Posted on 2 Mar 2026 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/essoar.177245874.45134081/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

Recent Temperature and Energy Imbalance Trends Point to Higher Estimates of Future Warming

Gergana Gyuleva[1], Erich Markus Fischer[2], Reto Knutti[2], and Sebastian Sippel[3]

[1]Swiss Federal Institute of Technology in Zurich
[2]ETH Zurich
[3]University of Leipzig

March 02, 2026

Abstract

Climate models simulate a wide range of 21st century warming for a given forcing scenario. Constraining this uncertainty is a central challenge in climate science because of its implications for climate policy and adaptation. The Transient Climate Response (TCR) is a key idealized metric used to quantify future warming in response to an exponentially increasing CO_2 concentration. Climate models span a range of 1.3–3 K for TCR. In attempts to constrain this range, emergent constraints on TCR based on historical temperature trends consistently pointed towards TCR values at the lower end of the range of models. However, recent evidence from trends in the short-wave and long-wave components of Earth's energy imbalance (EEI) at the top-of-atmosphere suggests that models with higher TCR lie closer to the observed EEI trends. Here, we reconcile this discrepancy and provide a revised range for TCR of 1.9–2.6 K. We show that previous temperature-based constraints were biased low due to internal variability, using a statistical variability-filtering approach. We then provide an EEI-based constraint and demonstrate that EEI trends have a strong potential to constrain future warming, due to the much larger inter-model spread in EEI compared to surface temperature. When considering the recent 2001–2024 period, our results show that both surface temperature and EEI trends support higher TCR values than previously estimated, making it increasingly difficult to exclude high sensitivity models entirely.

Hosted file

`SM_Constraint.docx` available at https://authorea.com/users/886710/articles/1391344-recent-temperature-and-energy-imbalance-trends-point-to-higher-estimates-of-future-warming

Recent Temperature and Energy Imbalance Trends Point to Higher Estimates of Future Warming

G. Gyuleva[1], E. Fischer[1], R. Knutti[1], S. Sippel[2]

[1]Institute of Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[2]Institute for Meteorology, Leipzig University, Leipzig, Germany

Key Points:

- Recent surface warming and top-of-atmosphere energy imbalance point to higher future warming than previously estimated.
- Previous constraints on TCR using observed surface temperature trends are biased low due to internal variability in recent decades.
- The energy imbalance has a strong potential to constrain future warming due to large model spread in top-of-atmosphere radiative fluxes.

Corresponding author: Gergana Gyuleva, gergana.gyuleva@env.ethz.ch

Abstract

Climate models simulate a wide range of 21st century warming for a given forcing scenario. Constraining this uncertainty is a central challenge in climate science because of its implications for climate policy and adaptation. The Transient Climate Response (TCR) is a key idealized metric used to quantify future warming in response to an exponentially increasing CO_2 concentration. Climate models span a range of 1.3–3 K for TCR. In attempts to constrain this range, emergent constraints on TCR based on historical temperature trends consistently pointed towards TCR values at the lower end of the range of models. However, recent evidence from trends in the short-wave and long-wave components of Earth's energy imbalance (EEI) at the top-of-atmosphere suggests that models with higher TCR lie closer to the observed EEI trends. Here, we reconcile this discrepancy and provide a revised range for TCR of 1.9–2.6 K. We show that previous temperature-based constraints were biased low due to internal variability, using a statistical variability-filtering approach. We then provide an EEI-based constraint and demonstrate that EEI trends have a strong potential to constrain future warming, due to the much larger inter-model spread in EEI compared to surface temperature. When considering the recent 2001–2024 period, our results show that both surface temperature and EEI trends support higher TCR values than previously estimated, making it increasingly difficult to exclude high sensitivity models entirely.

Plain Language Summary

Predicting how much the Earth will warm as carbon dioxide concentrations rise is a major challenge for climate science, with consequences for climate policy and adaptation. Different global climate models produce widely varying projections of future warming, even when using the same scenario. Previous attempts to narrow down this range based on observed 20th century warming suggested that Earth might warm less than some models predicted. However, recent data of the amount of heat energy trapped within the Earth system—the energy imbalance—indicates that Earth may actually be warming faster, aligned with models that forecast higher temperatures. We partially resolve this contradiction by showing that during the recent observational period, both temperature and the energy imbalance point to higher future warming values. We also show how previous analyses were skewed by natural, short-term climate fluctuations, known as internal variability. By filtering out this natural variability and using trends of temperature and the energy imbalance, we provide a revised very likely range of future warming. We show that the energy imbalance is a potentially strong new line of evidence to narrow our uncertainty about future warming. Our findings indicate that it becomes increasingly difficult to exclude models with strong future warming.

1 Introduction

Global climate models (GCMs) simulate a wide range of future warming in the 21st century, even when run with the same scenario of anthropogenic and natural forcings (P. Forster et al., 2021). Reducing the uncertainty in future warming is a key challenge in climate science: it is the basis for assessing carbon budgets and shaping climate policy (Matthews et al., 2021), and it informs the development of climate adaptation measures from global to regional and local levels.

The Transient Climate Response (TCR) is an idealized metric used to quantify future warming simulated by GCMs. The TCR of a model is the global surface air temperature (GSAT) increase at the time of CO_2 doubling in a simulation where the CO_2 concentration increases at 1% per year. TCR takes values between 1.3 K and 3 K across GCMs participating in Phase 6 of the Coupled Model Intercomparison Project (CMIP6) (black dots in Fig. 1).

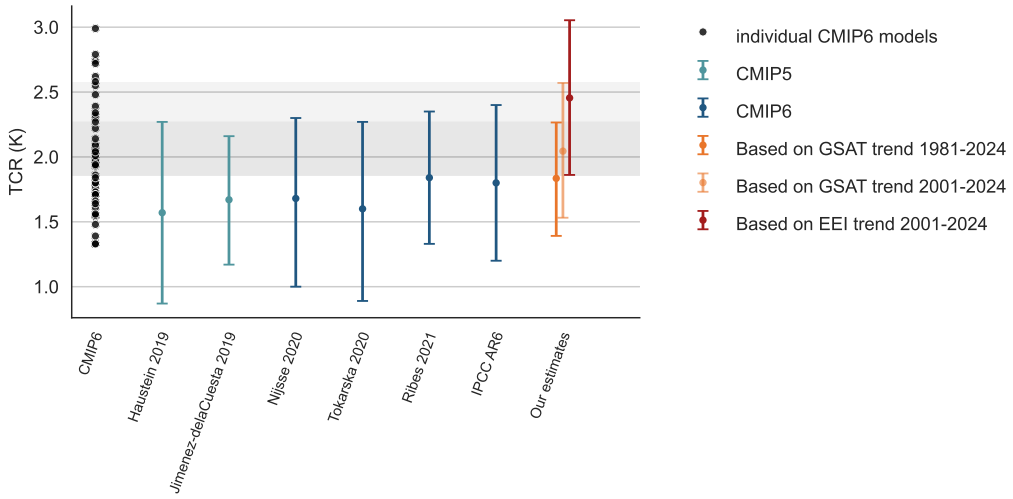

Figure 1. Overview of TCR constraints in the published literature (Haustein et al., 2019; Jiménez-de-la Cuesta & Mauritsen, 2019; Nijsse et al., 2020; Tokarska et al., 2020; Ribes et al., 2021; P. Forster et al., 2021), and our three constrained estimates, two of which are based on GSAT trends, and the other on EEI trends. Gray shading indicates the overlap region between the EEI-based range and the two GSAT-based ranges. Bars show 90% confidence ranges.

CMIP6 models simulated a wider range of future warming with a higher multi-model mean compared to CMIP5 (Zelinka et al., 2020). Some of the efforts to narrow the range led to a constrained range of 1.2–2.4 K using multiple lines of evidence (P. Forster et al., 2021; Sherwood et al., 2020). A key contributor to this assessment were emergent constraints: emergent constraints are based on a present-day observable that, when computed in climate models, exhibits a relationship with the respective model's TCR. A typical emergent constraint on TCR is based on a regression of simulated historical GSAT trends onto TCR across a multi-model ensemble of GCMs. One can then use the regression to predict based on the observable—for example the true observed historical GSAT change—and thereby obtain a prediction interval for TCR (Jiménez-de-la Cuesta & Mauritsen, 2019; Nijsse et al., 2020; Tokarska et al., 2020). Virtually all such published temperature-based constraints exclude the fastest warming models, narrowing the upper bound of the very likely (90% confidence) TCR range to below 2.4 K (Fig.1).

Recently, several studies (Armour et al., 2024; Liang et al., 2024; Myhre et al., 2025; Vogt et al., 2025) suggested that the constrained ranges are actually biased low, for two main reasons. The majority of studies point to the potential effect of sea-surface temperatures (SST) in modulating GSAT via radiative feedbacks, a phenomenon known as the pattern effect (Rugenstein et al., 2016): for example, surface warming in the Eastern equatorial Pacific leads to a positive feedback similar to an El Niño event. However, while models simulate a more El-Niño-like warming pattern with climate change, the observed warming pattern has actually been more La-Niña-like, with a cooling in the Eastern equatorial pacific (Ceppi & Gregory, 2017; Dong et al., 2019; Olonscheck et al., 2020; Seager et al., 2019; Watanabe et al., 2021; Wills et al., 2022).

This discrepancy between observed and modeled SST patterns in recent decades could potentially bias the constraint. If internal variability in surface warming patterns caused recent warming to be less strong than simulated by GCMs, then constraints relying on the observed warming trend might be biased low (Dong et al., 2021; Zhou et al., 2021). Liang et al. (2024) show this effect for the influence of Pacific variability on

the constraints. However, also if a systematic model bias were responsible for the discrepancy in the observed and modeled SST patterns, the constraints might still be biased low (Armour et al., 2024). This results from the fact that emergent constraints can only be valid if there is no common bias that all GCMs share in the variable used for the constraint, when compared with observations (Sanderson et al., 2021). The central question arising from this discussion is how much the observed SST pattern affected and still affects GSAT-based constraints. We quantify this using a statistical method trained on coupled GCMs (Gyuleva et al., 2025), making our work complementary to the study of Armour et al. (2024) who demonstrate this effect via the use of atmosphere-only simulations.

Independently of possible SST pattern effects, recent studies have identified evidence from Earth's Energy Imbalance (EEI) which possibly suggests even higher future warming (Mauritsen et al., 2025; Myhre et al., 2025; Vogt et al., 2025). This is in apparent contradiction to the existing GSAT constraints, which constrained the TCR range to lower values. Mauritsen et al. (2025) show that the observed energy imbalance is rising twice as fast as projected by climate models. Myhre et al. (2025) find that the net EEI across models is not related to future warming, however the partitioning into shortwave (SW) and long-wave (LW) components is: models exhibiting stronger future warming tend to have stronger positive (downward) SW EEI trends and stronger negative LW EEI trends (see Sec. 2.2.2). When taking into account the observed split of the EEI into its SW and LW components, observations appear to agree much better with the stronger warming models (Myhre et al., 2025). This raises the natural question whether it is possible to leverage the recently available twenty-four years of CERES satellite observations of EEI (Loeb et al., 2018) to constrain future warming, as EEI is believed to be a "more robust metric of the rate of global climate change than GSAT" (P. Forster et al., 2021).

In this study, we take into account both of the above effects and provide a revised range for TCR. First, we improve on existing GSAT-based constraints by extending the data up to 2024, and by removing an estimate of the effect of unforced SST pattern variability from the observed GSAT (and EEI) trend. We then incorporate evidence from EEI into an emergent constraint on TCR and show how EEI data jointly with GSAT can be leveraged to constrain future warming. We reconcile the apparent discrepancy between temperature-based and EEI-based evidence for constraining future warming by considering the evolution in both variables in more recent decades. Finally, we assess how the two main sources of uncertainty, one stemming from the internal variability of the climate system, the other from our structural uncertainty in the physics of climate models, contribute individually and jointly to the total uncertainty about future warming.

2 Data and methods

2.1 Data

We use GCM output from CMIP6 *historical* (1850–2014) and *ssp245* (2015–2100) simulations (Eyring et al., 2016; O'Neill et al., 2016; Tebaldi et al., 2021) for the constraints (Sec. 2.2) and for training the statistical models to remove interannual variability (Sec. 2.5). The *historical* simulations span the period 1850–2014 and are forced with all known forcings for this period. The *ssp245* simulation is a future scenario simulation branching off from the *historical* in 2015 and going until 2100, corresponding to a world in a 'middle of the road' socioeconomic scenario with a continued increase in CO_2 emissions (O'Neill et al., 2016). TCR values of GCMs are taken from Smith, Nicholls, et al. (2021).

For surface air temperature we use the CMIP6 variable *tas*. Net SW EEI at the TOA is calculated from the difference of the SW downward flux (*rsdt*) minus the reflected

short-wave upward flux (*rsut*). Net LW EEI is given by the variable *rlut*. All fluxes are defined as positive downward.

We use annual gridded surface air temperature anomalies from 1940 to 2024 from ERA5 (Hersbach et al., 2020), re-gridded to a 5°×5° resolution. The gridded ERA5 data are used as predictors for the regression models that filter internal variability (See Sec.2.5). For observed global surface air temperature (GSAT) we use the area-weighted average of the girdded ERA5 data. For short-wave (SW) and long-wave (LW) EEI data we use CERES EBAF-TOA satellite data available for 2001-2024 (Loeb et al., 2018).

2.2 Emergent constraints on TCR

We create two constraints on TCR: one relating historical GSAT trends to TCR, and one relating SW and LW EEI trends at the top-of-atmosphere (TOA) to TCR. In both cases, we fit a linear relationship between simulated GSAT trends (or SW/LW EEI trends) over a certain period and TCR across models, and then compare this with the observed trend for this period. We use the periods 1981–2024 and 2001–2024. For EEI, we can only use the period 2001–2024, as this is the available CERES period. For GSAT, we use also the longer 1981–2024 period, as longer trends are more robust and allow comparisons with previous work related to constraints which used trends starting in or around 1981 (Armour et al., 2024; Nijsse et al., 2020; Tokarska et al., 2020).

In order to create the constraint with periods extending post 2014, we extend the *historical* simulations by the *ssp-245* simulations, as this is deemed to be closest to the real world evolution (Hausfather, 2025). However, this might introduce a bias in the constraint if the observed post-2014 forcing strongly deviates from the *ssp-245* forcing. We compare forcing estimates for *ssp245* simulations (Smith, Hall, et al., 2021) with estimates of post-2014 observed forcing as published in the Indicators of Global Climate Change, (Smith et al., 2025) in order to test comparability of *ssp245* simulations and observations. The ratio of simulated to observed total and in particular anthropogenic forcing does not deviate systematically post 2014 (Fig. S1). This justifies the use of *ssp245* as an extension of *historical* simulations up to 2024. Potentially relevant regional biases (e.g., in aerosol forcing) can still not be excluded, and their impact on global variables such as EEI and GSAT is impossible to quantify given the available data.

2.2.1 GSAT-based constraint

In order for an emergent constraint to be valid, the relationship between future warming and the constraint variable needs to be based on physical principles (Armour et al., 2024; Caldwell et al., 2018). In the case of the GSAT-based constraint, fitting a linear relationship between models' historical GSAT trends and their TCR can be understood as 'models that warm more in the past will warm more in the future', assuming similar forcing agents in the two periods. This assumption is not perfect but justifiable, as CO_2 forcing dominates in recent observed trends (P. M. Forster et al., 2025; Smith et al., 2025). The physical reasoning for the existence of this constraint is that the strength of climate feedback in a model is an intrinsic model property, and has its origins in one-dimensional energy balance models (Hansen et al., 1985). Today, it is the basis of simple climate models, e.g. (Smith et al., 2018). Numerous previous constraints have been based on the relationship between past and future warming (Knutti et al., 2017).

A caveat of emergent constraints on TCR using past warming is the potential effect of the discrepancy between observed and modeled SST patterns. Armour et al. (2024) show that if this is owed to a systematic model bias, such constraints are largely invalid. Variability could also be responsible for the mismatch between observed and modeled SST patterns, rather than only model bias. We employ a statistical model that quantifies the effect of variability on GSAT and SW and LW EEI (Sec. 2.5). In the case of

GSAT, the model identifies a cooling influence of the surface temperature pattern over the last decades (Gyuleva et al., 2025), regardless of whether this is due to natural variability or possibly a forced accumulation of typically cooling internal-variability patterns in recent decades. This allows us to partially account for the bias induced by pattern effects. We note that our results are based on the assumption that variability accounts for some part of the SST-pattern discrepancy.

2.2.2 EEI-based constraint

The emergent constraint based on SW and LW EEI trends rests on a relationship to future warming recently identified by Myhre et al. (2025): Models with stronger negative SW EEI trends have stronger positive LW EEI trends, and vice-versa. We find a similar relationship with TCR (Fig. 2a). This relationship in the partitioning of the net EEI into its SW and LW components is physically justified: models with a higher TCR warm faster and therefore result in stronger SW EEI effects (e.g., positive SW cloud feedbacks, melting of glaciers and sea-ice) while at the same time exhibiting stronger LW EEI trends due to an enhanced Planck feedback (Myhre et al., 2025). Models with a weak TCR on the other hand, warm slowly in the transient run and may even exhibit still positive LW EEI trends, as the Planck feedback has not yet gained dominance over the increased trapping of LW radiation due to CO_2 increase.

The fact that all models lie close to a line in Fig. 2a is a result of the fact that their net EEI trend (sum of SW and LW) is well-constrained. However, this may be because of GCM tuning on the net EEI (Hourdin et al., 2017). The line would be a perfect line if all models simulated an identical net EEI trend. In contrast, the spread along the line, which is the partitioning of the net EEI into its SW and LW components, is large. Importantly, Myhre et al. (2025) suggest that the spread in the LW and SW partitioning correlates with climate sensitivity, and here we show that it also correlates well with TCR (Fig. 2b): by performing principal component analysis of the trends in Fig. 2a (i.e., by projecting each point orthogonally onto the straight line which goes through the point cloud such as to minimize the orthogonal distance from each point to the cloud), we obtain the relationship in Fig. 2b. From now on we refer to the first principal component loading of the SW/LW EEI trends as $\mathrm{EEI}_{\mathrm{PC1}}$. We will use this relationship to obtain an emergent constraint on TCR based on the observed EEI.

2.2.3 Fitting the constraints

We use ordinary least squares (OLS) regression with TCR as target and ensemble mean GSAT trends or $\mathrm{EEI}_{\mathrm{PC1}}$ trends as predictors for the two constraints. For the EEI-based constraint, the trends are computed on TOA net SW EEI and TOA net LW EEI separately, after which they are transformed to their value in the $\mathrm{EEI}_{\mathrm{PC1}}$ space. The transformation is done by fitting a one-component principal component analysis to the simulated SW EEI and LW EEI trends for the given time period from all members of all GCMs used for the constraint.

We use GCMs with at least five *historical* members. This is a somewhat arbitrary choice and a trade-off between reducing the noise from internal variability and including as many GCMs as possible. Fig. S2 shows the Pearson correlation between TCR and the two predictors GSAT and $\mathrm{EEI}_{\mathrm{PC1}}$ as a function of the minimum number of members and as a function of the number of GCMs included in the constraint. The qualitative conclusions do not depend on the choice of the number of members.

We fit the OLS regression for the constraint by giving equal weight to each model family. This avoids an over-representation of individual modeling centers providing more than one version of their model. The results largely remain unchanged when omitting the model-family weighting.

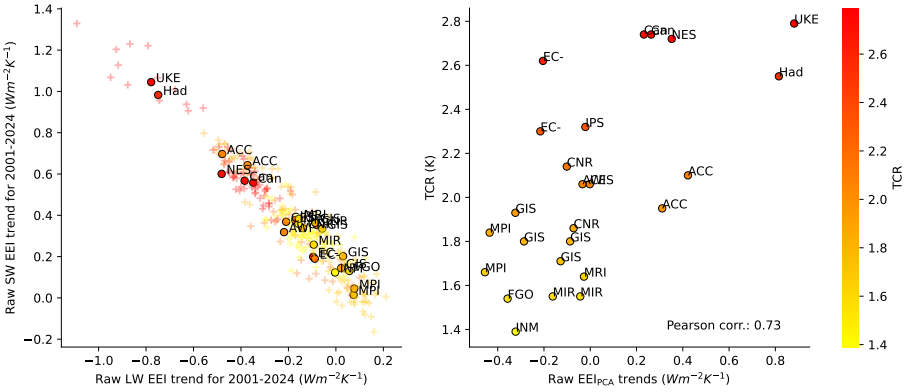

Figure 2. a) Partitioning of the net top-of-atmosphere EEI trend for 2001-2004 into its net SW and net LW components for 2001–2024. Fluxes are defined as positive downwards. Dots represent model means, crosses the individual ensemble members. Model family abbreviations can be found in Table S1. The figure is adapted for TCR from a similar figure in Myhre et al. (2025) showing the relationship with equilibrium climate sensitivity. (b): First principal component of (a) and its relationship with TCR.

In order to account for the effects of variability in the constraints, we perform a variability-filtering on the raw GSAT and SW/LW EEI trends, similar to the method in Gyuleva et al. (2025) (see Section 2.5). The main idea is that variability-filtering is a process that transforms the desired quantities GSAT, LW EEI and SW EEI into a space where their relationship with TCR still remains equally strong and physically justified, yet the noise due to variability is reduced. Note that its aim is to reduce as much of the uncertainty due to internal variability as possible, and not to provide an accurate estimate of what the true 'forced' variables are. For simplicity we still refer to variability-filtered quantities as 'forced' from now on, as opposed to those computed on the raw data ('raw').

In theory, noise in the predictor variables GSAT and EEI_{PC1} could bias the regression slope towards zero, a phenomenon known as regression dilution. Previous constraints have also used OLS regression, arguing that the OLS fit rather describes "the relationship between simulated warming and future warming than the relationship between forced warming and future warming" (Tokarska et al., 2020). In our case, since we apply variability filtering which aims to reduce this noise, we argue that a good amount of the noise due to variability in the predictors should have been removed, although the statistical model is not perfect. Hence, we use OLS regression similar to previous constraints studies on TCR (Tokarska et al., 2020).

Best-estimates of TCR are obtained by predicting TCR using the fitted OLS regression with the observed raw or forced GSAT or EEI_{PC1} trend as predictor.

2.3 Prediction interval in the constraint

To compute a prediction interval for the constraints we use the predictive distribution arising from a linear regression. The probability of a prediction y given a new observation x^* is normally distributed with a variance σ_y^2 given by

$$\sigma_y = \frac{N}{N-2}\text{MSE}\sqrt{1 + \frac{1}{N} + \frac{(x-\bar{x})^2}{N\sigma_x^2}}, \quad (1)$$

where N is the sample size, MSE the mean squared error, and $\bar{x}$ and σ_x^2 the mean and variance of the predictors (Sippel et al., 2024).

To account for uncertainty in the observed trends, we estimate the variance of (forced or raw) trends for a given period by computing them on all members of GCMs with more than 10 members. We then center all trends of a GCM to have zero mean and pool the trends of all GCMs together. We estimate the density of this trend distribution via a normal distribution while weighting the samples to ensure equal weight for each GCM regardless of sample size (Fig. S3 shows the trend distributions with the normal distribution fit). The uncertainty estimates remain virtually unchanged if a non-parametric density estimation is fitted to the modeled trend distribution instead of the normal approximation. We therefore use the normal approximation for simplicity.

To obtain the uncertainty in the prediction of TCR around the observed trend, we sample 100,000 times from the estimated density of modeled trends. For each such sample x^*, we compute the predictive distribution using Eq. 1 and sample from it once to obtain a corresponding y^*. The 90% uncertainty interval in our TCR estimate is then given by the 5%-95% quantiles of all sampled y^*.

2.4 Decomposing the uncertainty into variability and model error

We decompose the total prediction error into its contributions from internal variability and structural model error. The uncertainty originating from GCM structural errors would correspond to the error coming only from the scatter in the points—equivalent to the prediction uncertainty remaining if we assume perfect knowledge of the observed trend. We therefore quantify it by assuming zero noise due to variability in the observed trend and computing the 5-95% prediction interval based on Eq. 1 at the actual observed trend.

The noise from internal variability, on the other hand, would correspond to the uncertainty in the prediction if we knew the perfect relationship of the predictor with TCR, and sampled only the uncertainty in the observed trend. We therefore quantify the uncertainty from internal variability by assuming a perfect linear relationship between TCR and the predictor and perform the following perfect-model test: for each GCM with more than 10 members, we leave this particular GCM out when fitting the constraint regression on all other GCMs, and then predict the TCR for the left-out GCM on each of its members. We compute the variance of these predictions for each GCM and then pool the variances across GCMs. For a 5-95% confidence interval we use the corresponding quantiles of a normal distribution with variance equal to the pooled variance. Again, the results remain largely unchanged if a non-parametric density estimation is fitted rather than a normal distribution.

2.5 Ridge regression for filtering internal variability

Following the method of Gyuleva et al. (2025), we train a ridge regression model to predict the contribution of variability to annual GSAT, and to annual SW and LW EEI. We give a brief overview of the method here, details can be found in Gyuleva et al. (2025).

Global variables GSAT (CMIP6 variable name: *tas*), LW EEI (CMIP6 variable name: -*rsut*), and SW EEI (CMIP6 variable name: *rsdt* - *rsut*) are calculated as area-weighted global averages of the gridded model output.

For each of the global variables GSAT, SW EEI, and LW EEI, we define variability in *historical* and *ssp245* simulations as the deviation from the ensemble mean in a GCM. We then obtain a vector $\mathbf{y}$ with entries for each year of each member of the selected GCMs containing the variability contribution to the global variable in that year.

We train a ridge regression model to predict $\mathbf{y}$ by using the corresponding ensemble member's annual gridded GSAT anomalies as predictors. We then seek an optimal vector of regression coefficients $\hat{\boldsymbol{\beta}}$, which for a given year $\mathbf{x_i}$ predicts the variability in the global variable for the given year, $\hat{y}_i$ via the scalar product:

$$\hat{y}_i = \mathbf{x_i}^T \cdot \hat{\boldsymbol{\beta}} \tag{2}$$

The training data consist of a matrix X and response vector $\mathbf{y}$. The rows of X contain the individual $\mathbf{x_i}$ for each year and ensemble member. The components of $\mathbf{y}$ contain the corresponding annual internal variability contribution. The columns of X are the different predictors (grid cells). We standardize the columns of X to have mean zero and variance one before training.

The optimal coefficient vector $\hat{\boldsymbol{\beta}}$ is found by minimizing the ridge loss over the training data, which consists of the squared prediction error as in linear regression, and of an additional penalty term which aims to shrink the individual components of $\hat{\boldsymbol{\beta}}$ towards zero:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} ||\mathbf{y} - X\boldsymbol{\beta}||_2^2 + \lambda ||\boldsymbol{\beta}||_2^2$$

The penalty term reduces the variance of the estimator in the case of highly correlated predictors, as is the case for gridded geospatial data, and performs very well on climate data (Sippel et al., 2020, 2021; Gyuleva et al., 2025). The strength of the penalty is governed by the hyperparameter λ, which is estimated using cross-validation, see Fig. S4 (Gyuleva et al., 2025).

With the three fitted ridge models, we predict for each year of each ensemble member of each GCM the corresponding variability contributions to annual GSAT, LW EEI and SW EEI. We then obtain variability-filtered ensemble members of GSAT, LW EEI and SW EEI by subtracting these contributions from the original ensemble member.

To filter variability from observations, we use annual ERA5 gridded surface air temperature anomalies as predictors. Predictions with the three ridge regression models yield estimates of the annual variability contribution to observed GSAT, LW EEI and SW EEI. We subtract this variability estimate from the annual observational GSAT, LW EEI and SW EEI value to obtain 'forced' estimates.

The performance of the GSAT ridge regression model is discussed in detail in Gyuleva et al. (2025) and displayed in Figs. S5-S6 for SW and LW EEI variability. The SW and LW EEI ridge regression model overall shows a good performance in capturing both interannual and decadal variability in SW and LW EEI fluxes (Fig. S7-S8), justifying its use in this work.

3 Results & Discussion

3.1 Comparison of EEI and GSAT constraints

Emergent constraints based on raw historical GSAT trends for 1981-2024, and raw historical $\mathrm{EEI}_{\mathrm{PC1}}$ trends for 2001-2024, produce very different TCR ranges: 1.2–2.2 K for the GSAT-based constraint and 1.8–3.0 K for the EEI-based constraint (Fig. 3a, b). The GSAT-based best estimate of 1.7 K lies just below the EEI-based lower bound, while

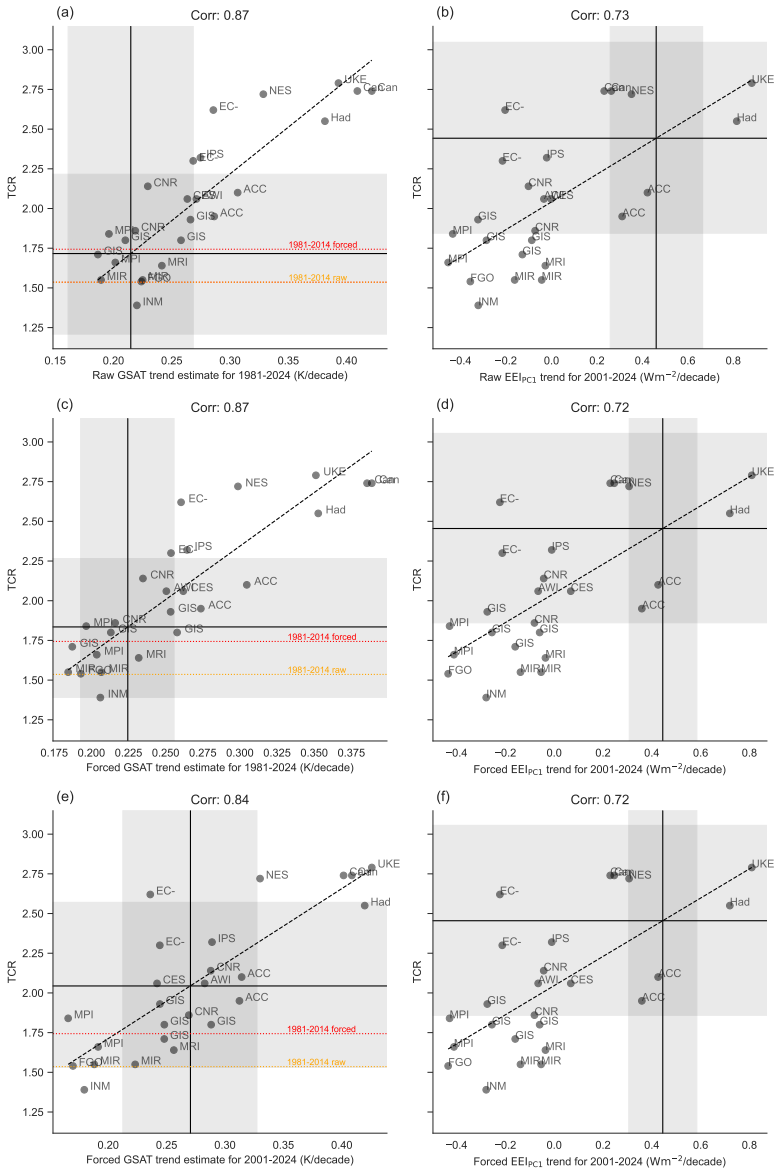

Figure 3. Emergent constraints based on (a): raw GSAT trends for 1981–2024; (b): raw EEI_{PC1} trends for 2001–2024, (c): forced GSAT trends for 1981–2024; (d) and (f): forced EEI_{PC1} trends based on 2001–2024, (e): forced GSAT trends based on 2001–2024. Dots represent model means, labels show the model family (see Table S1). Solid vertical lines show the observational estimate for the quantity on the horizontal axis estimated from ERA5 surface air temperature data for GSAT and CERES data for the EEI_{PC1}. Solid horizontal lines show the corresponding TCR best-estimate. Shadings indicate the two-sided 90% confidence range (5^{th}-95^{th} quantiles) of the estimated distribution of corresponding trends from large ensembles on the horizontal axis, and the 90% prediction interval (5^{th}-95^{th} quantiles of the predictive distribution) for TCR on the vertical axis (see Sec. 2.3). Dotted horizontal lines indicate the raw (orange) versus forced (red) GSAT-based constraint using the period 1981–2014, which formed the constraint in Tokarska et al. (2020). Note that panels (d) and (f) are identical for illustration purposes.

the EEI_{PC1}-based best-estimate is 2.4 K—outside of the very likely range for the GSAT-based constraint (horizontal black lines in Fig. 3a, b). No GCM simulates both a GSAT trend and an EEI_{PC1} trend compatible with the observed trend within the 90% confidence range. Further, the TCR best estimate from either constraint is not included in the other constraint's prediction interval. This mismatch could have several reasons: a) observational error, b) internal variability influencing the raw trends (vertical black lines in Fig. 3a,b), c) an inherent bias in the way models simulate EEI and GSAT trends, or d) a shift of processes and feedbacks taking place in observations between the two overlapping trend periods used, 1981–2024 and 2001–2024.

We find that using a different GSAT or GMST dataset has little effect, and the difference between different surface temperature datasets is much smaller than the uncertainty from variability (Lenssen et al., 2024; Rohde & Hausfather, 2020; Morice et al., 2021). For EEI data, such a test is impossible as CERES is the only currently available satellite product. Without a valid alternative dataset to compare with, there may be large uncertainty and potential biases in CERES. However, the time evolution of EEI from CERES is remarkably similar to the independent ARGO dataset using ocean heat uptake as a proxy for EEI (Johnson et al., 2016; Riser et al., 2016), providing confidende in the trend estimated from CERES.

The incompatibility of the raw GSAT-based constraint over 1981–2024 and the raw EEI-based constraint over 2001–2024 could be partially owed to variability acting in different ways during the two periods: if variability overall cooled the GSAT trend relative to the models' distribution during 1981–2024, whereas it pushed the EEI_{PC1} trend upward relative to the models' during 2001-2024, this could produce two apparently incompatible regions for the TCR predictions. We would also not expect any of the model means to capture such a shift, as variability is largely absent from the model mean.

We test this by computing the constraints on variability-filtered data (Fig. 3c,d). Variability-filtering leads to an upward correction of the 1981–2024 GSAT trend by 0.010 K/decade, and an upward correction of 0.015 Wm^{-2}/decade for the EEI_{PC1} trend (-0.027 Wm^{-2}/decade for the SW EEI trend and $+0.026$ Wm^{-2}/decade for the LW EEI trend). The effect on the TCR best estimate is an upward correction of 0.1 K for the GSAT constraint and a negligible effect for the EEI constraint. Note that the models' position in the constraint plot also shift during variability filtering, so any shift in the observed trend's position is only relevant if it changes the position relative to the models'. However, the variability effect on the GSAT trend was much larger (0.026 K/decade) for the period 1981–2014 used in Tokarska et al. (2020), leading to a 0.21 K higher TCR estimate if accounting for variability (orange and red dotted lines in Fig. 3a). This is likely due to variability-driven cooling during the hiatus period as identified in Gyuleva et al. (2025). Previous estimates based on trends ending in 2014 are thus biased low by around 0.3 K according to our analysis, partly due to variability acting during the 1981–2014 period, and partly due to the extension of the trend period up to 2024 (compare horizontal black line with the orange and red lines in Fig. 3c).

Variability filtering does not resolve the partial incompatibility of simulated GSAT and EEI trends, as the uncertainty bands have now become smaller but have not shifted substantially. Even when attempting to filter variability with our method, no model is able to fall within the uncertainty remaining around both of the two observed trends, GSAT and EEI_{PC1}. Therefore, either an inherent model bias in the link between GSAT and EEI in GCMs, or a change in processes and feedbacks in observations between the two time periods 1981-2024 and 2001-2024, or a combination of both, is needed to explain the mismatch.

Our results provide some evidence for both. When performing the GSAT constraint on the more recent period 2001-2024 used for the EEI constraint (Fig. 3e), the mismatch is partially resolved: the TCR best-estimate is now 2.0 K, compared to 1.8 K for the 1981–

2024 period. There is now one model family (ACC, see Table S1), which simulates both GSAT and EEI trends within the uncertainty due to variability remaining after the filtering, and the TCR best estimates of the GSAT and EEI constraints are compatible with each other's prediction intervals.

This poses the question whether the GSAT-based constraint based on 2001–2024 is incompatible with the one based on 1981–2024. We test this by seeing how much TCR-best-estimate predictions vary for different initial condition members of the same model in a perfect-model setup (i.e., when leaving the model out for fitting the constraint relationship, see Sec 2.4). Both for the forced and the raw TCR best estimate based on 1981–2024, the best estimate based on 2001–2024 lies within the estimated 95% confidence region from the perfect-model test (Fig.4a, compare yellow and orange error bars for the two periods). We therefore cannot exclude variability as a cause of the shift in the GSAT-based constraint when shifting between the two periods.

Interpreting the partial reconciliation of GSAT and EEI-based constraints when using the same period 2001–2024 is not straightforward: according to our results, accounting for variability does not reconcile the mismatch. However, even after variability filtering uncertainty from the initial conditions still remains (gray shading around vertical lines in Fig. 3c-f), as our method is not perfect. It is also possible that applying the regression model on observations, which come from a different distribution, introduces an unknown bias. Although our ridge regression models perform very well when tested on model families not used in training (Figs. S5-S8 and Gyuleva et al. (2025)), we still cannot exclude an 'unknown unknown' in terms of the distribution shift between models and observations. Nonetheless, considering the estimated magnitude of variability in models about the observed trend, it can be seen that even if the 1981–2024 GSAT constraint in Fig. 3c shifted to the rightmost region of the 90% confidence band, and the 2001–2024 EEI-based constraint in Fig. 3d to the left-most region of its 90% confidence band, there would still hardly be models simulating both GSAT and EEI consistently with each other.

This gives us some confidence that variability is most likely not the main cause of the discrepancy. Since accounting for variability does not seem to resolve the discrepancy observed in Figures 3a,b and 3c,d, then we should expect this discrepancy to persist regardless of the period in consideration if it is owed to a systematic model bias in the GSAT-EEI coupling. The fact that it partly becomes resolved by aligning the two periods suggests that processes in the real world might have changed between 1981–2024 and 2001–2024 in such a way as to make higher future warming more likely. Even so, aligning the periods of the constraint to 2001–2024 does not fully resolve the discrepancy, so a part of it may still be owed to model biases, some of which have been identified in literature (Olonscheck & Rugenstein, 2024).

3.2 Limits and opportunities of emergent constraints on TCR

Next, we decompose the sources of uncertainty in the constraints. Fig. 4b shows the 90% prediction interval for the constraints (blue bars). The prediction interval incorporates two types of uncertainty: uncertainty from internal variability, i.e., the uncertainty in the position of the vertical line (observed trend) in Figs. 3a-f; and structural model uncertainty leading to an imperfect relationship between TCR and the constraint variable (here, GSAT or EEI_{PCA} trends). This uncertainty is represented by the scatter in the points around the regression line in Fig. 3. Note that the two types of uncertainty do not contribute linearly to the joint uncertainty (see Eq.1), so we do not expect the orange and green bars to add up to the blue. We also note that part of the structural uncertainty in the scatter of the points might still be due to internal variability, as some models only provide few or even a single member to compute their TCR val-

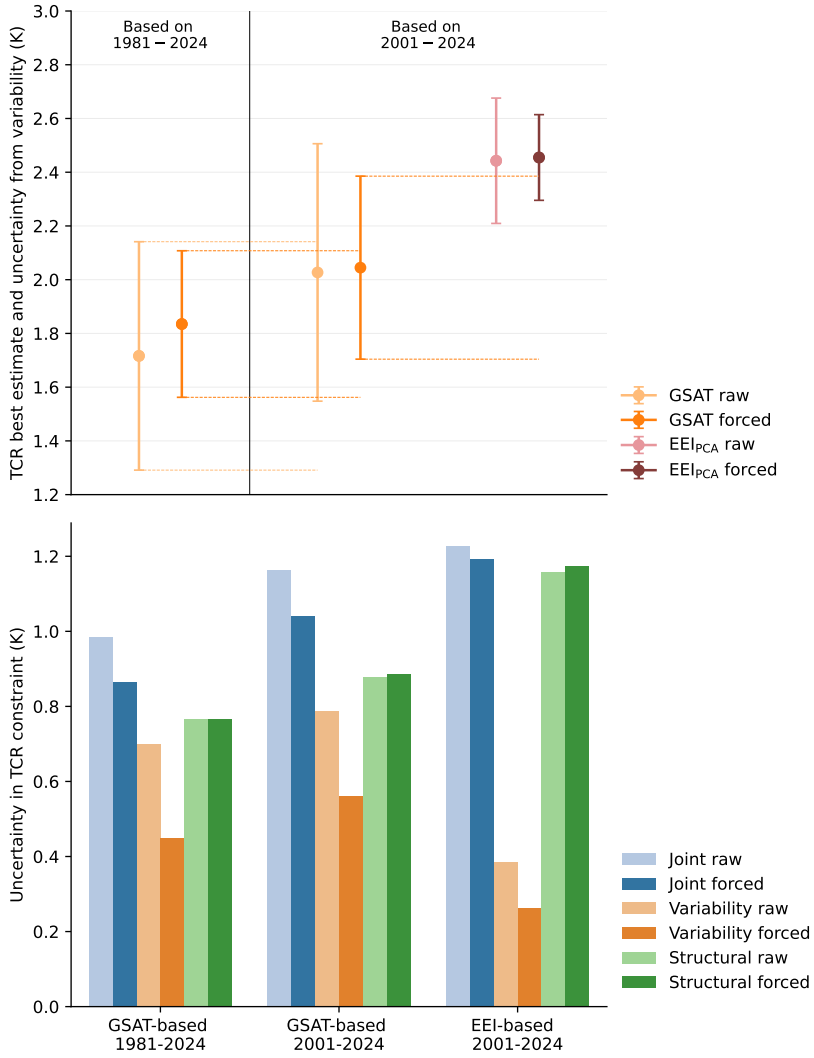

Figure 4. (a): Best-estimate of TCR and 2.5%-97.5% uncertainty range due to internal variability only. The error bars represent two times the standard deviation of TCR predictions across initial condition ensembles, computed from the pooled variance across models in a perfect-model setup (see Sec. 2.4). (b) Break-down of the total prediction uncertainty (cf., gray horizontal shading in Fig.3) into uncertainty coming only from internal variability or GCM structural errors. The bars represent 90% confidence ranges (the two-sided 5%–95% range).

ues, and even in ensembles with at least five members, substantial noise may still remain after computing the ensemble mean.

The joint uncertainty is highest for the EEI-based constraint and lowest for the long period GSAT-based constraint. However, the dominant sources of uncertainty differ between GSAT- and EEI-based constraints. Whereas for GSAT, variability and structural uncertainty contribute roughly equally to the joint uncertainty, for the EEI-based constraint structural uncertainty is larger than the relatively small contribution from variability.

Our results suggest that the uncertainty in the still short observed 2001–2024 EEI trend is not a limiting factor for better constraining TCR. This is because the difference in the ensemble means of EEI_{PCA} trends between GCMs is much larger than the differences within one GCM (Fig. S4). This can be seen in the narrow uncertainty around the observed trend in Figs. 3b,d,f compared to the larger uncertainty about the GSAT trend in Figs. 3a,c,e, in relation to the overall horizontal spread of ensemble means. This has two implications: first, the currently available CERES observational period is sufficiently long to derive a valid constraint, and waiting longer will not help the signal to emerge. Second, due to the much larger differences between GCMs compared to within GCMs, EEI-based constraints have much more power in discriminating between GCMs, provided that the relationship with TCR is strong enough. Structural model uncertainty dominates the EEI constraint and progress is needed there in order to reduce the uncertainty in EEI-based TCR constraints.

Filtering variability in GSAT leads to an overall larger reduction in total uncertainty than removing variability in EEI. This is because the two sources of uncertainty contribute non-linearly to the joint uncertainty and a reduction of the larger of the two has a much bigger effect on the total. Due to the much larger effect of variability on the GSAT-based constraint, longer trend periods in GSAT may well improve our overall uncertainty, yet progress is needed in reducing both structural and internal variability uncertainty in order to improve the constraint. We note that even after variability-filtering, substantial uncertainty remains, which is likely irreducible (Huber et al., 2014).

We note that most models are directly or indirectly tuned to match the observed temperature record and the net energy imbalance (Hourdin et al., 2017). However, to our knowledge there is no systematic tuning of models to SW or LW fluxes separately. This might explain why models tend to agree much more in GSAT and in net EEI, than they do in the SW-versus-LW partitioning of net EEI. Our results therefore also challenge the process of model tuning: if we tune models to an observable, then this particular observable loses its power in constraining models. Models might be then artificially forced into a state where they are 'right', but for the wrong reasons (Kiehl, 2007; Knutti, 2008).

3.3 Assessing a likely TCR range

Based on the above, we attempt a revision of the TCR best-estimate. We believe that both GSAT-based constraints and the new EEI-based constraints provide important evidence to constrain TCR. If we assume that the "better" GCMs tend to be those that simulate both metrics close to the observed trends within the uncertainties, our new informed guess for a very likely TCR range should be based on an overlap of the GSAT- and EEI-based constraints. The new estimate for the lower bound of the very likely TCR range is therefore 1.9 K, as this is the lowest overlap point between the EEI-based and GSAT-based TCR prediction intervals. This lower bound is higher than all previous constraint-based TCR best-estimates in the literature (Fig. 1).

Defining a likely upper bound for TCR is more difficult: if we were to take a statistically conservative approach, we would use the upper bound of the 1981-2024 GSAT-based constraint, because GSAT-based constraints are strongly impacted by internal variability in shorter trend periods. This gives an upper bound of 2.3 K, which still lies within the range of previous GSAT-based TCR estimates from the CMIP6 era (Fig. 1). However, our results provide some evidence for a shift of the rate of observed warming in recent decades, which is closer to that of the faster-warming GCMs. Based on this reasoning, a better physically justified upper bound would be the upper bound of the GSAT-based constraint based on the recent period 2001-2024. This would result in an upper bound on TCR of 2.6 K, which is higher than previous TCR estimates' upper bounds.

Since we do not have evidence to rule out the 2.6 K upper bound, we propose a revised TCR range of 1.9–2.6 K (total gray shading in Fig. 1).

We acknowledge that two main factors might bias our results. First, despite the overall similarity between the *ssp-245* and real world forcing estimates (Smith, Hall, et al., 2021; Smith et al., 2025), differences in the regional distribution of climate forcers, particularly aerosols, might have impacted the observed EEI or GSAT differently than the forcing in the models (Hodnebrog et al., 2024; Samset et al., 2025). The updated forcing used for the upcoming 7$^{\text{th}}$ cycle of CMIP might help shed light on this and reveal if relevant biases due to diverging forcing are affecting our results.

Second, like most emergent constraint studies, we assume no strong structural bias common to all models relative to observations (Sanderson et al., 2021). Armour et al. (2024) identify such a bias for GSAT-based constraints, and we acknowledge that this might affect our work. However, we believe this work is justified nonetheless for two reasons: first, the SST pattern bias identified by Armour et al. (2024) is for the period 1981–2014, without including the data up to 2024. Recent years have been marked by the remarkable 2023–2024 surge in GSAT and a strong El Niño event (Cattiaux et al., 2024; Raghuraman et al., 2024; Seeber et al., 2026; Terhaar et al., 2025). It is thus not clear whether the evolution of the SST pattern from 2014 onward still points to a substantial bias between models and observations with a potential to bias the constraints. Second, we counteract the effect of the SST pattern via our variability-filtering approach. Although the filtering is not perfect, it accounts for some of the effect that SST pattern feedbacks have on GSAT and the EEI, even if the specific evolution may not be represented by most models.

4 Conclusions

Our results introduce two improvements to previous emergent constraint studies based on historical trends, by accounting for the effect of variability on observed trends, and by considering a new metric based on the EEI.

We show that observations of TOA energy fluxes may provide a complementary piece of evidence in constraining future warming, particularly due to the large spread between models. A constraint based on observed EEI points to higher future warming than the previously published temperature-based constraints. This discrepancy between temperature- and EEI-based constraints is partially resolved when aligning the periods used for the constraint to the last two decades—providing some evidence for a possible shift of observed climate processes towards stronger warming with increasing CO_2 concentration.

We also reveal that past temperature-based constraints on TCR were biased low due to a cooling influence of variability in previous decades. More recent temperature trends point to higher future warming, although variability cannot be excluded as a cause of this shift.

We show that internal variability is a larger source of uncertainty for temperature-based constraints compared to EEI-based constraints, due to the much smaller spread between models in GSAT compared to EEI. With this in mind, the coming years may, by making the observational GSAT trend longer and so more robust, confirm or reject the hypothesis that observations are shifting towards stronger warming models than previously expected. For the EEI, the large spread within GCMs means that internal variability in observed trends is not the dominant source of uncertainty, which implies that the EEI could serve as a potentially powerful new metric to use in constraints of future warming.

With the advent of the CMIP7 generation of models, tools for constraining the spread in warming projections are needed. We provide a tool for accounting for variability, and a new fundamental constraint metric. A revision of the TCR range upwards has implications for future policy making. An update to the calibration of simple climate models, carbon budgets, and climate models may be required, if the faster-warming models are indeed the more realistic ones.

Open Research Section

All data used in this study are from publicly available sources. Climate model simulations from CMIP6 are available at https://esgf-node.llnl.gov/projects/cmip6/ and were accessed through the ETH Zurich Next-Generation Archive (Brunner et al., 2020). 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 who support CMIP6 and ESGF.

ERA5 reanalysis data (Hersbach et al., 2020) was accessed via
https://doi.org/10.24381/cds.f17050d7

CERES satellite data with TOA radiative fluxes (Loeb et al., 2018) was accessed via
https://ceres.larc.nasa.gov/data/

GMST observational data sources used for sensitivity tests of our results (not present in any figures) are listed below.
Berkeley Earth (Rohde & Hausfather, 2020):
https://berkeleyearth.org/data/
GISTEMP v4 (Lenssen et al., 2024):
https://data.giss.nasa.gov/gistemp/
HadCRUT5 (Morice et al., 2021):
https://www.metoffice.gov.uk/hadobs/hadcrut5/data/HadCRUT.5.0.2.0/download.html
The code used for reproducing the analysis and all figures in the study is available at Zenodo via the DOI 10.5281/zenodo.18631613 (Gyuleva et al., 2026).

Conflict of interest disclosure statement

The authors declare there are no conflicts of interest for this manuscript.

Acknowledgments

We thank Chris Smith for insights and access to forcing data. We thank Nina Effenberger, Robin Noyelle and Konstantin Weber for insightful discussions about the project. G.G., S.S. and R.K. acknowledge the project 'Constraints on near-term warming projections via distributionally robust statistical and machine learning' (COPE; grant agreement C22-02, funded by the Swiss Data Science Center). S.S. acknowledges funding provided by the German Research Foundation through the Heinz Maier-Leibnitz Prize 2024 and the European Union's Horizon Europe programme via the project 'Artificial Intelligence for Enhanced Representation of Processes and Extremes in Earth System Models' (AI4PEX; grant agreement 101137682). G.G. and R.K. acknowledge funding by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101003536 (ESM2025–Earth System Models for the Future).

References

Armour, K. C., Proistosescu, C., Dong, Y., Hahn, L. C., Blanchard-Wrigglesworth, E., Pauling, A. G., ... Gregory, J. M. (2024, March). Sea-surface temperature pattern effects have slowed global warming and biased warming-based constraints on climate sensitivity. *Proceedings of the National Academy of Sciences*, *121*(12), e2312093121. Retrieved 2025-08-05, from https://www.pnas.org/doi/10.1073/pnas.2312093121 doi: 10.1073/pnas.2312093121

Brunner, L., Hauser, M., Lorenz, R., & Beyerle, U. (2020, March). The ETH Zurich CMIP6 next generation archive: technical documentation. Retrieved 2026-02-11, from https://zenodo.org/records/3734128 doi: 10.5281/zenodo.3734128

Caldwell, P. M., Zelinka, M. D., & Klein, S. A. (2018, May). Evaluating Emergent Constraints on Equilibrium Climate Sensitivity. *Journal of Climate*, *31*(10), 3921–3942. Retrieved 2026-01-22, from https://journals.ametsoc.org/view/journals/clim/31/10/jcli-d-17-0631.1.xml doi: 10.1175/JCLI-D-17-0631.1

Cattiaux, J., Ribes, A., & Cariou, E. (2024). How Extreme Were Daily Global Temperatures in 2023 and Early 2024? *Geophysical Research Letters*, *51*(19), e2024GL110531. Retrieved 2024-10-14, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2024GL110531 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2024GL110531) doi: 10.1029/2024GL110531

Ceppi, P., & Gregory, J. M. (2017). Relationship of tropospheric stability to climate sensitivity and Earth's observed radiation budget. *Proceedings of the National Academy of Sciences*, *114*(50), 13126–13131.

Dong, Y., Armour, K. C., Proistosescu, C., Andrews, T., Battisti, D. S., Forster, P. M., ... Shiogama, H. (2021). Biased Estimates of Equilibrium Climate Sensitivity and Transient Climate Response Derived From Historical CMIP6 Simulations. *Geophysical Research Letters*, *48*(24), e2021GL095778. Retrieved 2023-07-06, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2021GL095778 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021GL095778) doi: 10.1029/2021GL095778

Dong, Y., Proistosescu, C., Armour, K. C., & Battisti, D. S. (2019). Attributing historical and future evolution of radiative feedbacks to regional warming patterns using a Green's function approach: The preeminence of the western Pacific. *Journal of Climate*, *32*(17), 5471–5491.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016, May). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. Retrieved 2023-07-12, from https://gmd.copernicus.org/articles/9/1937/2016/ doi: 10.5194/gmd-9-1937-2016

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., ... Zhang, H. (2021). The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity. In V. Masson-Delmotte 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. 923–1054). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. (Type: Book Section)

Forster, P. M., Smith, C., Walsh, T., Lamb, W. F., Lamboll, R., Cassou, C., ... Zhai, P. (2025, June). Indicators of Global Climate Change 2024: annual update of key indicators of the state of the climate system and human influence. *Earth System Science Data*, *17*(6), 2641–2680. Retrieved 2026-02-

13, from https://essd.copernicus.org/articles/17/2641/2025/ doi: 10.5194/essd-17-2641-2025

Gyuleva, G., Knutti, R., & Sippel, S. (2025). Combination of Internal Variability and Forced Response Reconciles Observed 2023–2024 Warming. *Geophysical Research Letters*, *52*(14), e2025GL115270. Retrieved 2025-08-18, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2025GL115270 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2025GL115270) doi: 10.1029/2025GL115270

Gyuleva, G., Sippel, S., Knutti, R., & Fischer, E. (2026, February). *Code for the study "Recent Temperature and Energy Imbalance Trends Point to Higher Estimates of Future Warming" (submitted to Earth's Future on Feb. 13th 2026).* Zenodo. Retrieved 2026-02-13, from https://zenodo.org/records/18631614 doi: 10.5281/zenodo.18631614

Hansen, J., Russell, G., Lacis, A., Fung, I., Rind, D., & Stone, P. (1985, August). Climate Response Times: Dependence on Climate Sensitivity and Ocean Mixing. *Science*, *229*(4716), 857–859. Retrieved 2025-10-27, from https://www.science.org/doi/10.1126/science.229.4716.857 doi: 10.1126/science.229.4716.857

Hausfather, Z. (2025, January). An assessment of current policy scenarios over the 21st century and the reduced plausibility of high-emissions pathways - Zeke Hausfather, 2025. *Dialogues on Climate Change*. Retrieved 2026-01-30, from https://journals.sagepub.com/doi/10.1177/29768659241304854

Haustein, K., Otto, F. E. L., Venema, V., Jacobs, P., Cowtan, K., Hausfather, Z., ... Schurer, A. P. (2019, August). A Limited Role for Unforced Internal Variability in Twentieth-Century Warming. Retrieved 2024-10-15, from https://journals.ametsoc.org/view/journals/clim/32/16/jcli-d-18-0555.1.xml doi: 10.1175/JCLI-D-18-0555.1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... others (2020). *ERA5 monthly averaged data on single levels from 1940 to present [Dataset].* (Copernicus Climate Change Service (C3S) Climate Data Store (CDS)) doi: 10.24381/cds.f17050d7

Hodnebrog, , Myhre, G., Jouan, C., Andrews, T., Forster, P. M., Jia, H., ... Schulz, M. (2024, April). Recent reductions in aerosol emissions have increased Earth's energy imbalance. *Communications Earth & Environment*, *5*(1), 166. Retrieved 2025-11-06, from https://www.nature.com/articles/s43247-024-01324-8 doi: 10.1038/s43247-024-01324-8

Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., ... Williamson, D. (2017, March). The Art and Science of Climate Model Tuning. *Bulletin of the American Meteorological Society*, *98*(3), 589–602. Retrieved 2026-01-20, from https://journals.ametsoc.org/view/journals/bams/98/3/bams-d-15-00135.1.xml doi: 10.1175/BAMS-D-15-00135.1

Huber, M., Beyerle, U., & Knutti, R. (2014). Estimating climate sensitivity and future temperature in the presence of natural climate variability. *Geophysical Research Letters*, *41*(6), 2086–2092. Retrieved 2026-01-20, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2013GL058532 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2013GL058532) doi: 10.1002/2013GL058532

Jiménez-de-la Cuesta, D., & Mauritsen, T. (2019). Emergent constraints on Earth's transient and equilibrium response to doubled CO\textsubscript2 from post-1970s global warming. *Nature Geoscience*, *12*(11), 902–905.

Johnson, G. C., Lyman, J. M., & Loeb, N. G. (2016, July). Improving estimates of Earth's energy imbalance. *Nature Climate Change*, *6*(7), 639–640. Retrieved 2026-01-26, from https://www.nature.com/articles/nclimate3043 doi: 10.1038/nclimate3043

Kiehl, J. T. (2007). Twentieth century climate model response and climate sensitivity. *Geophysical Research Letters*, *34*(22). Retrieved 2025-11-07, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2007GL031383 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007GL031383) doi: 10.1029/2007GL031383

Knutti, R. (2008). Why are climate models reproducing the observed global surface warming so well? *Geophysical Research Letters*, *35*(18). Retrieved 2026-01-26, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2008GL034932 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2008GL034932) doi: 10.1029/2008GL034932

Knutti, R., Rugenstein, M. A., & Hegerl, G. C. (2017). Beyond equilibrium climate sensitivity. *Nature Geoscience*, *10*(10), 727–736.

Lenssen, N., Schmidt, G. A., Hendrickson, M., Jacobs, P., Menne, M. J., & Ruedy, R. (2024). A NASA GISTEMPv4 Observational Uncertainty Ensemble [Dataset]. *Journal of Geophysical Research: Atmospheres*, *129*(17), e2023JD040179. Retrieved 2024-12-16, from https://data.giss.nasa.gov/gistemp/ (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2023JD040179) doi: 10.1029/2023JD040179

Liang, Y., Gillett, N. P., & Monahan, A. H. (2024, June). Accounting for Pacific climate variability increases projected global warming. *Nature Climate Change*, *14*(6), 608–614. Retrieved 2025-07-04, from https://www.nature.com/articles/s41558-024-02017-y doi: 10.1038/s41558-024-02017-y

Loeb, N. G., Doelling, D. R., Wang, H., Su, W., Nguyen, C., Corbett, J. G., ... Kato, S. (2018, January). Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product. *Journal of Climate*, *31*(2), 895–918. Retrieved 2025-11-26, from https://journals.ametsoc.org/view/journals/clim/31/2/jcli-d-17-0208.1.xml doi: 10.1175/JCLI-D-17-0208.1

Matthews, H., D., Tokarska, K. B., Rogelj, J., Smith, C. J., MacDougall, A. H., Haustein, K., ... Knutti, R. (2021, January). An integrated approach to quantifying uncertainties in the remaining carbon budget. *Communications Earth & Environment*, *2*(1), 7. Retrieved 2026-02-03, from https://www.nature.com/articles/s43247-020-00064-9 doi: 10.1038/s43247-020-00064-9

Mauritsen, T., Tsushima, Y., Meyssignac, B., Loeb, N. G., Hakuba, M., Pilewskie, P., ... Zelinka, M. D. (2025). Earth's Energy Imbalance More Than Doubled in Recent Decades. *AGU Advances*, *6*(3), e2024AV001636. Retrieved 2025-05-16, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2024AV001636 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2024AV001636) doi: 10.1029/2024AV001636

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., ... Simpson, I. R. (2021). An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set [Dataset]. *Journal of Geophysical Research: Atmospheres*, *126*(3), e2019JD032361. Retrieved 2024-12-16, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2019JD032361 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019JD032361) doi: 10.1029/2019JD032361

Myhre, G., Hodnebrog, , Loeb, N., & Forster, P. M. (2025, June). Observed trend in Earth energy imbalance may provide a constraint for low climate sensitivity models. *Science*, *388*(6752), 1210–1213. Retrieved 2025-07-04,

from https://www.science.org/doi/10.1126/science.adt0647 doi: 10.1126/science.adt0647

Nijsse, F. J., Cox, P. M., & Williamson, M. S. (2020). Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models. *Earth System Dynamics*, *11*(3), 737–750.

Olonscheck, D., & Rugenstein, M. (2024). Coupled Climate Models Systematically Underestimate Radiation Response to Surface Warming. *Geophysical Research Letters*, *51*(6), e2023GL106909. Retrieved 2025-11-06, from https://onlinelibrary.wiley.com/doi/abs/10.1029/2023GL106909 (_eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2023GL106909) doi: 10.1029/2023GL106909

Olonscheck, D., Rugenstein, M., & Marotzke, J. (2020). Broad consistency between observed and simulated trends in sea surface temperature patterns. *Geophysical Research Letters*, *47*(10), e2019GL086773.

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., ... others (2016). The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482.

Raghuraman, S. P., Soden, B., Clement, A., Vecchi, G., Menemenlis, S., & Yang, W. (2024, October). The 2023 global warming spike was driven by the El Niño–Southern Oscillation. *Atmospheric Chemistry and Physics*, *24*(19), 11275–11283. Retrieved 2024-10-14, from https://acp.copernicus.org/articles/24/11275/2024/ doi: 10.5194/acp-24-11275-2024

Ribes, A., Qasmi, S., & Gillett, N. P. (2021, January). Making climate projections conditional on historical observations. *Science Advances*, *7*(4), eabc0671. Retrieved 2025-03-28, from https://www.science.org/doi/10.1126/sciadv.abc0671 doi: 10.1126/sciadv.abc0671

Riser, S. C., Freeland, H. J., Roemmich, D., Wijffels, S., Troisi, A., Belbéoch, M., ... Jayne, S. R. (2016, February). Fifteen years of ocean observations with the global Argo array. *Nature Climate Change*, *6*(2), 145–153. Retrieved 2026-01-26, from https://www.nature.com/articles/nclimate2872 doi: 10.1038/nclimate2872

Rohde, R. A., & Hausfather, Z. (2020, December). The Berkeley Earth Land/Ocean Temperature Record [Dataset]. *Earth System Science Data*, *12*(4), 3469–3479. Retrieved 2025-07-04, from https://essd.copernicus.org/articles/12/3469/2020/ doi: 10.5194/essd-12-3469-2020

Rugenstein, M. A. A., Caldeira, K., & Knutti, R. (2016). Dependence of global radiative feedbacks on evolving patterns of surface heat fluxes. *Geophysical Research Letters*, *43*(18), 9877–9885. Retrieved 2023-07-13, from https://onlinelibrary.wiley.com/doi/abs/10.1002/2016GL070907 (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2016GL070907) doi: 10.1002/2016GL070907

Samset, B., Wilcox, L., Allen, R., Stjern, C., Lund, M., Ahmadi, S., ... Westervelt, D. (2025, February). *China's aerosol cleanup has contributed strongly to the recent acceleration in global warming.* Research Square. Retrieved 2025-05-05, from https://www.researchsquare.com/article/rs-6005409/v1 (ISSN: 2693-5015) doi: 10.21203/rs.3.rs-6005409/v1

Sanderson, B. M., Pendergrass, A. G., Koven, C. D., Brient, F., Booth, B. B. B., Fisher, R. A., & Knutti, R. (2021, August). The potential for structural errors in emergent constraints. *Earth System Dynamics*, *12*(3), 899–918. Retrieved 2025-10-29, from https://esd.copernicus.org/articles/12/899/2021/ doi: 10.5194/esd-12-899-2021

Seager, R., Cane, M., Henderson, N., Lee, D.-E., Abernathey, R., & Zhang, H. (2019, July). Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. *Nature Climate Change*, *9*(7),

517–522. Retrieved 2023-05-29, from https://www.nature.com/articles/s41558-019-0505-x (Number: 7) doi: 10.1038/s41558-019-0505-x

Seeber, S., Schumacher, D. L., Gudmundsson, L., & Seneviratne, S. I. (2026, January). The observed September 2023 temperature jump was nearly impossible under standard anthropogenic forcing. *Communications Earth & Environment*. Retrieved 2026-02-11, from https://www.nature.com/articles/s43247-026-03178-8 doi: 10.1038/s43247-026-03178-8

Sherwood, S., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., ... others (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, *58*(4), e2019RG000678.

Sippel, S., Kent, E. C., Meinshausen, N., Chan, D., Kadow, C., Neukom, R., ... Knutti, R. (2024, November). Early-twentieth-century cold bias in ocean surface temperature observations. *Nature*, *635*(8039), 618–624. Retrieved 2026-02-13, from https://www.nature.com/articles/s41586-024-08230-1 doi: 10.1038/s41586-024-08230-1

Sippel, S., Meinshausen, N., Fischer, E. M., Székely, E., & Knutti, R. (2020, January). Climate change now detectable from any single day of weather at global scale. *Nature Climate Change*, *10*(1), 35–41. Retrieved 2023-06-03, from https://www.nature.com/articles/s41558-019-0666-7 (Number: 1) doi: 10.1038/s41558-019-0666-7

Sippel, S., Meinshausen, N., Székely, E., Fischer, E., Pendergrass, A. G., Lehner, F., & Knutti, R. (2021). Robust detection of forced warming in the presence of potentially large climate variability. *Science advances*, *7*(43), eabh4429.

Smith, C., Forster, P. M., Allen, M., Leach, N., Millar, R. J., Passerello, G. A., & Regayre, L. A. (2018, June). FAIR v1.3: a simple emissions-based impulse response and carbon cycle model. *Geoscientific Model Development*, *11*(6), 2273–2297. Retrieved 2026-01-22, from https://gmd.copernicus.org/articles/11/2273/2018/ doi: 10.5194/gmd-11-2273-2018

Smith, C., Hall, B., Dentener, F., Ahn, J., Collins, W., Jones, C., ... Simpson, I. (2021, August). *IPCC Working Group 1 (WG1) Sixth Assessment Report (AR6) Annex III Extended Data.* Zenodo. Retrieved 2025-11-03, from https://zenodo.org/records/5705391 doi: 10.5281/zenodo.5705391

Smith, C., Nicholls, Z., Armour, K., Collins, W., Forster, P., Meinshausen, M., ... Watanabe, M. (2021). The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity Supplementary Material. In V. Masson-Delmotte 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.* Retrieved from Availablefromhttps://www.ipcc.ch/ (Type: Book Section)

Smith, C., Walsh, T., Gillett, N., Hauser, M., Krummel, P., Lamb, W., ... Forster, P. (2025, June). *Indicators of Global Climate Change 2024.* Zenodo. Retrieved 2025-11-03, from https://zenodo.org/records/15639576 doi: 10.5281/zenodo.15639576

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., ... Ziehn, T. (2021, March). Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6. *Earth System Dynamics*, *12*(1), 253–293. Retrieved 2025-01-10, from https://esd.copernicus.org/articles/12/253/2021/ doi: 10.5194/esd-12-253-2021

Terhaar, J., Burger, F. A., Vogt, L., Frölicher, T. L., & Stocker, T. F. (2025, March). Record sea surface temperature jump in 2023–2024 unlikely but not unexpected. *Nature*, *639*(8056), 942–946. Retrieved 2025-05-06, from https://www.nature.com/articles/s41586-025-08674-z doi: 10.1038/s41586-025-08674-z

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F.,

& Knutti, R. (2020, March). Past warming trend constrains future warming in CMIP6 models. *Science Advances*, *6*(12), eaaz9549. Retrieved 2023-05-25, from https://www.science.org/doi/full/10.1126/sciadv.aaz9549 doi: 10.1126/sciadv.aaz9549

Vogt, L., de Lavergne, C., Sallée, J.-B., Kwiatkowski, L., Frölicher, T. L., & Terhaar, J. (2025, October). Increased future ocean heat uptake constrained by Antarctic sea ice extent. *Earth System Dynamics*, *16*(5), 1453–1482. Retrieved 2025-11-04, from https://esd.copernicus.org/articles/16/1453/2025/ doi: 10.5194/esd-16-1453-2025

Watanabe, M., Dufresne, J.-L., Kosaka, Y., Mauritsen, T., & Tatebe, H. (2021). Enhanced warming constrained by past trends in equatorial Pacific sea surface temperature gradient. *Nature Climate Change*, *11*(1), 33–37.

Wills, R. C., Dong, Y., Proistosecu, C., Armour, K. C., & Battisti, D. S. (2022). Systematic Climate Model Biases in the Large-Scale Patterns of Recent Sea-Surface Temperature and Sea-Level Pressure Change. *Geophysical Research Letters*, *49*(17), e2022GL100011.

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., ... Taylor, K. E. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, *47*(1), e2019GL085782.

Zhou, C., Zelinka, M. D., Dessler, A. E., & Wang, M. (2021). Greater committed warming after accounting for the pattern effect. *Nature Climate Change*, *11*(2), 132–136.