D. Risser, M. F. Wehner, Attributable human-induced changes in the likelihood and magnitude of the observed extreme precipitation during hurricane Harvey. *Geophys. Res. Lett.* **44**, 12457–12464 (2017).
12. J. S. Risbey, D. B. Irving, D. T. Squire, R. J. Matear, D. P. Monselesan, M. J. Pook, N. Ramesh, D. Richardson, C. R. Tozer, A large ensemble illustration of how record-shattering heat records can endure. *Environ. Res. Climate* **2**, 035003 (2023).
13. 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).
14. K. Emanuel, Assessing the present and future probability of Hurricane Harvey's rainfall. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12681–12684 (2017).
15. D. E. Rupp, S. Li, P. W. Mote, N. Massey, S. N. Sparrow, D. C. H. Wallom, Influence of the ocean and greenhouse gases on severe drought likelihood in the Central United States in 2012. *J. Climate* **30**, 1789–1806 (2017).
16. S. F. Kew, S. Y. Philip, G. J. van Oldenborgh, G. van der Schrier, F. E. L. Otto, R. Vautard, The exceptional summer heat wave in Southern Europe 2017. *Bull. Am. Meteorol. Soc.* **100**, S49–S53 (2019).
17. S. Philip, S. Kew, G. J. van Oldenborgh, F. Otto, R. Vautard, K. van der Wiel, A. King, F. Lott, J. Arrighi, R. Singh, M. van Aalst, A protocol for probabilistic extreme event attribution analyses. *Adv. Stat. Clim. Meteorol. Oceanogr.* **6**, 177–203 (2020).
18. 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).
19. M. Zachariah, T. Arulalan, K. AchutaRao, F. Saeed, R. Jha, M. K. Dhasmana, A. Mondal, R. Bonnet, R. Vautard, S. Philip, S. Kew, M. Vahlberg, R. Singh, J. Arrighi, D. Heinrich, L. Thalheimer, C. P. Marghidan, A. Kapoor, M. van Aalst, E. Raju, S. Li, J. Sun, G. Vecchi, W. Yang, M. Hauser, D. L. Schumacher, S. I. Seneviratne, L. J. Harrington, F. E. L. Otto, Attribution of 2022 early-spring heatwave in India and Pakistan to climate change: lessons in assessing vulnerability and preparedness in reducing impacts. *Environ. Res. Climate* **2**, 045005 (2023).
20. 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).
21. T. G. Shepherd, A common framework for approaches to extreme event attribution. *Curr. Clim. Change Rep.* **2**, 28–38 (2016).
22. A. D. Jones, D. Rastogi, P. Vahmani, A. M. Stansfield, K. A. Reed, T. Thurber, P. A. Ullrich, J. S. Rice, Continental United States climate projections based on thermodynamic modification of historical weather. *Sci. Data* **10**, 664 (2023).
23. D. Rastogi, F. Lehner, M. Ashfaq, Revisiting recent U.S. heat waves in a warmer and more humid climate. *Geophys. Res. Lett.* **47**, e2019GL086736 (2020).
24. K. A. Reed, A. M. Stansfield, M. F. Wehner, C. M. Zarzycki, Forecasted attribution of the human influence on Hurricane Florence. *Sci. Adv.* **6**, eaaw9253 (2020).
25. K. A. Reed, M. F. Wehner, Real-time attribution of the influence of climate change on extreme weather events: A storyline case study of Hurricane Ian rainfall. *Environ. Res. Climate* **2**, 043001 (2023).
26. P. Yiou, R. Vautard, P. Naveau, C. Cassou, Inconsistency between atmospheric dynamics and temperatures during the exceptional 2006/2007 fall/winter and recent warming in Europe. *Geophys. Res. Lett.* **34**, L21808 (2007).
27. J. Cattiaux, R. Vautard, C. Cassou, P. Yiou, V. Masson-Delmotte, F. Codron, Winter 2010 in Europe: A cold extreme in a warming climate. *Geophys. Res. Lett.* **37**, L20704 (2010).
28. R. García-Herrera, J. M. Garrido-Perez, D. Barriopedro, C. Ordóñez, S. M. Vicente-Serrano, R. Nieto, L. Gimeno, R. Sorí, P. Yiou, The European 2016/17 drought. *J. Climate* **32**, 3169–3187 (2019).
29. K. E. Trenberth, J. T. Fasullo, T. G. Shepherd, Attribution of climate extreme events. *Nat. Clim. Chang.* **5**, 725–730 (2015).
30. W. Li, L. Li, M. Ting, Y. Liu, Intensification of Northern Hemisphere subtropical highs in a warming climate. *Nat. Geosci.* **5**, 830–834 (2012).
31. D. E. Horton, N. C. Johnson, D. Singh, D. L. Swain, B. Rajaratnam, N. S. Diffenbaugh, Contribution of changes in atmospheric circulation patterns to extreme temperature trends. *Nature* **522**, 465–469 (2015).
32. D. L. Swain, D. E. Horton, D. Singh, N. S. Diffenbaugh, Trends in atmospheric patterns conducive to seasonal precipitation and temperature extremes in California. *Sci. Adv.* **2**, e1501344 (2016).

33. E. Rousi, K. Kornhuber, G. Beobide-Arsuaga, F. Luo, D. Coumou, Accelerated western European heatwave trends linked to more-persistent double jets over Eurasia. *Nature* **13**, 3851 (2022).
34. F. V. Davenport, N. S. Diffenbaugh, Using machine learning to analyze physical causes of climate change: A case study of U.S. midwest extreme precipitation. *Geophys. Res. Lett.* **48**, e2021GL093787 (2021).
35. T. G. Shepherd, Atmospheric circulation as a source of uncertainty in climate change projections. *Nat. Geosci.* **7**, 703–708 (2014).
36. O. Bellprat, F. Doblas-Reyes, Attribution of extreme weather and climate events overestimated by unreliable climate simulations. *Geophys. Res. Lett.* **43**, 2158–2164 (2016).
37. S. Sippel, N. Meinshausen, A. Merrifield, F. Lehner, A. G. Pendergrass, E. Fischer, R. Knutti, Uncovering the forced climate response from a single ensemble member using statistical learning. *J. Climate* **32**, 5677–5699 (2019).
38. E. A. Barnes, B. Toms, J. W. Hurrell, I. Ebert-Uphoff, C. Anderson, D. Anderson, Indicator patterns of forced change learned by an artificial neural network. *J. Adv. Model. Earth Syst.* **12**, e2020MS002195 (2020).
39. S. Sippel, N. Meinshausen, E. Székely, E. Fischer, A. G. Pendergrass, F. Lehner, R. Knutti, Robust detection of forced warming in the presence of potentially large climate variability. *Sci. Adv.* **7**, eabh4429 (2021).
40. Y.-G. Ham, J.-H. Kim, S.-K. Min, D. Kim, T. Li, A. Timmermann, M. F. Stuecker, Anthropogenic fingerprints in daily precipitation revealed by deep learning. *Nature* **622**, 301–307 (2023).
41. A. R. Gottlieb, J. S. Mankin, Evidence of human influence on Northern Hemisphere snow loss. *Nature* **625**, 293–300 (2024).
42. N. S. Diffenbaugh, E. A. Barnes, Data-driven predictions of the time remaining until critical global warming thresholds are reached. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2207183120 (2023).
43. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**, 436–444 (2015).
44. Y.-G. Ham, J.-H. Kim, J.-J. Luo, Deep learning for multi-year ENSO forecasts. *Nature* **573**, 568–572 (2019).
45. V. Jacques-Dumas, F. Ragone, P. Borgnat, P. Abry, F. Bouchet, Deep learning-based extreme heatwave forecast. *Front. Clim.* 10.3389/fclim.2022.789641 , (2021).
46. E. Racah, C. Beckham, T. Maharaj, S. E. Kahou, M. Prabhat, C. Pal, Extreme weather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. *Adv. Neural Inf. Process. Syst.* **30**, 3405–3416 (2017).
47. D. J. Gagne II, S. E. Haupt, D. W. Nychka, G. Thompson, Interpretable deep learning for spatial analysis of severe hailstorms. *Mon. Weather Rev.* **147**, 2827–2845 (2019).
48. K. K. Prabhat, M. Mudigonda, S. Kim, L. Kapp-Schwoerer, A. Graubner, E. Karaismailoglu, L. von Kleist, T. Kurth, A. Greiner, A. Mahesh, K. Yang, C. Lewis, J. Chen, A. Lou, S. Chandran, B. Toms, W. Chapman, K. Dagon, C. A. Shields, T. O'Brien, M. Wehner, W. Collins, ClimateNet: An expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather. *Geosci. Model Dev.* **14**, 107–124 (2021).
49. B. Pan, K. Hsu, A. AghaKouchak, S. Sorooshian, Improving precipitation estimation using convolutional neural network. *Water Resour. Res.* **55**, 2301–2321 (2019).
50. L. Sun, Y. Lan, Statistical downscaling of daily temperature and precipitation over China using deep learning neural models: Localization and comparison with other methods. *Int. J. Climatol.* **41**, 1128–1147 (2021).
51. Y. Han, G. J. Zhang, X. Huang, Y. Wang, A moist physics parameterization based on deep learning. *J. Adv. Model. Earth Syst.* **12**, e2020MS002076 (2020).
52. T. Bolton, L. Zanna, Applications of deep learning to ocean data inference and subgrid parameterization. J. *Adv. Model. Earth Syst.* **11**, 376–399 (2019).
53. S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, W. Samek, On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLOS ONE* **10**, e0130140 (2015).
54. A. Mamalakis, E. A. Barnes, I. Ebert-Uphoff, Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience. *Artif. Intell. Earth Syst.* **1**, e220012 (2022).
55. B. A. Toms, E. A. Barnes, I. Ebert-Uphoff, Physically interpretable neural networks for the geosciences: Applications to earth system variability. *J. Adv. Model. Earth Syst.* **12**, e2019MS002002 (2020).
56. J. H. Friedman, Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1536 (2001).
57. G. Zhang, M. Wang, K. Liu, Deep neural networks for global wildfire susceptibility modelling. *Ecol. Indic.* **127**, 107735 (2021).
58. J. T. Trok, F. V. Davenport, E. A. Barnes, N. S. Diffenbaugh, Using machine learning with partial dependence analysis to investigate coupling between soil moisture and near-surface temperature. *J. Geophys. Res.* **128**, e2022JD38365 (2023).
59. S. Rahmstorf, D. Coumou, Increase of extreme events in a warming world. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 17905–17909 (2011).
60. D. Mitchell, C. Heaviside, S. Vardoulakis, C. Huntingford, G. Masato, B. P. Guillod, P. Frumhoff, A. Bowery, D. Wallom, M. Allen, Attributing human mortality during extreme heat waves to anthropogenic climate change. *Environ. Res. Lett.* **11**, 074006 (2016).
61. K. Wehrli, B. P. Guillod, M. Hauser, M. Leclair, S. I. Seneviratne, Identifying key driving processes of major recent heat waves. *J. Geophys. Res.* **124**, 11746–11765 (2019).
62. C. Gessner, E. M. Fischer, U. Beyerle, R. Knutti, Very rare heat extremes: Quantifying and understanding using ensemble reinitialization. *J. Climate* **34**, 6619–6634 (2021).
63. E. Bercos-Hickey, T. A. O'Brien, M. F. Wehner, L. Zhang, C. M. Patricola, H. Huang, M. D. Risser, Anthropogenic contributions to the 2021 pacific northwest heatwave. *Geophys. Res. Lett.* **49**, e2022GL099396 (2022).
64. D. L. Schumacher, M. Hauser, S. I. Seneviratne, Drivers and mechanisms of the 2021 pacific northwest heatwave. *Earth's Future* **10**, e2022EF002967 (2022).
65. E. M. Fischer, U. Beyerle, L. Bloin-Wibe, C. Gessner, V. Humphrey, F. Lehner, A. G. Pendergrass, S. Sippel, J. Zeder, R. Knutti, Storylines for unprecedented heatwaves based on ensemble boosting. *Nat. Commun.* **14**, 4643 (2023).
66. M. F. Wehner, K. A. Reed, Operational extreme weather event attribution can quantify climate change loss and damages. *PLOS Clim.* **1**, e0000013 (2022).
67. H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. De Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, C. Lupu, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume, J.-N. Thépaut, The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.* **146**, 1999–2049 (2020).
68. Z. Xu, Y. Han, C.-Y. Tam, Z.-L. Yang, C. Fu, Bias-corrected CMIP6 global dataset for dynamical downscaling of the historical and future climate (1979–2100). *Sci. Data* **8**, 1–11 (2021).
69. D. E. Horton, C. B. Skinner, D. Singh, N. S. Diffenbaugh, Occurrence and persistence of future atmospheric stagnation events. *Nat. Clim. Chang.* **4**, 698–703 (2014).
70. M. Steininger, K. Kobs, P. Davidson, A. Krause, A. Hotho, Density-based weighting for imbalanced regression. *Mach. Learn.* **110**, 2187–2211 (2021).
71. T. Developers, *TensorFlow* (2021); https://zenodo.org/record/5593257.
72. H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Horányi, J. Muñoz Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, J.-N. Thépaut, ERA5 hourly data on pressure levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), (2023); https://doi.org/10.24381/cds.bd0915c6.
73. J. Muñoz-Sabater, E. Dutra, A. Agustí-Panareda, C. Albergel, G. Arduini, G. Balsamo, S. Boussetta, M. Choulga, S. Harrigan, H. Hersbach, B. Martens, D. G. Miralles, M. Piles, N. J. Rodríguez-Fernández, E. Zsoter, C. Buontempo, J.-N. Thépaut, ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **13**, 4349–4383 (2021).

Acknowledgments: We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies that support CMIP6 and ESGF. Computational resources were provided by the Stanford Research Computing Center and the Stanford Doerr School Center for Computation. **Funding:** E.A.B. was supported, in part, by the Regional and Global Model Analysis program area of the U.S. Department of Energy's (DOE) Office of Biological and Environmental Research (BER) as part of the Program for Climate Model Diagnosis and Intercomparison project. J.T.T. and N.S.D. acknowledge support from Stanford University.
Author contributions: Conceptualization: J.T.T. and N.S.D. Methodology: J.T.T., E.A.B., F.V.D., and N.S.D. Software: J.T.T. Visualization: J.T.T., E.A.B., F.V.D., and N.S.D. Writing—original draft: J.T.T. Writing—review and editing: J.T.T., E.A.B., F.V.D., and N.S.D. Funding acquisition: N.S.D. Supervision: N.S.D. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The ERA5 (*72*) and ERA5-Land (*73*) data are available from the Copernicus Climate Change Service Climate Data Store and can be accessed at from their website at https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels and https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land, respectively. The CMIP6 data used in this analysis are available from the Earth System Grid Federation and can be accessed from their website at https://esgf-node.llnl.gov/search/cmip6/. Analysis code is available in a Zenodo archive (https://doi.org/10.5281/zenodo.12745552).

Submitted 10 November 2023
Accepted 15 July 2024
Published 21 August 2024
10.1126/sciadv.adl3242