

Geophysical Research Letters®



RESEARCH LETTER

10.1029/2024GL111832

Key Points:

- We train CNNs to predict peak global warming given the map of recent annual temperatures and total additional CO₂ emissions
- Even if net-zero emissions are reached mid-century, mean warming is virtually certain to exceed 1.5°C, with even odds of 2°C
- There is high likelihood of individual years that are at least 0.5°C hotter than 2023 even in the most ambitious decarbonization scenario

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

N. S. Diffenbaugh,
diffenbaugh@stanford.edu

Citation:

Diffenbaugh, N. S., & Barnes, E. A. (2024). Data-driven predictions of peak warming under rapid decarbonization. *Geophysical Research Letters*, 51, e2024GL111832. <https://doi.org/10.1029/2024GL111832>

Received 7 AUG 2024

Accepted 5 NOV 2024

Author Contributions:

Conceptualization: Noah S. Diffenbaugh, Elizabeth A. Barnes

Data curation: Noah S. Diffenbaugh, Elizabeth A. Barnes

Formal analysis: Elizabeth A. Barnes

Methodology: Noah S. Diffenbaugh, Elizabeth A. Barnes

Software: Elizabeth A. Barnes

Visualization: Noah S. Diffenbaugh, Elizabeth A. Barnes

Writing – original draft: Noah S. Diffenbaugh

Writing – review & editing: Elizabeth A. Barnes

© 2024. The Author(s).

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](https://creativecommons.org/licenses/by/4.0/), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Data-Driven Predictions of Peak Warming Under Rapid Decarbonization

Noah S. Diffenbaugh¹  and Elizabeth A. Barnes² 

¹Doerr School of Sustainability, Stanford University, Stanford, CA, USA, ²Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA

Abstract The severe impacts associated with recent record-setting annual global temperatures elevate the need to accurately predict the hottest conditions that could occur even if the most ambitious decarbonization goals are achieved. We use convolutional neural networks (CNNs) to predict peak global warming from recent observed temperature maps and future cumulative CO₂ emissions. For the SSP1-1.9 decarbonization scenario there is >99% probability that mean global warming exceeds 1.5°C, approximately even odds that it reaches 2°C, and ~90% probability that the hottest year globally exceeds 2023 by at least 0.5°C. Further, for the SSP2-4.5 decarbonization scenario, there is >90% probability that the hottest annual global temperature anomaly is twice the 2023 anomaly. That our framework makes highly accurate out-of-sample predictions of the hottest historical year provides confidence in the predicted future probabilities, suggesting substantial risks from the extreme local conditions that are likely to result from globally hot years during rapid decarbonization.

Plain Language Summary Calendar year 2023 was the hottest year on record globally, reaching ~1.5°C above the pre-industrial. Many national, sub-national and non-state actors have articulated ambitious decarbonization goals to stabilize the global temperature. However, the intensifying impacts as individual years have approached 1.5°C have heightened the need to more accurately predict not just the mean warming but also the hottest years that could occur even in the context of rapid decarbonization. We train neural networks on an ensemble of global climate models and then use historical observations as input to the trained networks, thus constraining the uncertainty in climate model projections by using the current state of the climate system as the basis for a truly out-of-sample prediction. For the historical period, we find that, despite a wide range of climate sensitivities across global climate models, the neural networks make highly accurate predictions of the hottest historical year when given observed climate patterns as out-of-sample inputs. Predicting future warming under different cumulative emissions, we find that even if net-zero emissions are achieved mid-century, mean warming is virtually certain to exceed 1.5°C and has even odds of reaching 2°C, with high likelihood of individual years that are at least 0.5°C hotter than 2023.

1. Introduction

The United Nations Paris Agreement codified the international goals of “holding the increase in the global average temperature to well below 2°C above pre-industrial levels and pursuing efforts to limit the temperature increase to 1.5°C above pre-industrial levels” (UNFCCC, 2015), leading to ambitious decarbonization goals focused on a mid-21st-century net-zero emissions time horizon (Matthews et al., 2020; UNEP, 2023). While global emissions continue to increase, the rate of increase has slowed (Friedlingstein et al., 2023). As a result, although a gap remains with mid-century net-zero goals, the global emissions trajectory is currently consistent with scenarios in which emissions decrease during the 21st century, and is well below those in which emissions continue to increase (Friedlingstein et al., 2023; Gidden et al., 2019; Riahi et al., 2017). These developments have generated substantial interest in quantifying uncertainty in the global warming trajectory under decarbonization, including the probability that mean global warming exceeds 1.5°C or 2.0°C (e.g., IPCC, 2018; Lee et al., 2021; Rogelj et al., 2017).

Like the Paris Agreement, much of the existing research is framed around the relationship between cumulative emissions and global temperature change. However, although the first-order relationship is robust (Lee et al., 2021; Matthews & Caldeira, 2008; Matthews et al., 2009; Matthews & Wynes, 2022), a number of uncertainties remain, including in the exact amount of warming for a given level of cumulative emissions (e.g., Matthews & Wynes, 2022), the exact carbon budget for a given level of warming (e.g., Lamboll et al., 2023;

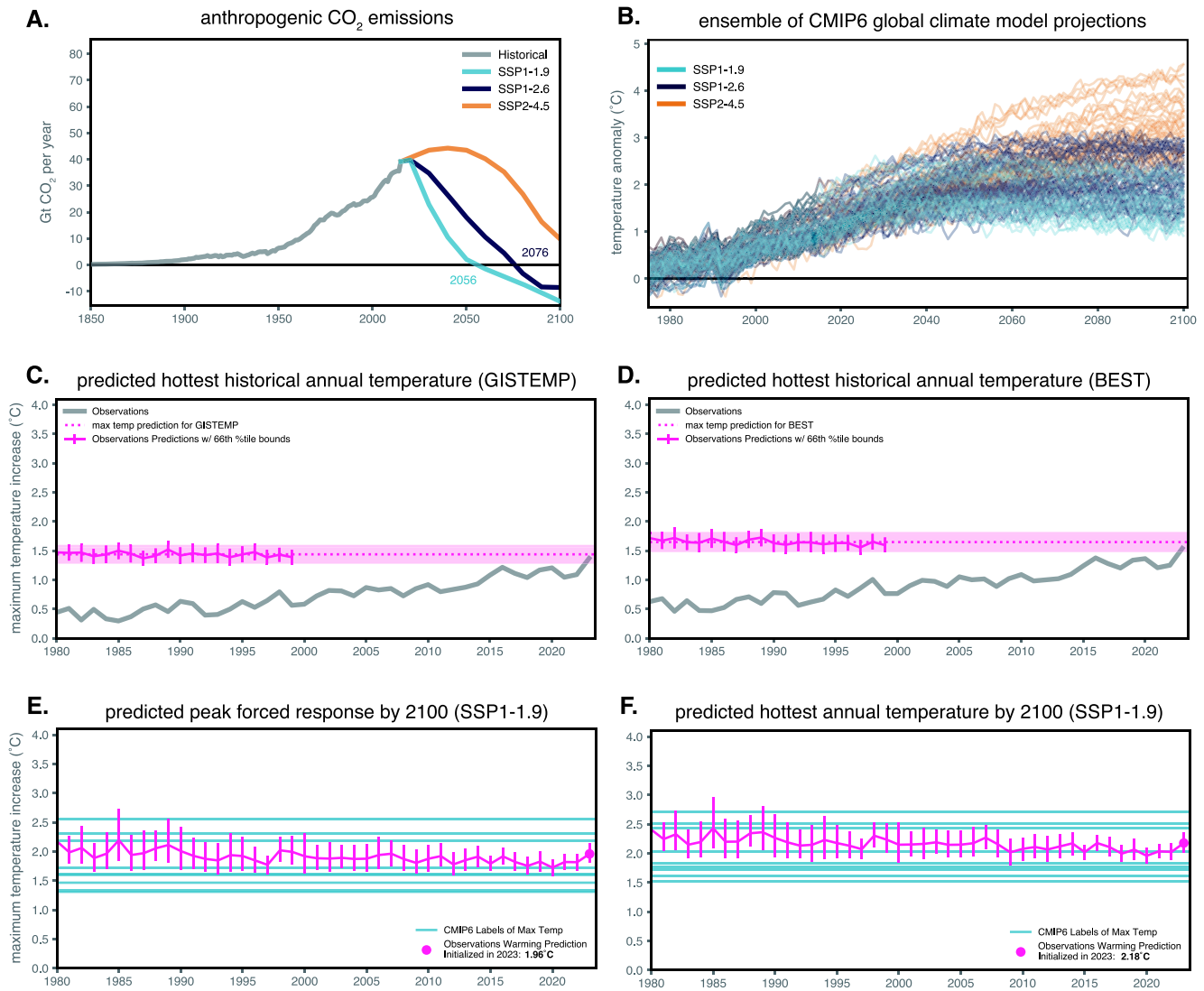


Figure 1. Timeseries of predicted peak warming. (a) Anthropogenic CO₂ emissions over the historical period and decarbonization scenarios. (b) Global temperature anomalies from the pre-industrial baseline (1850–1899) for 5-member global climate model ensembles. (c) CNN predictions of the maximum annual temperature reached by 2023, initialized in each year using the observed temperature map from NASA (GISTEMP). Vertical bars denote the predictions initialized in that year with their 66th percentile uncertainty bounds (“likely” range in the IPCC uncertainty guidance). Dotted line denotes the average prediction made using inputs from 1980 to 1999, and shading denotes its uncertainty. (d) As in (c), but for Berkeley Earth (BEST) observations. (e) Initialized predictions of the future peak forced temperature response from CNNs trained across the SSP1-1.9, SSP1-2.6 and SSP2-4.5 scenarios and initialized with Berkeley Earth observations (and SSP1-1.9 cumulative emissions). Blue lines denote the peak forced response across the climate model testing set. (f) As in (e) but for the maximum annual temperature by 2100.

Rogelj et al., 2019), and the peak and equilibrium warming in a given decarbonization scenario (e.g., Lee et al., 2021, Figure 1).

The recognition that emissions reductions may not avoid a particular warming threshold has led to the concept of “overshoot”, in which the global temperature initially exceeds the threshold, reaches peak warming, and then falls back below the threshold (e.g., IPCC, 2018). Like the broader literature on climate stabilization, overshoot research has primarily focused on mean global temperature (Betts et al., 2023; IPCC, 2018). However, “internal” climate variability creates substantial uncertainty in the climate trajectory along a decarbonization pathway (e.g., Diffenbaugh et al., 2023). While this uncertainty is most pronounced at local and regional scales, global temperature variability will cause specific thresholds to be exceeded by individual years before being exceeded by the forced response (e.g., Betts et al., 2023; Lee et al., 2021; Smith et al., 2018). Further, the breadth and intensity of impacts in extremely hot years such as 2023—a record-setting excursion that approached 1.5°C above the pre-

industrial (R. Rohde, 2024; Schmidt, 2024; Figure 1)—have heightened the need to accurately predict not just the forced global temperature response but also the hottest individual years that could occur even if the most ambitious decarbonization goals are achieved.

With that motivation, we use convolutional neural networks (CNNs) to quantify the probability of peak global warming under different decarbonization scenarios. We build on recent work using neural networks to predict the time until a given forced temperature response is reached (Diffenbaugh & Barnes, 2023). A key advantage of this framework is the capacity to constrain climate model uncertainty using the observed climate as the basis for an out-of-sample prediction. In the current study, we train CNNs to predict the distribution of peak forced temperature response and maximum annual temperature from (a) maps of annual temperature simulated by an ensemble of global climate models (GCMs) and (b) remaining cumulative CO₂ emissions. To quantify uncertainty, the CNN architecture is constructed to predict the parameters of a sinh-arcsinh conditional distribution (SHASH (Barnes et al., 2023; C. Jones & Pewsey, 2019; M. C. Jones & Pewsey, 2009)), and we train multiple CNNs using different initial random seeds to produce derived distributions that capture multiple sources of uncertainty. Once the CNNs are trained, we use observed maps of recent annual temperature anomalies to predict the distribution of peak warming under different levels of future cumulative emissions.

2. Materials and Methods

2.1. Data Sets

To train the CNNs across a range of decarbonization pathways, we pool results from multiple realizations of multiple GCMs in the SSP1-1.9, SSP1-2.6 and SSP2-4.5 decarbonization scenarios archived in Phase 6 of the Coupled Model Intercomparison Project (CMIP6 (Eyring et al., 2016); Figure 1, Table S1 in Supporting Information S1). As in Diffenbaugh and Barnes (2023), we select those GCMs that have archived ≥ 5 realizations in order to capture the influence of internal climate variability, but only include 5 realizations from each in order to weight those GCMs evenly. One limitation is that only a few GCMs have archived at least 5 realizations in the CMIP6 scenarios that reach net-zero during the 21st century (SSP1-1.9 and SSP1-1.6; Table S1 in Supporting Information S1). Given the importance of including a broad range of climate sensitivities, we expand the training data set by also including one realization of each available GCM, as in the IPCC's "one model, one vote" approach.

Within this "grand ensemble", we define the peak forced response for each ensemble member separately. We fit each realization's global mean temperature timeseries from 2000 to 2100 with a cubic function, and the maximum of this cubic fit defines the peak forced response (Figure S2A in Supporting Information S1). The maximum annual temperature is similarly defined separately for each realization, except that no cubic fit is applied (Figure S2B in Supporting Information S1).

We use the Berkeley Earth gridded temperature observations as our primary observational data set. Although the global temperature synthesis is very consistent between observational data sets (Gulev et al., 2021), there can be differences in the exact spatial patterns in a given year (e.g., NCEI, 2024; R. Rohde, 2024). However, the neural network predictions of the evolution of the global temperature response are consistent between gridded data sets (Diffenbaugh & Barnes, 2023), and we find very close agreement in the prediction of the historical maximum annual temperature using gridded data sets from Berkeley Earth (R. A. Rohde & Hausfather, 2020) and NASA (GISTEMP, 2021) (Figure 1).

For all observed and simulated data sets, we interpolate the annual temperature data to a common $2.5^\circ \times 2.5^\circ$ grid (IPCC, 2013). Global mean temperature anomalies (the predictand) are computed as deviations from the 1850–1899 mean of each realization. Surface temperature anomalies (the predictors) for both the GCMs and observations are computed as the difference from each realization's 1975–2004 mean.

For calculating cumulative emissions, we first concatenate the historical emissions data from (Friedlingstein et al., 2022) with the SSP emissions data for the remainder of the 21st century (Gidden et al., 2019; IIASA, 2018; Riahi et al., 2017), and then follow the SSP extensions of (Meinshausen et al., 2020) by linearly interpolating at an annual resolution out to 2250 (Figure S7A in Supporting Information S1). For a given year, the cumulative emissions remaining are then computed as the sum over all positive emissions from 1850 to 2250 minus the cumulative sum-to-date. Since SSP1-1.9 and SSP1-2.6 both reach net-zero over the 21st century, this definition leads to negative emissions remaining prior to 2100 (Figure S7b in Supporting Information S1).

2.2. Convolutional Neural Networks (CNNs)

2.2.1. Architecture and Training of the CNN

We train a CNN to take single maps of annual-mean near-surface temperature (“tas”) and the additional cumulative CO₂ emissions remaining, and predict the peak warming (as an anomaly from 1850 to 1899) that will occur by 2100 (Figure S1 in Supporting Information S1). Following previous work, we train the CNN to predict the four parameters of a sinh-arcsinh (SHASH) distribution (Barnes et al., 2023; C. Jones & Pewsey, 2019; M. C. Jones & Pewsey, 2009): mu; sigma; gamma; and tau. Thus, for each input into the network, the network predicts a full conditional distribution of the warming remaining through 2100 (Figure S3 in Supporting Information S1). From this, we compute the predicted peak warming as the sum of the current year’s temperature anomaly plus the predicted remaining warming. Although the SHASH takes four parameters, we have found that training the CNN to predict only the first three (and freezing tau = 1.0) produces similar results to training all four (see Figure S5 in Supporting Information S1).

The CNN has an initial circular padding layer that accounts for the discontinuity in longitude at the Prime Meridian. From there, the input map is fed through three convolutional layers with 32 kernels each of size 5 × 5, 3 × 3 and 3 × 3, respectively. Each convolutional layer is followed by a max pooling layer with kernel size 2 × 2 and stride of 2 to reduce the dimensionality. The final max pooling layer is followed by a flatten layer to which the input cumulative CO₂ emissions remaining scalar is appended. We separate the predictions of the SHASH parameters, thereby allowing the network to learn different strategies if necessary. Each parameter pathway consists of three dense layers of 10 units each. The input cumulative CO₂ emissions remaining are once again concatenated to the output of the dense block and the resulting tensor is fed into a final dense block of 5 units. The output is processed through a rescaling layer prior to predicting each parameter. We use the rectified linear unit (ReLU) for all activations except for the final output, which uses a hyperbolic tangent (Tanh) activation. The full CNN setup depicted in Figure S1 in Supporting Information S1 contains a total of 273,992 trainable parameters. However, we freeze the fourth parameter (tau) as its inclusion did not improve our results (compare the “default” to “hp9” and “hp10” in Figure S5 in Supporting Information S1). Thus, the final CNN architecture contains 210,326 trainable parameters.

The training set is created by randomly selecting two realizations from each GCM that has at least 5 realizations (“multi-member GCMs”), and randomly selecting 80% of the single-member realizations (which include the first realization from the multi-member GCMs). The validation set, which is used for early stopping, is created by randomly selecting one of the remaining members of the multi-member GCMs and 10% of the remaining single-member realizations. The testing set is then created by combining all of the remaining realizations, which amounts to one from the multi-member GCMs and 10% of the single-member realizations. The CNNs are trained to minimize the negative log-likelihood loss using the Adam optimizer with a batch size of 32, learning rate of 1.0e−05, weight decay of zero, and eps parameter of 1.0e−07. We employ early stopping when the validation loss no longer decreases after a patience of 20 epochs to help avoid overfitting.

We conduct a number of tests to explore the sensitivity of our results to the CNN architecture choices by training 5 different CNNs for each network hyperparameter configuration: “hp1-10” (Figure S5 in Supporting Information S1). The five CNNs for each configuration differ only by their random seed, which sets both the training/validation/testing split as well as the random initialization of the weights. The architecture used here (“default”) results in low performance metrics such as mean absolute error and loss, although it is reassuring that other hyperparameter choices lead to similar predictions for the observations (Figure S5 in Supporting Information S1).

2.2.2. Uncertainty Quantification

The network predicts a SHASH distribution of the remaining warming for each input, providing a measure of uncertainty conditioned on the input. However, training different CNNs that vary only in the random seed used to split the data and initialize the network can result in a range of observational predictions (Figures S4, S5 in Supporting Information S1). Thus, there is an additional source of uncertainty not directly captured by each predicted SHASH (Barnes et al., 2023; Haynes et al., 2023). To account for this, we train 15+ CNNs with differing seeds for each peak warming definition (i.e., peak forced response and maximum annual temperature) and select the first 15 CNNs whose testing loss is less than −0.1 (Figure S5 in Supporting Information S1 provides metrics for each of the 15+ individual SHASH distribution predictions.). We then create a derived distribution by

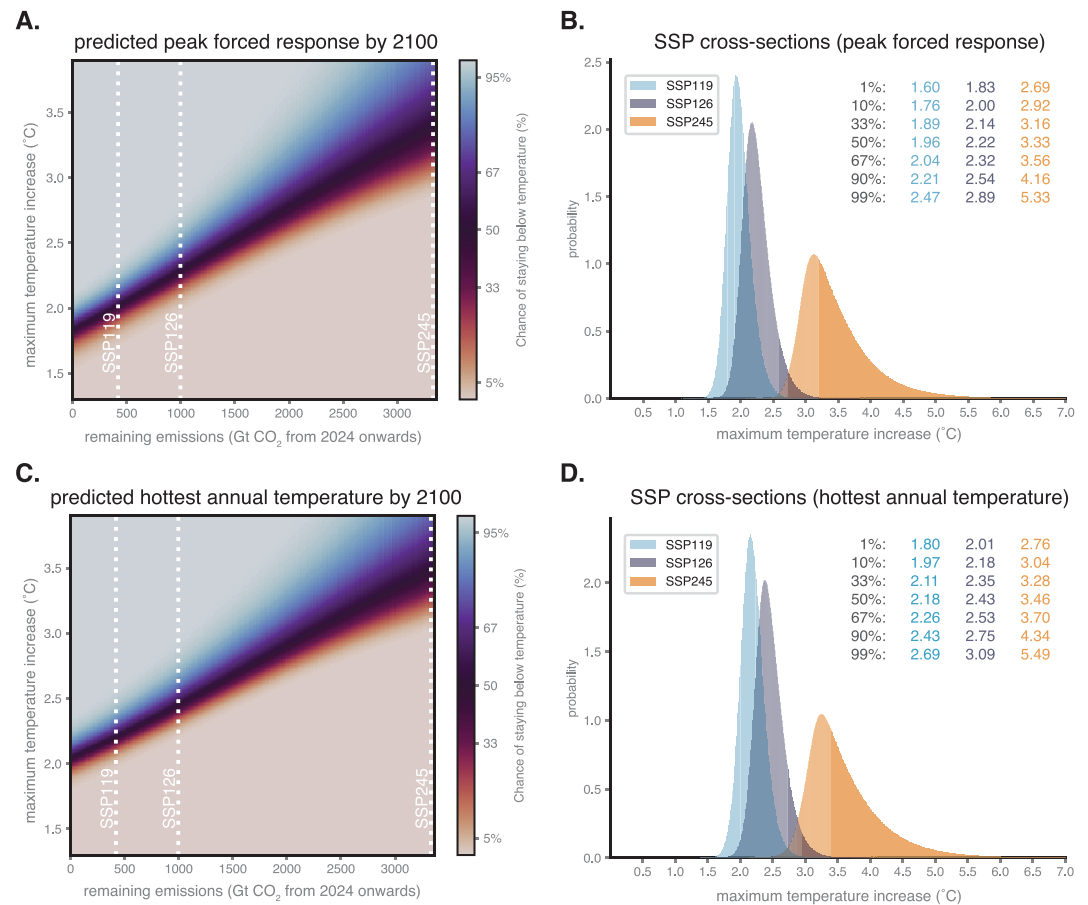


Figure 2. Probability of peak warming in response to cumulative emissions. (a) Probability that the peak forced temperature response (compared to the 1850–1899 pre-industrial baseline) remains below a range of thresholds as a function of remaining cumulative emissions, predicted by CNNs initialized using observations in 2023. Vertical dotted lines denote the remaining emissions under three decarbonization scenarios. (b) The CNN-predicted derived distributions for the peak forced response by 2100, corresponding to the cross-sections shown as vertical dotted lines in (a). (c) As in (a) but for the maximum annual temperature by 2100. (d) As in (b), but for the maximum annual temperature by 2100 (i.e., corresponding to the cross-sections shown as vertical dotted lines in (c)).

combining the predictions across all 15 CNNs. Specifically, for each prediction on the observations (i.e., each set of inputs) we obtain 15 SHASH distributions of the remaining warming. We then take 50,000 random draws from each predicted SHASH distribution resulting in $15 \times 50,000 = 750,000$ random values. These 750,000 values thus comprise our predicted derived distribution from which we can plot the resulting probability density function and calculate probabilities (Figure 2). All results discussed in the text are quantified from this derived distribution. Individual SHASH predictions and metrics are presented in Figures S3–S5 in Supporting Information S1.

Our decision to employ a derived distribution for quantifying uncertainty is supported by the fact that the testing uncertainty (and error) introduced by the choice in training/validation/testing split is much larger than the uncertainty due to the random initialization of the network weights (Figure S6 in Supporting Information S1). Since we employ a “one-model, one-vote” approach for the GCM ensemble, every training/validation/testing split is equally valid presuming that the network does a reasonable job training (i.e., meets our criteria that the loss be less than -0.1). Because each of the 15 CNNs are equally valid, we choose to combine all 15 predictions into a single derived distribution to capture the uncertainty from the data split rather than choose one CNN, which would fail to capture this additional source of climate model uncertainty.

3. Results

We first test the accuracy of our prediction framework during the historical period. We train multiple CNNs on the CMIP6 Historical simulations (extending to 2023 using the SSP1-1.9 simulations). Then, for each historical year, we use the observed maps of annual temperature anomalies and remaining cumulative CO₂ emissions (through 2023) to predict the maximum annual temperature that will be reached by 2023. The observations-based prediction is quite stable through time, and is very similar when using inputs from NASA (“GISTEMP”; Figure 1c) or Berkeley Earth (“BEST”; Figure 1d). Further, the actual hottest historical year falls very close to the mean of CNN predictions made using inputs from 1980 to 1999, and within the “likely” range (66% in the IPCC uncertainty guidance (Mastrandrea et al., 2011)) for both observed temperature data sets.

We next use the CNNs trained across the SSP decarbonization scenarios to predict the distribution of peak warming given 2023 temperature observations and varying levels of cumulative emissions (Figure 2). We find that the timeseries of peak warming predicted by the CNNs is robust to the initialization year, particularly for the past two decades (Figures 1e and 1f). Further, the uncertainties in the observations-based CNN predictions are substantially smaller than the uncertainty in the GCM simulations on which the CNNs are trained, particularly for the rapid decarbonization scenarios (Figures 1e, 1f, S8 in Supporting Information S1).

Although the median prediction is generally proportional to the cumulative emissions, the distributions become more positively skewed as cumulative emissions increase, with the upper tail increasing at a faster rate than the lower tail (Figures 2a and 2c). For SSP1-1.9, which reaches net-zero CO₂ emissions in the mid-2050s followed by deep negative emissions, the CNN-predicted peak forced response by 2100 is “virtually certain” (Mastrandrea et al., 2011) to exceed 1.60°C (median prediction: 1.96°C), and the maximum annual temperature is “virtually certain” to exceed 1.80°C (median prediction: 2.18°C). For SSP1-2.6, which reaches net-zero CO₂ emissions in the mid-2070s, the peaked forced response is “virtually certain” to exceed 1.83°C (median prediction: 2.22°C), and the maximum annual temperature is “virtually certain” to exceed 2.01°C (median prediction: 2.43°C). For SSP2-4.5, which does not reach net-zero in the 21st century, the peak forced response is “virtually certain” to exceed 2.69°C (median prediction: 3.33°C), and the maximum annual temperature is “virtually certain” to exceed 2.76°C (median prediction: 3.46°C).

The magnitude of predicted peak warming suggests that, even if rapid decarbonization is achieved, regional conditions are likely to be considerably more severe than what people and ecosystems have experienced to date. For example, for an annual global temperature anomaly of ~2°C, regional conditions are highly likely to be hotter than in 2023, including a strong possibility of regional temperature anomalies that are at least double the 2023 regional anomaly (Figure 3a). Likewise, for a single climate realization whose maximum annual global temperature anomaly reaches ~2°C, the maximum annual daily precipitation in the hottest years exceeds the baseline maximum over most of the globe, including widespread increases of at least 25% over continental areas and even larger increases over much of the tropics (Figure 3c). However, such rapid decarbonization is far from guaranteed (UNEP, 2023), and there is very high likelihood that the maximum annual global temperature anomaly reaches 2.5°C for additional cumulative emissions of 2,000 GtCO₂ and 3.0°C for additional cumulative emissions of 3,500 GtCO₂ (Figure 2c). For an annual global temperature anomaly of ~3°C, regional conditions are virtually certain to be hotter than in 2023 over most illustrative regions, including a strong possibility of regional temperature anomalies that are at least triple the 2023 regional anomaly (Figure 3a).

4. Discussion

Our results suggest virtual certainty (>99% in the IPCC uncertainty guidance (Mastrandrea et al., 2011)) that reaching net-zero CO₂ emissions by mid-century will not hold the global forced temperature response below 1.5°C. While there are a variety of definitions for global temperature thresholds (e.g., Betts et al., 2023), the forced response is among the most conservative. Hence, our results raise the likelihood of a broad suite of impacts that are expected to occur beyond the 1.5°C threshold, including on terrestrial and marine ecosystems, human health, livelihoods, economic growth, and food and water security (IPCC, 2018).

In addition, the median CNN-predicted peak forced response by 2100 for SSP1-1.9 emissions is 1.96°C (Figure 1), and our results suggest that other ambitious decarbonization pathways that hold future cumulative emissions to ~500 GtCO₂ are also only ~50% likely to avoid a peak forced response of 2°C (Figure 2a). Absent adaptation, this level of warming poses “high” to “very high” risks to natural and human systems across the

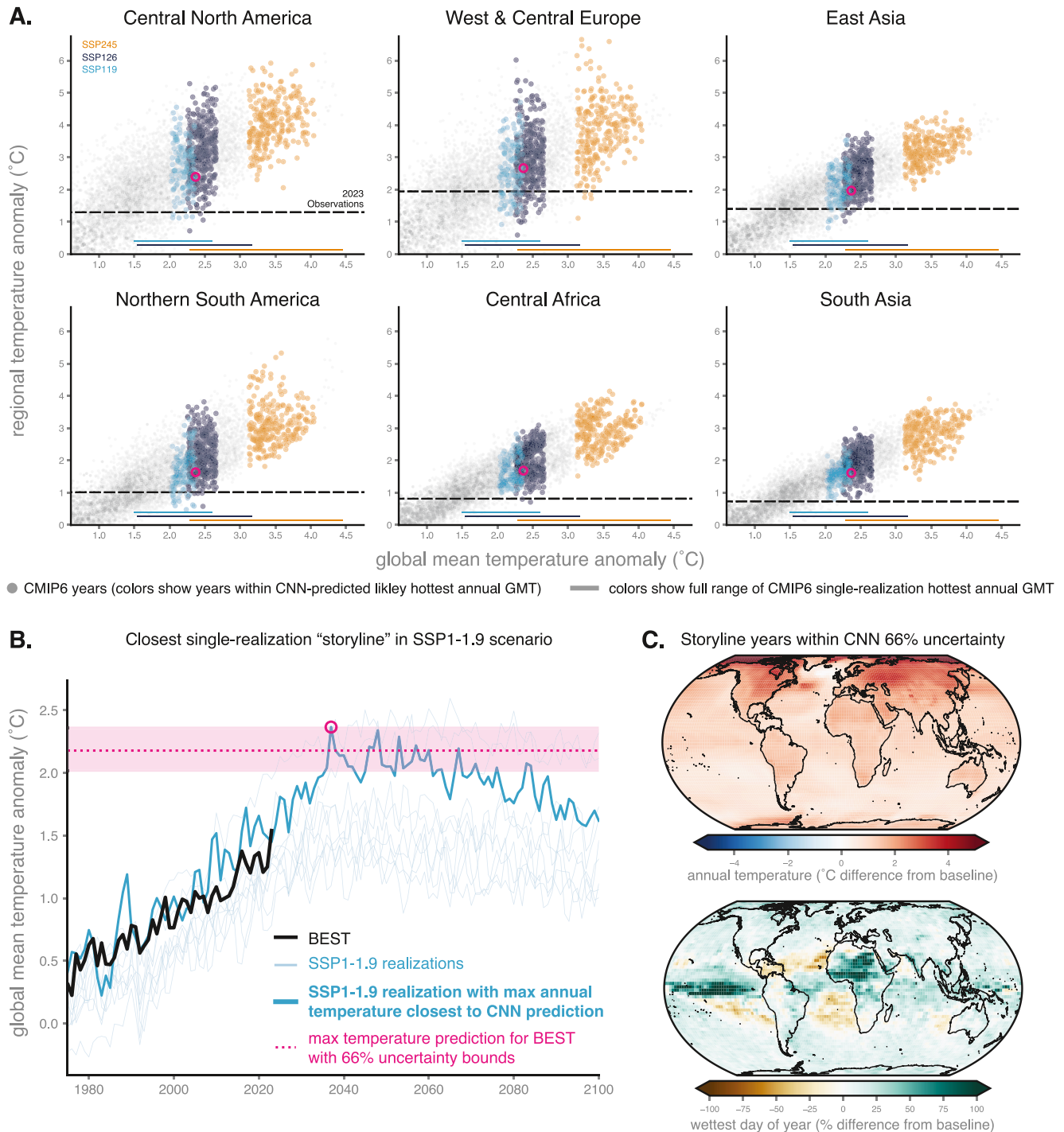


Figure 3. Regional climate risks associated with maximum annual global temperature under rapid decarbonization. (a) Global and regional annual temperature anomalies (from the 1850–1899 baseline) across the CMIP6 single-realization climate model ensembles. Colors show years where the global anomaly falls within the IPCC “likely” range of the CNN predictions (66th percentile uncertainty bounds). Horizontal lines at the base of each panel denote the full range across the respective CMIP6 single-realization ensembles. Dashed black bars denote the observed regional temperature anomaly in 2023; pink circles denote the point associated with the single-realization storyline identified in (b). (b) Global-mean temperature anomalies for individual CMIP6 SSP1-1.9 realizations (light blue). Black line denotes observations; bold blue line denotes the CMIP6 realization whose hottest annual temperature is closest to the CNN-predicted value for SSP1-1.9 emissions. Dotted pink line shows the median CNN prediction, and pink shading denotes the IPCC “likely” range. (c) Composite maps of annual temperature and wettest day of the year for all 27 years where the realization’s global-mean temperature falls within the IPCC “likely” range of CNN predictions (shown in (b)). Composite values are calculated relative to the pre-industrial baseline (1850–1999).

comprehensive areas assessed by the IPCC (IPCC, 2022). Further, our results suggest that the peak forced response is very likely to exceed 2°C if additional cumulative emissions exceed 1,000 GtCO₂, and is likely to exceed 3°C if additional cumulative emissions exceed 3,500 GtCO₂ (Figure 2a). Absent adaptation, 3°C of mean global warming poses “very high” risks (IPCC, 2022).

Given the importance of extreme events for climate change impacts (e.g., IPCC, 2012), the magnitude of individual hot years is also critical for understanding risks under different trajectories of decarbonization. Our results suggest that the hottest annual global temperature anomaly is very likely to be at least 1.97°C—and likely to be at least 2.11°C—even in the most ambitious decarbonization scenario (Figure 2d). These values for the hottest annual global temperature anomaly are ~0.2°C hotter than the equivalent-likelihood peak forced response (Figure 2b), reflecting the important influence of internal climate variability. This difference of ~0.2°C is generally consistent across the probability distribution for both the SSP1-1.9 and SSP1-2.6 cumulative emissions (Figures 2b and 2d), and is also consistent with the difference between the mean global warming (~1.3°C) and hottest historical year (~1.5°C) in observations (e.g., R. Rohde, 2024).

The CNNs' predicted probabilities include higher likelihood of mean warming than reported by the IPCC (Lee et al., 2021) and some other assessments (e.g., Lamboll et al., 2023; Matthews & Wynes, 2022; Rogelj et al., 2019). For example, for SSP1-1.9, the lowest 1% of the CNNs' predictions for the peak forced response (1.60°C; Figure 2b) is similar to the median IPCC assessment based on multiple lines of evidence (1.6°C for the mid-21st century of SSP1-1.9 (Lee et al., 2021)), and the median CNN prediction (1.96°C) is similar to the 95th percentile IPCC assessment (2.0°C for the mid-21st century of SSP1-1.9). Likewise, whereas the median CNN prediction of the peak forced response is ~2°C for additional emissions of ~500 GtCO₂ from 2024 onwards (Figures 2a), Matthews and Wynes (2022) reported median warming between 1.5°C and 1.6°C for a remaining budget of ~500 GtCO₂ from 2022 onwards, and Lamboll et al. (2023) reported median likelihood of 1.5°C for ~250 GtCO₂ from 2023 onwards, and 2°C for ~1,200 GtCO₂.

The primary difference between our framework and previous approaches is that our framework uses the CNN to refine the climate model uncertainty based on the observed map of recent temperature anomalies (Figures 1e and 1f). Our framework is thus another approach to constraining the climate model ensemble with observations, using the power of the CNN to learn which indicators are most reliable for accurately predicting the peak warming. However, it is possible for inaccuracies to be introduced into our prediction framework. In particular, the fact that the CNNs are trained on GCM simulations means that the networks can only learn relationships that are represented in the climate models. While we are careful to train on combinations of multiple GCMs and multiple realizations from individual GCMs, biases that are consistent across GCMs may be learned by the CNNs. This could lead to errors in predictions based on the observations. Thus, one test to gauge trust in our approach is to apply our framework to the historical period, reserving the observed climate data as input for out-of-sample predictions. The fact that our framework accurately predicts both the timing of historical global warming (Diffenbaugh & Barnes, 2023) and the magnitude of the hottest historical year (Figure 1) provides confidence in the accuracy of the predicted probabilities of future peak warming (Figure 2).

Finally, we follow the IPCC and numerous other studies in focusing on cumulative CO₂ emissions (e.g., IPCC, 2021; Lamboll et al., 2023; Matthews et al., 2009; Matthews & Wynes, 2022; Rogelj et al., 2019). However, non-CO₂ emissions also impact the climate system (e.g., Andreae et al., 2005; Fu et al., 2020; Lee et al., 2021), with implications for warming levels associated with decarbonization targets (Li et al., 2022; Matthews et al., 2020). Although the uncertainty in the relationship between global temperature change and cumulative CO₂ emissions follows a similar proportionality between the SSP scenarios (IPCC, 2021), differences in non-CO₂ emissions between the SSPs (e.g., Meinshausen et al., 2020) likely contribute to the uncertainty in the magnitude of the proportional relationship. The differences in non-CO₂ emissions is much larger for the higher-forcing SSPs (e.g., SSP3-7.0 and SSP5-8.5) than for the decarbonization SSPs that we use (SSP1-1.9, SSP1-2.6 and SSP2-4.5) (Meinshausen et al., 2020). However, for the smaller carbon budgets associated with more rapid reductions in CO₂ emissions, the trajectory of non-CO₂ emissions will be an important influence on the actual global and regional temperatures that occur (e.g., Lamboll et al., 2023; Matthews et al., 2020).

5. Conclusions

We train CNNs to predict the distribution of peak warming from the map of observed recent surface temperature anomalies and future cumulative CO₂ emissions. We find that even in the most ambitious decarbonization

scenarios it is “virtually certain” that the world will exceed 1.5°C of mean warming, and that there are approximately even odds of reaching 2°C. Further, we find high likelihood of individual years that are at least 0.5°C hotter than the record-setting global anomaly in 2023, even if net-zero CO₂ emissions are reached mid-century. This result suggests that even the most ambitious emissions reductions will still yield further intensification of severe heat, heavy precipitation, intensified drought, and other extreme climate conditions. Given the robust empirical evidence for non-linear sensitivity in many natural and human systems (e.g., Abatzoglou & Williams, 2016; Burke et al., 2015; Davenport et al., 2021; Diffenbaugh et al., 2021; Hsiang et al., 2017), substantial adaptation is likely to be required if dangerous impacts are to be avoided.

Data Availability Statement

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling and the participating climate modeling groups for producing and making available their model output to the CMIP6 archive, which is publicly available on the Earth System Grid Federation (ESGF) at the following link: <https://aims2.llnl.gov/search>. We acknowledge Berkeley Earth for making their gridded temperature observations publicly available at the following link: <https://berkeleyearth.org/data/>. We acknowledge NASA GISS for making their gridded temperature observations publicly available at the following link: <https://data.giss.nasa.gov/gis-temp/>. Code is available on GitHub and archived on Zenodo (Barnes, 2024).

Acknowledgments

We thank two anonymous reviewers. NSD acknowledges support from Stanford University. EAB's contributions were funded, in part, by the Regional and Global Model Analysis program area of the U.S. DOE Office of Biological and Environmental Research (BER) as part of the Program for Climate Model Diagnosis and Intercomparison project.

References

- Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences*, 113(42), 11770–11775. <https://doi.org/10.1073/pnas.1607171113>
- Andreae, M. O., Jones, C. D., & Cox, P. M. (2005). Strong present-day aerosol cooling implies a hot future. *Nature*, 435(7046), 1187–1190. <https://doi.org/10.1038/nature03671>
- Barnes, E. A. (2024). eabarnes1010/shash_peak_warming_public. *Zenodo*. <https://doi.org/10.5281/zenodo.14045959>
- Barnes, E. A., Barnes, R. J., & DeMaria, M. (2023). Sinh-arcsinh-normal distributions to add uncertainty to neural network regression tasks: Applications to tropical cyclone intensity forecasts. *Environmental Data Science*, 2, e15. <https://doi.org/10.1017/eds.2023.7>
- Betts, R. A., Belcher, S. E., Hermanson, L., Klein Tank, A., Lowe, J. A., Jones, C. D., et al. (2023). Approaching 1.5°C: How will we know we've reached this crucial warming mark? *Nature*, 624(7990), 33–35. <https://doi.org/10.1038/d41586-023-03775-z>
- Burke, M., Hsiang, S. M., & Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature*, 527(7577), 235–239. <https://doi.org/10.1038/nature15725>
- Davenport, F. V., Burke, M., & Diffenbaugh, N. S. (2021). Contribution of historical precipitation change to US flood damages. *Proceedings of the National Academy of Sciences of the United States of America*, 118(4). <https://doi.org/10.1073/pnas.2017524118>
- Diffenbaugh, N. S., & Barnes, E. A. (2023). Data-driven predictions of the time remaining until critical global warming thresholds are reached. *Proceedings of the National Academy of Sciences*, 120(6), e2207183120. <https://doi.org/10.1073/pnas.2207183120>
- Diffenbaugh, N. S., Barnes, E. A., & Keys, P. W. (2023). Probability of continued local-scale warming and extreme events during and after decarbonization. *Environmental Research: Climate*, 2(2), 21003. <https://doi.org/10.1088/2752-5295/acf2f2>
- Diffenbaugh, N. S., Davenport, F. V., & Burke, M. (2021). Historical warming has increased U.S. crop insurance losses. *Environmental Research Letters*, 16(8), 84025. <https://doi.org/10.1088/1748-9326/ac1223>
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model Intercomparison project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., et al. (2023). Global carbon budget 2023. *Earth System Science Data*, 15(12), 5301–5369. <https://doi.org/10.5194/essd-15-5301-2023>
- Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., et al. (2022). Global carbon budget 2022. *Earth System Science Data*, 14(11), 4811–4900. <https://doi.org/10.5194/essd-14-4811-2022>
- Fu, B., Gasser, T., Li, B., Tao, S., Ciais, P., Piao, S., et al. (2020). Short-lived climate forcings have long-term climate impacts via the carbon-climate feedback. *Nature Climate Change*, 10(9), 851–855. <https://doi.org/10.1038/s41558-020-0841-x>
- Gidden, M. J., Riahi, K., Smith, S. J., Fujimori, S., Luderer, G., Kriegler, E., et al. (2019). Global emissions pathways under different socioeconomic scenarios for use in CMIP6: A dataset of harmonized emissions trajectories through the end of the century. *Geoscientific Model Development*, 12(4), 1443–1475. <https://doi.org/10.5194/gmd-12-1443-2019>
- GISTEMP. (2021). GISS surface temperature analysis (GISTEMP), version 4. Retrieved from <https://data.giss.nasa.gov/gistemp/>
- Gulev, S. K., Thorne, P. W., Ahn, J., Dentener, F. J., Domingues, C. M., Gerland, S., et al. (2021). Changing State of the Climate System. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 287–422). Cambridge University Press. <https://doi.org/10.1017/9781009157896.004>
- Haynes, K., Lagerquist, R., McGraw, M., Musgrave, K., & Ebert-Uphoff, I. (2023). Creating and evaluating uncertainty estimates with neural networks for environmental-science applications. *Artificial Intelligence for the Earth Systems*, 2(2), 220061. <https://doi.org/10.1175/AIES-D-22-0061.1>
- Hsiang, S., Kopp, R., Jina, A., Rising, J., Delgado, M., Mohan, S., et al. (2017). Estimating economic damage from climate change in the United States. *Science*, 356(6345), 1362–1369. <https://doi.org/10.1126/science.aal4369>
- IIASA. (2018). SSP database (shared socioeconomic pathways) - version 2.0. Retrieved from <https://tntcat.iiasa.ac.at/SspDb/dsd>
- IPCC. (2012). Managing the risks of extreme events and disasters to advance climate change adaptation. In C. B. Field, V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, et al. (Eds.), *A special report of Working Groups I and II of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

- IPCC. (2013). Annex I: Atlas of global and regional climate projections. In *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1311–1394). Cambridge University Press.
- IPCC. (2018). Summary for policymakers. In M. Allen, M. Babiker, Y. Chen, & H. C. de Coninck (Eds.), *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*. Cambridge University Press.
- IPCC. (2021). Summary for policymakers. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, et al. (Eds.), *Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- IPCC. (2022). Summary for policy makers. In H. O. Pörtner, D. C. Roberts, H. Adams, C. Adler, P. Aldunce, et al. (Eds.), *Climate change 2022: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC.
- Jones, C., & Pewsey, A. (2019). The sinh-arcsinh normal distribution. *Significance*, 16(2), 6–7. <https://doi.org/10.1111/j.1740-9713.2019.01245.x>
- Jones, M. C., & Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika*, 96(4), 761–780. <https://doi.org/10.1093/biomet/asp053>
- Lamboll, R. D., Nicholls, Z. R. J., Smith, C. J., Kikstra, J. S., Byers, E., & Rogelj, J. (2023). Assessing the size and uncertainty of remaining carbon budgets. *Nature Climate Change*, 13(12), 1360–1367. <https://doi.org/10.1038/s41558-023-01848-5>
- Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J. P., et al. (2021). Future global climate: Scenario-based projections and near-term information. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 553–672). Cambridge University Press. <https://doi.org/10.1017/9781009157896.006>
- Li, L., Zhang, Y., Zhou, T., Wang, K., Wang, C., Wang, T., et al. (2022). Mitigation of China's carbon neutrality to global warming. *Nature Communications*, 13(1), 5315. <https://doi.org/10.1038/s41467-022-33047-9>
- Mastrandrea, M. D., Mach, K. J., Plattner, G.-K., Edenhofer, O., Stocker, T. F., Field, C. B., et al. (2011). The IPCC AR5 guidance note on consistent treatment of uncertainties: A common approach across the working groups. *Climatic Change*, 108(4), 675–691. <https://doi.org/10.1007/s10584-011-0178-6>
- Matthews, H. D., & Caldeira, K. (2008). Stabilizing climate requires near-zero emissions. *Geophysical Research Letters*, 35(4), L04705. <https://doi.org/10.1029/2007GL032388>
- Matthews, H. D., Gillett, N. P., Stott, P. A., & Zickfeld, K. (2009). The proportionality of global warming to cumulative carbon emissions. *Nature*, 459(7248), 829. <https://doi.org/10.1038/nature08047>
- Matthews, H. D., Tokarska, K. B., Nicholls, Z. R. J., Rogelj, J., Canadell, J. G., Friedlingstein, P., et al. (2020). Opportunities and challenges in using remaining carbon budgets to guide climate policy. *Nature Geoscience*, 13(12), 769–779. <https://doi.org/10.1038/s41561-020-00663-3>
- Matthews, H. D., & Wynes, S. (2022). Current global efforts are insufficient to limit warming to 1.5°C. *Science*, 376(6600), 1404–1409. <https://doi.org/10.1126/science.abc3378>
- Meinshausen, M., Nicholls, Z. R. J., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., et al. (2020). The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geoscientific Model Development*, 13(8), 3571–3605. <https://doi.org/10.5194/gmd-13-3571-2020>
- NCEI. (2024). Monthly global climate report for annual 2023. Retrieved from <https://www.ncei.noaa.gov/access/monitoring/monthly-report/global/202313>
- Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., et al. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, 42, 153–168. <https://doi.org/10.1016/j.gloenvcha.2016.05.009>
- Rogelj, J., Forster, P. M., Kriegler, E., Smith, C. J., & Séférian, R. (2019). Estimating and tracking the remaining carbon budget for stringent climate targets. *Nature*, 571(7765), 335–342. <https://doi.org/10.1038/s41586-019-1368-z>
- Rogelj, J., Schleussner, C.-F., & Hare, W. (2017). Getting it right matters: Temperature goal interpretations in geoscience research. *Geophysical Research Letters*, 44(20), 10610–10626. <https://doi.org/10.1002/2017GL075612>
- Rohde, R. (2024). Global temperature report for 2023. Retrieved from <https://berkeleyearth.org/global-temperature-report-for-2023/>
- Rohde, R. A., & Hausfather, Z. (2020). The Berkeley Earth land/ocean temperature record. *Earth System Science Data Discussions*. <https://doi.org/10.5194/essd-2019-259>
- Schmidt, G. (2024). Climate models can't explain 2023's huge heat anomaly — We could be in uncharted territory. *Nature*, 627(8004), 467. <https://doi.org/10.1038/d41586-024-00816-z>
- Smith, D. M., Scaife, A. A., Hawkins, E., Bilbao, R., Boer, G. J., Caian, M., et al. (2018). Predicted chance that global warming will temporarily exceed 1.5°C. *Geophysical Research Letters*, 45(21), 11811–11819. <https://doi.org/10.1029/2018GL079362>
- UNEP. (2023). Emissions Gap Report 2023: Broken Record – temperatures hit new highs, yet world fails to cut emissions (again). Nairobi. Retrieved from <https://www.unep.org/resources/emissions-gap-report-2023>
- UNFCCC. (2015). Adoption of the Paris agreement. Paris, France. Retrieved from <https://unfccc.int/resource/docs/2015/cop21/eng/l09.pdf>