

Greater future global warming inferred from Earth's recent energy budget

Patrick T. Brown¹ & Ken Caldeira¹

Climate models provide the principal means of projecting global warming over the remainder of the twenty-first century but modelled estimates of warming vary by a factor of approximately two even under the same radiative forcing scenarios. Across-model relationships between currently observable attributes of the climate system and the simulated magnitude of future warming have the potential to inform projections. Here we show that robust across-model relationships exist between the global spatial patterns of several fundamental attributes of Earth's top-of-atmosphere energy budget and the magnitude of projected global warming. When we constrain the model projections with observations, we obtain greater means and narrower ranges of future global warming across the major radiative forcing scenarios, in general. In particular, we find that the observationally informed warming projection for the end of the twenty-first century for the steepest radiative forcing scenario is about 15 per cent warmer (+0.5 degrees Celsius) with a reduction of about a third in the two-standard-deviation spread (−1.2 degrees Celsius) relative to the raw model projections reported by the Intergovernmental Panel on Climate Change. Our results suggest that achieving any given global temperature stabilization target will require steeper greenhouse gas emissions reductions than previously calculated.

Many relevant impacts of global climate change are expected to scale with the change in global mean surface air temperature (GMSAT)^{1,2} and thus there is great scientific and societal interest in projections of future warming. The primary tools used to project GMSAT over the remainder of the twenty-first century are coupled atmosphere–ocean–land global climate models (hereafter referred to as models) but there is substantial uncertainty inherent in model projections. The majority of the uncertainty associated with the change in GMSAT over the remainder of the twenty-first century is attributable to uncertainty in future anthropogenic greenhouse gas emissions (scenario uncertainty) but there is also substantial uncertainty in the magnitude of GMSAT change for a given emissions trajectory (response uncertainty)³. Narrowing scenario uncertainty is extremely difficult because it requires increased confidence in future technological and societal conditions. Narrowing response uncertainty, however, can potentially be achieved by using observations to weigh projections towards models that best represent the true climate system.

The response uncertainty for twenty-first-century warming originates primarily from differences in how models simulate Earth's radiative energy budget and its adjustment to warming⁴. Observations of the energy budget thus have the potential to help inform projections of future warming when across-model relationships can be found between currently observable attributes of the energy budget and the simulated magnitude of future warming. Such relationships have been identified^{5–17}, but it has remained unclear how best to account for potentially compensating relationships across different components of the energy budget and disparate relationships across space (see 'Choice of predictors and predictands' in Methods and Extended Data Fig. 1). Here, we address these issues by defining our predictand to be the ultimate variable of interest (change in GMSAT from the present to various times in the future and under various radiative forcing scenarios: ΔT ; see Methods) and by using as our predictors the full global spatial distribution of fundamental components of Earth's top-of-atmosphere energy budget—its outgoing (that is, reflected) shortwave radiation (OSR), outgoing longwave radiation (OLR) and net downward energy imbalance ($\downarrow N$).

We investigate three currently observable attributes of the predictor variables—mean climatology, the magnitude of the seasonal cycle, and the magnitude of monthly variability (Methods, Extended Data Fig. 7). The combination of these three attributes and the three variables (OSR, OLR and $\downarrow N$) result in a total of nine global predictor fields. Previous studies have indicated that seasonal^{15,13} and monthly^{7,18} variability in properties of Earth's climate system can be useful as predictors of ΔT because behaviour on these timescales is related to behaviour in long-term radiative feedbacks. Climatological predictors were used because the mean state of the climate system can affect the strength of radiative feedbacks^{6,10,15,19,20}. Finally, these fields were selected as potentially useful predictors because they have across-model spreads that tend to be substantially larger than the observational uncertainty in the Clouds and Earth's Radiant Energy System (CERES)²¹ observations (Methods, Extended Data Fig. 2), meaning that it is possible to use observations to discriminate between well and poor performing models.

Statistical procedure

We investigated the relationship between the nine predictor fields and the ΔT predictands using partial least squares (PLS) regression^{22–24} (see Methods). Observationally informed ΔT predictions were compared to the raw ΔT prediction using a prediction ratio metric ($\widehat{\Delta T} / \overline{\Delta T}$), where $\widehat{\Delta T}$ is the observationally informed prediction and $\overline{\Delta T}$ is the raw model-mean prediction (Extended Data Fig. 3). Figure 1 shows prediction ratios as a function of the number of PLS components used. Observations of each of the nine energy-budget predictor fields yielded prediction ratios very close to or above 1, provided that more than about one PLS component was used (Fig. 1a). Prediction ratios are also above 1, regardless of how many PLS components are used, when all the predictor fields are used simultaneously on the eight different ΔT predictands (Fig. 1b).

As with other techniques applied in related research, the PLS procedure is capable of overfitting predictors to predictands and identifying spurious correlations (Extended Data Fig. 4a)^{9,25,26}. To guard against these issues, we evaluated the predictive skill of the predictors (and thus

¹Department of Global Ecology, Carnegie Institution for Science, Stanford, California, USA.

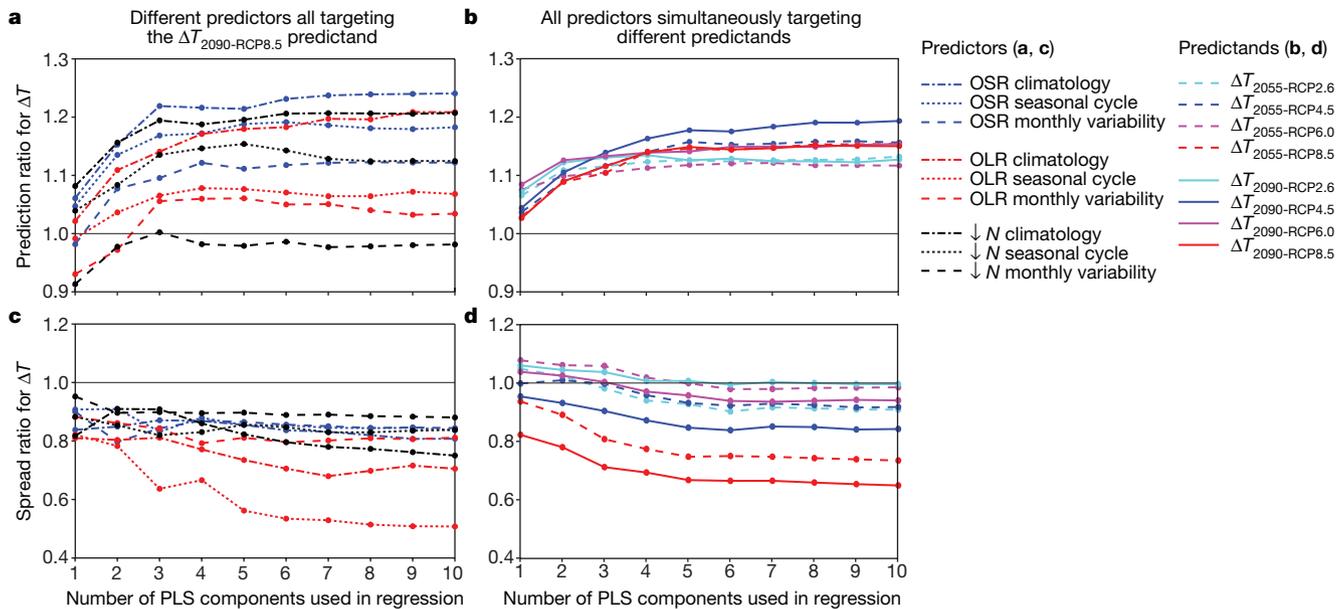


Figure 1 | Sensitivity of results to predictors or predictands used and to the number of PLS components used. **a**, Prediction ratios for the nine energy budget predictor fields, each individually targeting the $\Delta T_{2090-RCP8.5}$ predictand. **b**, As in **a** but using all nine of the energy-budget

predictor fields simultaneously while switching the predictand that is targeted. **c**, As in **a** but showing the spread ratios (equation (1)) using hold-one-out cross-validation. **d**, As in **b** but showing the spread ratios (equation (1)) using hold-one-out cross-validation.

the constrained spread) using leave-one-out cross-validation²⁷. With this method, each model in the ensemble in turn acted as a test model, with the remaining models designated as training models. PLS was performed on the training models and the resulting regression coefficients were used to predict a ΔT for the test model (Extended Data Fig. 3). In this procedure, the test model was treated as though it represented observations of the real climate system. However, unlike real observations, the ΔT for the test model was known and thus the skill of the PLS-based prediction could be evaluated. Specifically, predictive ability was assessed with the spread ratio where skill was measured relative to the root-mean-square error that would be achieved if the all-inclusive model mean $\overline{\Delta T}$ was used as the ΔT prediction for each model i of n total models:

$$\text{Spread ratio} = \sqrt{\frac{\sum_{j=1}^n (\widehat{\Delta T}_j - \Delta T_j)^2}{n}} / \sqrt{\frac{\sum_{i=1}^n (\Delta T_i - \overline{\Delta T})^2}{n}} \quad (1)$$

In equation (1), $\widehat{\Delta T}_j$ is the PLS prediction of ΔT for the j th test model (informed using only the $n - 1$ training models) and ΔT_j is the actual ΔT for the j th test model (see Methods and Extended Data Fig. 3). If a predictor field has any true skill in predicting ΔT and thus narrowing model response uncertainty, the spread ratio will be below 1.

Figure 1c shows the spread ratios, as a function of the number of PLS components used, for the nine energy-budget predictor fields targeting the $\Delta T_{2090-RCP8.5}$ predictand. (The representative concentration pathways (RCPs) are four greenhouse gas concentration trajectories adopted by the Intergovernmental Panel on Climate Change, IPCC.) We note that all of the nine energy-budget predictor fields have spread ratios below 1, indicating that each of the fields provides information that can help constrain $\Delta T_{2090-RCP8.5}$. When all predictor fields are used simultaneously, spread ratios are very near or below 1 for each of the eight predictands provided that more than about 3 PLS components are used (Fig. 1d).

In conjunction with cross-validation, we performed three additional experiments designed to expose any systematic biases in our methodology. These results indicate that the spread ratios below 1 and the ΔT prediction ratios above 1 in Fig. 1 are indeed a result of real underlying relationships between the predictors and predictands and are not an artefact of the statistical procedure (see ‘Testing for systematic bias in the statistical procedure’ in Methods, and Extended Data Fig. 4).

Overall, we find that the nine global energy-budget predictor fields tend to reduce the spread of projected ΔT (spread ratios below 1) and that observationally informed projections of ΔT are larger than the raw unconstrained model mean (prediction ratios above 1). This finding is robust to the number of PLS components used in the calculation, robust to which of the nine predictor fields are used (or if they are used simultaneously) and robust to which of the eight ΔT predictands is targeted ($\Delta T_{2055-RCP2.6}$, $\Delta T_{2055-RCP4.5}$, $\Delta T_{2055-RCP6.0}$, $\Delta T_{2055-RCP8.5}$, $\Delta T_{2090-RCP2.6}$, $\Delta T_{2090-RCP4.5}$, $\Delta T_{2090-RCP6.0}$, $\Delta T_{2090-RCP8.5}$). Both the spread ratios as well as the prediction ratios tend to asymptote after about seven PLS components are included and thus we use only seven components in the results displayed in Fig. 2 and Table 1.

Observationally informed ΔT projections

Leveraging the CERES satellite observations of the nine energy-budget predictor fields yields observationally informed projections of global warming that are higher than the unconstrained model means with roughly the same or smaller spreads for all RCP scenarios (Fig. 2, Table 1). In particular, 60%, 76%, 86% and 83% of the observationally informed ΔT distributions are greater than the raw unconstrained model mean for $\Delta T_{2090-RCP2.6}$, $\Delta T_{2090-RCP4.5}$, $\Delta T_{2090-RCP6.0}$, and $\Delta T_{2090-RCP8.5}$ respectively (Extended Data Fig. 5).

The observational constraints also affect the proportion of projections under various warming thresholds. For example, the proportion of projections that remain below 1.5°C, 2.0°C, 3.0°C and 4.0°C shifts from 44%, 21%, 64% and 38% in the raw distributions to 25%, 7%, 29% and 7% in the observationally informed distributions for RCP 2.6, RCP 4.5 RCP 6.0 and RCP 8.5, respectively (Table 1). It is also noteworthy that the observationally informed best estimate for warming by the end of the twenty-first century under the RCP 4.5 scenario is approximately the same as the raw best estimate for the RCP 6.0 scenario. This indicates that even if society were to decarbonize at a rate consistent with the RCP 4.5 pathway (which equates to cumulative CO₂ emissions about 800 gigatonnes less than that of the RCP 6.0 pathway²⁸), we should expect global temperatures to approximately follow the trajectory previously associated with RCP 6.0.

Finally, much previous research on reducing response uncertainty has made use of equilibrium climate sensitivity (ECS, the amount of warming after equilibration to a doubling of atmospheric CO₂ concentration from preindustrial values) as the predictand rather than the

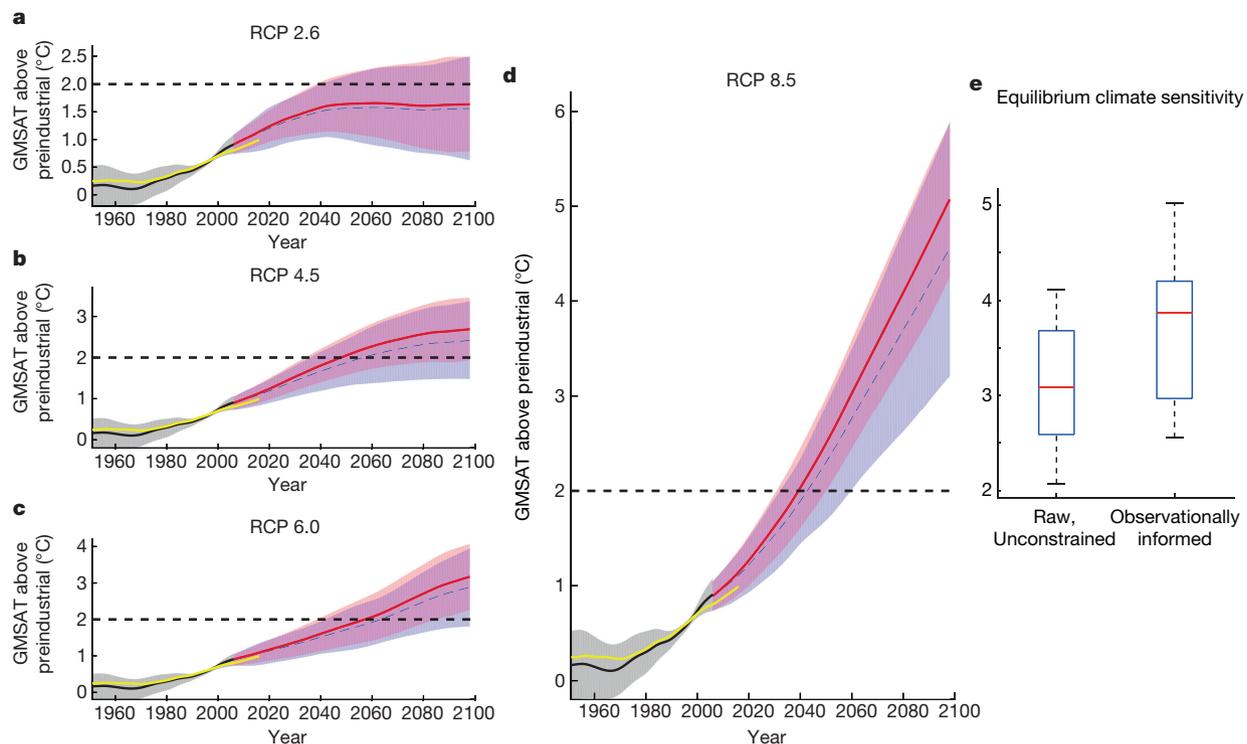


Figure 2 | Comparison of raw unconstrained and observationally informed projections of twenty-first-century global warming. **a**, Raw unconstrained (dashed blue line, blue shading) and observationally informed (solid red line, red shading) projections of twenty-first-century GMSAT based on the nine energy-budget predictor fields used simultaneously for the $\Delta T_{2090-RCP2.6}$ predictand. The yellow line

corresponds to observations from the Berkeley Earth Surface Temperature dataset⁴³. **b–d**, As in **a** but for the $\Delta T_{2090-RCP4.5}$, $\Delta T_{2090-RCP6.0}$, and $\Delta T_{2090-RCP8.5}$ predictands, respectively. **e**, Box and whisker plots for the raw and observationally informed distributions of equilibrium climate sensitivity to a doubling of CO₂. See Methods for more details. Time series are filtered⁴⁴.

transient warming magnitudes used here. Thus, in an effort to make the present work comparable to previous research, we also show the results of our procedure applied to a predictand of ECS values⁶. We find that the observationally informed ECS prediction has a mean value of 3.7°C (with a 25–75% interval of 3.0°C to 4.2°C) and that 68% of the observationally informed distribution of ECS is above the raw model mean of 3.1°C (Fig. 2e).

Physical mechanisms

Transient GMSAT response uncertainty is influenced by model differences in surface heat flux adjustments (and thus the efficiency by which energy is distributed to the deeper layers of the ocean)²⁹, as well as by model differences in the efficiency by which excess energy is radiated back to space. The latter process, which constitutes the dominant influence on GMSAT response uncertainty³⁰, is typically studied by analysing the strength of various radiative fast feedbacks. Thus, to

deduce the mechanisms through which the predictor fields influence the observationally informed ΔT projections, we perform additional PLS regressions in which we substitute the ΔT predictand with the modelled global magnitude of six components of the net radiative fast feedback: the shortwave cloud, longwave cloud, water vapour, surface albedo, lapse rate and Planck feedbacks³⁰. We find that the observationally informed prediction (using the nine predictor fields simultaneously) suggests a net feedback that is about 25% more positive than the raw model mean (increase of 0.27 W m⁻² K⁻¹, Extended Data Fig. 7d), which primarily emerges owing to a positive shift in the shortwave cloud feedback from a raw model mean of +0.02 W m⁻² K⁻¹ to an observationally informed estimate of +0.23 W m⁻² K⁻¹. The second-largest contribution to the positive shift in the net feedback comes from the surface albedo component which becomes about 14% more positive (increase of 0.05 W m⁻² K⁻¹) under the observational constraints (Extended Data Fig. 7d).

The spatial origins of the positive shift in the observationally constrained ΔT projections can be deduced by investigating the relationship between PLS loadings (see equation (10)) and CERES observations for each of the nine predictor fields (Fig. 3 and Extended Data Fig. 7c). PLS loadings denote tendencies in the predictor field that are associated with greater values in the predictand. Because Fig. 3 displays PLS loadings associated with a ΔT predictand, positive loadings denote where models with larger simulated values in their predictor field produce greater future global warming and negative loadings denote where models with smaller values in their predictor field produce greater future global warming. The extent to which observations imply larger ΔT projections can be inferred from considering how observational anomalies (represented as deviations from the model mean) project onto the PLS loading patterns (see contours in Fig. 3, Extended Data Fig. 6 and Extended Data Fig. 9).

Further insight into the mechanistic pathways by which the ΔT projections are constrained is gained by considering the similarity between

Table 1 | Updated results corresponding to tables 12.2 and 12.3 in chapter 12 of ref. 45

	RCP 2.6	RCP 4.5	RCP 6.0	RCP 8.5
Mean ΔT 2046–2065 $\pm 1\sigma$	1.7 \pm 0.3°C , 1.6 \pm 0.3°C	2.3 \pm 0.3°C , 2.0 \pm 0.3°C	2.1 \pm 0.3°C , 1.9 \pm 0.3°C	2.9 \pm 0.3°C , 2.6 \pm 0.3°C
Mean ΔT 2081–2100 $\pm 1\sigma$	1.7 \pm 0.4°C , 1.6 \pm 0.4°C	2.8 \pm 0.4°C , 2.4 \pm 0.5°C	3.2 \pm 0.4°C , 2.8 \pm 0.5°C	4.8 \pm 0.4°C , 4.3 \pm 0.7°C
$\Delta T > +1.0^\circ\text{C}$	95% , 94%	100% , 100%	100% , 100%	100% , 100%
$\Delta T > +1.5^\circ\text{C}$	75% , 56%	100% , 100%	100% , 100%	100% , 100%
$\Delta T > +2.0^\circ\text{C}$	15% , 22%	93% , 79%	100% , 100%	100% , 100%
$\Delta T > +3.0^\circ\text{C}$	0% , 0%	21% , 12%	71% , 36%	100% , 100%
$\Delta T > +4.0^\circ\text{C}$	0% , 0%	0% , 0%	0% , 0%	93% , 62%

The observationally informed projections increase the effective likelihood of warming beyond several relevant thresholds. Unboldface values correspond to the raw unconstrained model ensemble used in ref. 45 and boldface values correspond to the observationally informed projections of this study (using all nine energy budget predictor fields simultaneously). The top two rows show mean $\pm 1\sigma$ changes in global temperature and the bottom five rows show the proportion of the distribution of ΔT that falls above the given value of ΔT for the period 2081–2100.

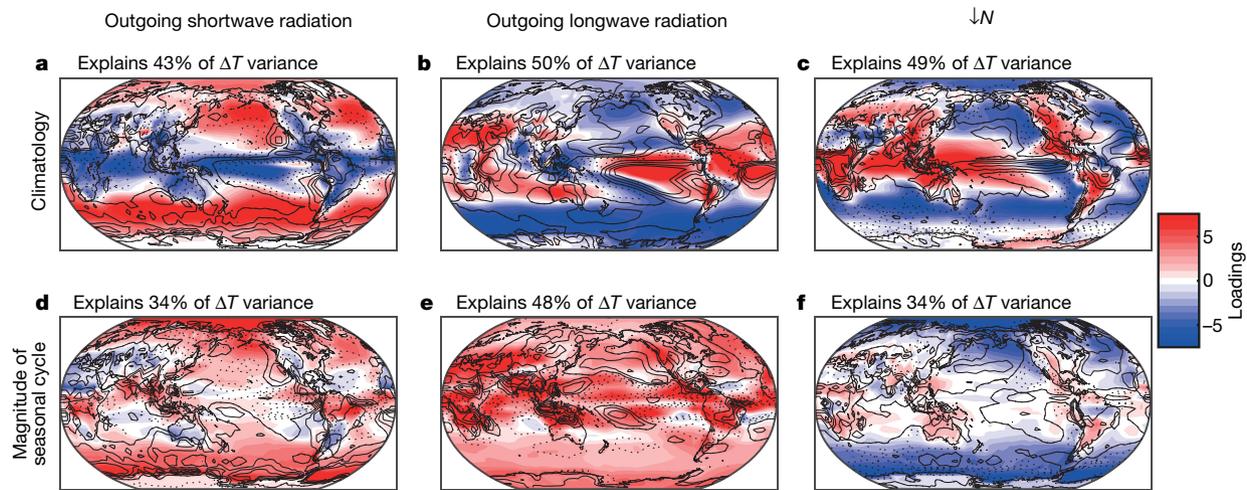


Figure 3 | Relationship between predictor fields and the magnitude of warming. The maps show PLS loadings for the first PLS component associated with the climatology (a–c) and magnitude-of-seasonal-cycle (d–f) predictor fields. Positive loadings indicate that models with larger values tend to simulate more twenty-first-century global warming and negative loadings indicate that models with smaller values tend to simulate more twenty-first-century global warming (see equation (10) in Methods).

the PLS loading patterns targeting the ΔT predictand (Fig. 3) and the PLS loading patterns targeting the magnitude of the radiative feedbacks (Extended Data Fig. 9). The PLS loading patterns associated with the predictor fields of climatology and magnitude of seasonal cycle that target ΔT (Fig. 3) are all most similar (highest pattern correlation and lowest spatial root-mean-square error) to the PLS loading patterns that target the magnitude of the shortwave cloud feedback (Extended Data Fig. 9aa–9af). The PLS loading patterns of the magnitude-of-seasonal-cycle predictor fields (Fig. 3d–f) are also similar to the PLS loading patterns targeting the magnitude of the surface albedo feedback (Extended Data Fig. 9dd–df). Since the shortwave cloud and surface albedo feedbacks seem to be the primary causes of the increased ΔT projections (Extended Data Fig. 7d), we focus on the climatology and magnitude-of-seasonal-cycle attributes of the energy budget and we discuss the magnitude-of-monthly-variability attribute in the Methods.

The climatological predictor field PLS loading patterns targeting the ΔT predictand are very similar to the analogous loading patterns targeting the magnitude of the global shortwave cloud feedback (pattern correlations of 0.92, 0.96 and 0.91 for OSR, OLR and $\downarrow N$ respectively; compare Fig. 3a–c to Extended Data Fig. 9aa–ac). Over the Southern Hemisphere's subtropical to high-latitude oceans, models that tend to simulate greater ΔT (hereafter referred to as more-sensitive models) simulate more climatological OSR (and thus more negative $\downarrow N$ despite less OLR; Fig. 4a). Larger climatological OSR in this region is associated with more-positive shortwave cloud feedback because it implies less potential for the projected local negative cloud optical depth feedback¹⁰. Similarly, larger climatological OSR over marine boundary layer regions in more-sensitive models (Fig. 3a and Extended Data Fig. 9aa) is associated with a larger net positive shortwave cloud feedback because it implies more potential for the projected local positive cloud-amount feedbacks^{31,32}.

Additionally, more-sensitive models tend to simulate less climatological OSR and less OLR (and thus larger positive $\downarrow N$) over the Indo-Pacific warm pool region (Fig. 3a–c). This supports the finding that more-sensitive models tend to have more climatological lower-tropospheric convective mixing in this region, which dehydrates the boundary layer (reducing OSR owing to reduced low-level cloud amount) and moistens the free troposphere (reducing OLR owing to increased water vapour and cloud greenhouse effects)⁶. Such models tend to have enhanced boundary layer dehydration as the climate warms and thus

The percentage of inter-model variance in $\Delta T_{2090-RCP8.5}$ explained by the first PLS component associated with each individual energy-budget field is indicated (see equation (21) in Methods). The difference between the observed CERES values and the model-mean values are contoured, with positive differences represented by solid contours and negative differences represented by dotted contours.

are characterized as having more-positive mixing-induced low cloud feedbacks⁶. The observational data project positively onto all of the above features (Fig. 3a–c and Fig. 4a), supporting the notion that observations imply a ΔT value greater than the model mean via shortwave cloud feedbacks that are more positive than the model mean.

The PLS loading patterns for the magnitude-of-seasonal-cycle predictors (Fig. 3d–f) are again most similar to the PLS loading patterns associated with the shortwave cloud feedback (Extended Data Fig. 9a) but also bear a resemblance to the loading patterns for the surface albedo feedback (spatial correlations of 0.79, 0.61 and 0.86 for OSR, OLR and $\downarrow N$ respectively; Extended Data Fig. 9dd–df). In particular, more-sensitive models have larger seasonal cycles in OSR over most of the world's oceans especially over high latitudes because more-sensitive models are characterized by more OSR during the summer months when incident solar radiation is highest (Figs 3d and 4b and d). Over the subtropical and mid-latitudes, this is mostly due to enhanced climatological cloud albedo but at high latitudes this is due to enhanced climatological surface snow and sea-ice albedo. The enhanced climatological albedo causes more-sensitive models to have smaller seasonal cycles in $\downarrow N$ at high latitudes (Figs 3f and 4b and d, Extended Data Fig. 9df). This feature is related to model sensitivity because enhanced climatological albedo provides more potential for long-term albedo feedback²⁰. Observations project positively onto these patterns, especially over the Southern Ocean (Figs 3d and f and 4b), further indicating that the larger-than-model-mean ΔT projections in Figs 1 and 2 and Table 1 are related to larger-than-model-mean shortwave cloud as well as surface albedo feedbacks.

Discussion

The constrained global warming projections presented here come with a number of important caveats. First, the unconstrained model values of ΔT do not span the complete uncertainty range. This is because there is a finite number of models, they are not comprehensive, and they do not sample the full uncertainty space of physical process representation³³. For example, a rapid nonlinear melting of the Greenland and Antarctic ice sheets³⁴ has some plausibility but is not represented in any of the models studied here and thus it has an effective probability of zero in both the raw unconstrained and observationally informed ΔT distributions. Furthermore, the models used here cannot be considered to be independent^{35,36} and thus the effective number of models

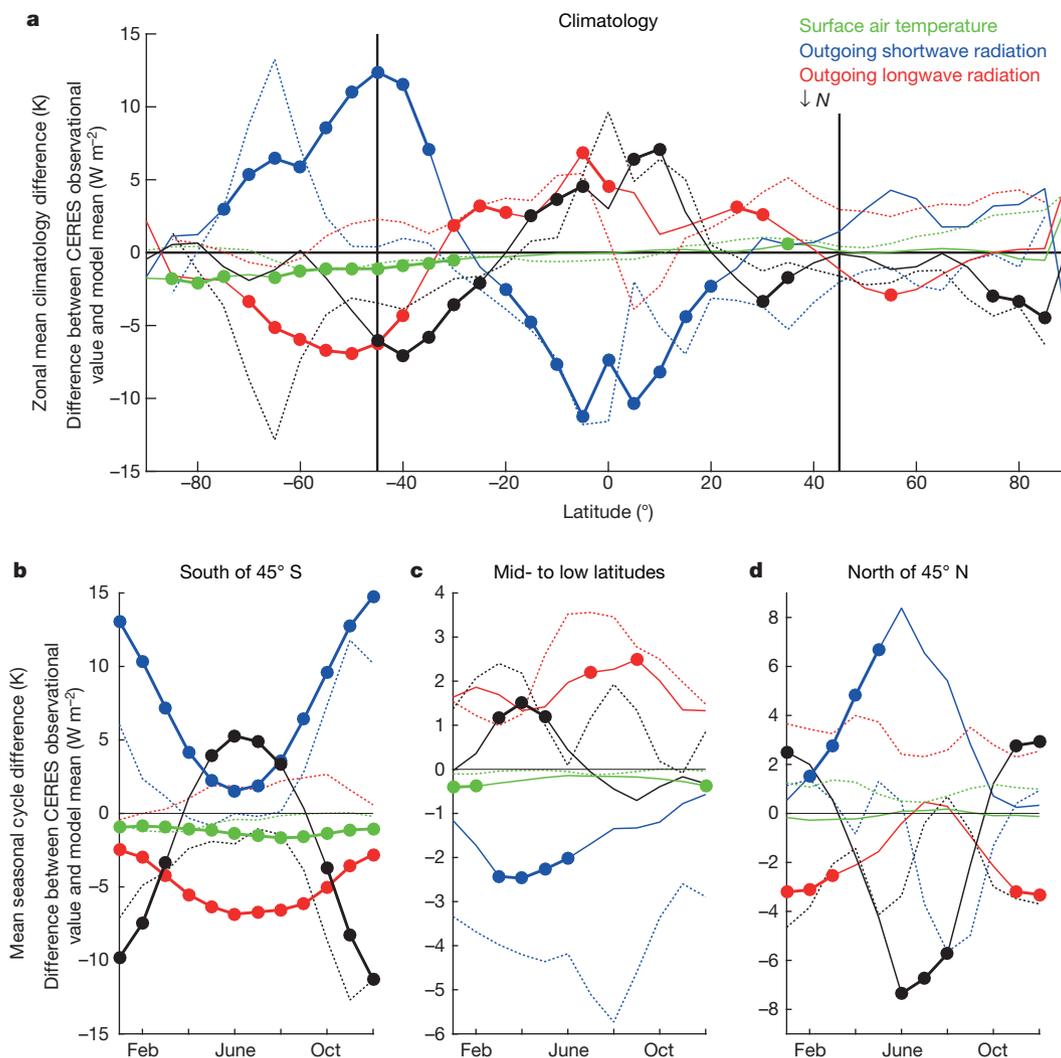


Figure 4 | Difference between more and less sensitive models in their simulation of the current climate. **a**, Zonal mean climatology difference between the more-sensitive models and less-sensitive models (solid lines) as well as the difference between the CERES observational values and the model mean (dotted lines). **b**, As in **a**, but for the mean seasonal cycle

over the Southern Hemisphere high latitudes (south of 45° S). **c**, As in **b** but over mid- to low latitudes (45° S– 45° N). **d**, As in **b** but over Northern Hemisphere high latitudes (north of 45° N). Filled circles and thicker lines indicate where the two groups of models were deemed to be statistically distinguishable (at the 90% level) based on a Student's *t*-test.

in the suite is less than the nominal number. Because of these considerations, the raw ΔT model spread is best thought of as a lower bound on total uncertainty^{30,33} and thus our observationally informed spread represents a reduction in this lower bound rather than a reduction in the upper bound.

Second, the model suite used here is diverse in terms of the level of sophistication of the simulation of, for example, atmospheric chemistry, carbon cycle processes, vegetation dynamics, and so on (Supplementary Table 1). This makes it more difficult to pinpoint the reasons for the spread in ΔT than it would be in a documented perturbed-physics-like model ensemble where only one aspect of model structure is altered at a time. Our statistical results suggest physical mechanisms (as discussed above), but these mechanisms should be interpreted as speculative rather than definitive.

Third, the CERES satellite observations used here to constrain the ΔT projections were used to some degree during the model development process³⁷. Ideally, observational data used to evaluate models would be completely independent of any data used in the development of the models. Unfortunately, owing to the limited length of high-quality observations, this is generally not possible for climate model evaluation. Thus, the model spread in the predictor fields may be artificially small owing to explicit efforts to reduce discrepancies between models and observations. Nevertheless, the model spread in the simulated

climatological energy budget components is much larger than observational uncertainty (Extended Data Fig. 2), indicating that it is possible to distinguish statistically between models that perform well and poorly.

The above caveats notwithstanding, our results indicate that observations of several diverse attributes of Earth's global energy budget indicate both individually and collectively that global warming is likely to be greater than that suggested by the unconstrained model suite. In particular, we find that the observationally informed end-of-twenty-first-century warming projection for the RCP 8.5 scenario is about 15% warmer with a reduction of about 33% in spread relative to the raw model projections. Another implication of our observationally informed projections is that the emissions associated with the RCP 4.5 scenario are likely to produce global warming more in line with that previously associated with the RCP 6.0 scenario (Table 1).

Finally, it is sometimes argued that the severity of model-projected global warming can be taken less seriously on the grounds that models fail to simulate the current climate sufficiently accurately³⁸. Our study confirms important model-observation discrepancies, indicating ample room for model improvement. However, we do not find that model errors can be taken as evidence that global warming is over-projected by climate models. On the contrary, our results add to a broadening collection of research indicating that models

that simulate today's climate best tend to be the models that project the most global warming over the remainder of the twenty-first century^{5,6,8,10–17,39–42}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 July; accepted 23 October 2017.

1. Tebaldi, C. & Arblaster, J. M. Pattern scaling: its strengths and limitations, and an update on the latest model simulations. *Clim. Change* **122**, 459–471 (2014).
2. Burke, M., Hsiang, S. M. & Miguel, E. Global non-linear effect of temperature on economic production. *Nature* **527**, 235–239 (2015).
3. Hawkins, E. & Sutton, R. The potential to narrow uncertainty in regional climate predictions. *Bull. Am. Meteorol. Soc.* **90**, 1095–1107 (2009).
4. Flato, G. *et al.* Evaluation of Climate Models. 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* (eds Stocker, T. F. *et al.*) (Cambridge Univ. Press, 2013).
5. Fasullo, J. T. & Trenberth, K. E. A less cloudy future: the role of subtropical subsidence in climate sensitivity. *Science* **338**, 792–794 (2012).
6. Sherwood, S. C., Bony, S. & Dufresne, J.-L. Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature* **505**, 37–42 (2014).
7. Dessler, A. E. A determination of the cloud feedback from climate variations over the past decade. *Science* **330**, 1523–1527 (2010).
8. Zhai, C., Jiang, J. H. & Su, H. Long-term cloud change imprinted in seasonal cloud variation: more evidence of high climate sensitivity. *Geophys. Res. Lett.* **42**, 8729–8737 (2015).
9. Klein, S. A. & Hall, A. Emergent constraints for cloud feedbacks. *Curr. Clim. Change Rep.* **1**, 276–287 (2015).
10. Trenberth, K. E. & Fasullo, J. T. Simulation of present-day and twenty-first-century energy budgets of the Southern Oceans. *J. Clim.* **23**, 440–454 (2010).
11. Su, H. *et al.* Weakening and strengthening structures in the Hadley Circulation change under global warming and implications for cloud response and climate sensitivity. *J. Geophys. Res.* **D 119**, 5787–5805 (2014).
12. Gordon, N. D. & Klein, S. A. Low-cloud optical depth feedback in climate models. *J. Geophys. Res.* **D 119**, 6052–6065 (2014).
13. Hall, A. & Qu, X. Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophys. Res. Lett.* **33**, (2006).
14. Qu, X., Hall, A., Klein, S. A. & Caldwell, P. M. On the spread of changes in marine low cloud cover in climate model simulations of the 21st century. *Clim. Dyn.* **42**, 2603–2626 (2014).
15. Volodin, E. M. Relation between temperature sensitivity to doubled carbon dioxide and the distribution of clouds in current climate models. *Izv., Atmos. Ocean. Phys.* **44**, 288–299 (2008).
16. Huber, M., Mahlstein, I., Wild, M., Fasullo, J. & Knutti, R. Constraints on climate sensitivity from radiation patterns in climate models. *J. Clim.* **24**, 1034–1052 (2011).
17. Tan, I., Storelvmo, T. & Zelinka, M. D. Observational constraints on mixed-phase clouds imply higher climate sensitivity. *Science* **352**, 224–227 (2016).
18. Dessler, A. E., Zhang, Z. & Yang, P. Water-vapor climate feedback inferred from climate fluctuations, 2003–2008. *Geophys. Res. Lett.* **35**, L20704 (2008).
19. Gregory, J. M., Andrews, T. & Good, P. The inconstancy of the transient climate response parameter under increasing CO₂. *Phil. Trans. R. Soc. Lond. A* **373**, <https://dx.doi.org/10.1098/rsta.2014.0417> (2015).
20. Hu, X. *et al.* Inter-model warming projection spread: inherited traits from control climate diversity. *Sci. Rep.* **7**, 4300 (2017).
21. Wielicki, B. A. *et al.* Clouds and the Earth's Radiant Energy System (CERES): an Earth observing system experiment. *Bull. Am. Meteorol. Soc.* **77**, 853–868 (1996).
22. de Jong, S. SIMPLS: an alternative approach to partial least squares regression. *Chemom. Intell. Lab. Syst.* **18**, 251–263 (1993).
23. Wold, H. in *Multivariate Analysis* 391–420 (Academic Press, 1966).
24. Smoliak, B. V., Wallace, J. M., Stoelinga, M. T. & Mitchell, T. P. Application of partial least squares regression to the diagnosis of year-to-year variations in Pacific Northwest snowpack and Atlantic hurricanes. *Geophys. Res. Lett.* **37**, (2010).
25. DelSole, T. & Shukla, J. Artificial skill due to predictor screening. *J. Clim.* **22**, 331–345 (2009).
26. Caldwell, P. M. *et al.* Statistical significance of climate sensitivity predictors obtained by data mining. *Geophys. Res. Lett.* **41**, 1803–1808 (2014).
27. Efron, B. & Gong, G. A leisurely look at the bootstrap, the jackknife, and cross-validation. *Am. Stat.* **37**, 36–48 (1983).
28. IPCC. Summary for Policymakers. 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* (eds Stocker, T. F. *et al.*) (Cambridge Univ. Press, 2013).
29. Baker, M. B. & Roe, G. H. The shape of things to come: why is climate change so predictable? *J. Clim.* **22**, 4574–4589 (2009).
30. Caldwell, P. M., Zelinka, M. D., Taylor, K. E. & Marvel, K. Quantifying the sources of inter-model spread in equilibrium climate sensitivity. *J. Clim.* **29**, 513–525 (2016).
31. Grise, K. M., Polvani, L. M. & Fasullo, J. T. Reexamining the relationship between climate sensitivity and the Southern Hemisphere radiation budget in CMIP models. *J. Clim.* **28**, 9298–9312 (2015).
32. Ceppi, P., Brient, F., Zelinka, M. D. & Hartmann, D. L. Cloud feedback mechanisms and their representation in global climate models. *Wiley Interdiscip. Rev. Clim. Change* **8**, e465 (2017).
33. Rowlands, D. J. *et al.* Broad range of 2050 warming from an observationally constrained large climate model ensemble. *Nat. Geosci.* **5**, 256–260 (2012).
34. Hansen, J. *et al.* Ice melt, sea level rise and superstorms: evidence from paleoclimate data, climate modeling, and modern observations that 2 °C global warming could be dangerous. *Atmos. Chem. Phys.* **16**, 3761–3812 (2016).
35. Tebaldi, C. & Knutti, R. The use of the multi-model ensemble in probabilistic climate projections. *Phil. Trans. R. Soc. A* **365**, 2053–2075 (2007).
36. Knutti, R., Masson, D. & Gettelman, A. Climate model genealogy: generation CMIP5 and how we got there. *Geophys. Res. Lett.* **40**, 1194–1199 (2013).
37. Schmidt, G. A. *et al.* Practice and philosophy of climate model tuning across six U.S. modeling centers. *Geosci. Model Dev. Discuss.* **2017**, 1–24 (2017).
38. Curry, J. A. *Climate Models for the Layman* (The Global Warming Policy Foundation, 2017).
39. Brient, F. & Schneider, T. Constraints on climate sensitivity from space-based measurements of low-cloud reflection. *J. Clim.* **29**, 5821–5835 (2016).
40. Myers, T. A. & Norris, J. R. Reducing the uncertainty in subtropical cloud feedback. *Geophys. Res. Lett.* **43**, 2144–2148 (2016).
41. Tian, B. Spread of model climate sensitivity linked to double-Intertropical Convergence Zone bias. *Geophys. Res. Lett.* **42**, 4133–4141 (2015).
42. Zhou, C., Zelinka, M. D., Dessler, A. E. & Klein, S. A. The relationship between interannual and long-term cloud feedbacks. *Geophys. Res. Lett.* **42**, 10463–10469 (2015).
43. Rohde, R. A. *et al.* A new estimate of the average Earth surface land temperature spanning 1753 to 2011. *Geoinform. Geostat.* **1**, 1–7 (2013).
44. Cleveland, W. S. Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.* **74**, 829–836 (1979).
45. Collins, M. *et al.* Long-term climate change: projections, commitments and irreversibility. 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* (eds Stocker, T. F. *et al.*) (Cambridge Univ. Press, 2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Z. Hausfather for discussions. This study was supported by the Fund for Innovative Climate and Energy Research and the Carnegie Institution for Science endowment. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for the Coupled Modelled Intercomparison Project (CMIP), and we thank the climate modelling groups for producing and making available their model output. For CMIP the US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

Author Contributions K.C. conceived the study. P.T.B. performed the analysis and wrote an initial draft of the manuscript. Both authors contributed to interpretation of results and refinement of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to P.T.B. (pbrown@carnegiescience.edu).

Reviewer Information *Nature* thanks T. L'Ecuyer and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Choice of predictors and predictands. There has been a recent emphasis in the literature on the practical goal of using currently observable aspects of the climate system (predictors) to constrain simulated long-term responses (predictands)^{9,13}. Emergent constraints are systematic across-model relationships between predictors and predictands that arise spontaneously owing to inherent differences in model structure (for example, choices made on how best to parameterize sub-grid scale processes). These across-model relationships can potentially be leveraged, in conjunction with observations of the predictors, to constrain the predictand. Emergent constraints for long-term GMSAT change are particularly compelling if the predictor has a straightforward causal relationship to GMSAT response and if there are reasonable grounds to believe that observable short-term variations are physically analogous to the changes relevant to long-term GMSAT change. Predictors associated with fast feedbacks (clouds, surface snow/ice albedo, and atmospheric water vapour/lapse rate changes) fit this description because they respond to temperature change on timescales from hours to years and because they exert a large impact on Earth's energy budget and thus GMSAT⁵⁻⁹.

However, an apparent emergent constraint on a fast feedback may not actually constrain GMSAT response owing to compensation in space or compensation between differing feedbacks. For example, some regions are characterized by strong across-model relationships between the magnitude of the seasonal cycle in the shortwave cloud radiative effect (CRE-SW)⁴⁵ and the magnitude of change in CRE-SW over the remainder of the twenty-first century (Extended Data Fig. 1a). Such relationships suggest that the emergent model spread in the magnitude of seasonal CRE-SW variability projects onto model spread in the magnitude of long-term CRE-SW change, which is the dominant contributor to model GMSAT response uncertainty³⁰. Since the predictor (short-term CRE-SW variability) and predictand (long-term CRE-SW response) both represent the behaviour of the same variable, drawing a physical connection between the two is relatively straightforward. This advantage in ease of physical interpretation, however, comes at the cost of being able to constrain the ultimate variable of interest (GMSAT change, Extended Data Fig. 1b). Thus, if the final goal is to constrain GMSAT change, it is most practical to focus on predictors that have a strong direct relationship with GMSAT change even if this makes drawing a physical connection less straightforward.

Another complicating factor in the identification of potential emergent constraints is that the strength and even the sign of the predictor–predictand relationship may depend on location (compare Extended Data Fig. 1c and d). This is partly because the magnitude and influence of fast feedbacks are spatially dependent, so feedbacks at different locations can exert differing leverage on GMSAT^{47,48}. Additionally, the climate system is dynamically linked through horizontal energy transport, so modelled feedback strength at a given location inevitably influences modelled feedback strength remotely⁴⁹⁻⁵¹. Therefore, any method intent on constraining GMSAT change should be able to account holistically for relationships that vary over the entire spatial extent of Earth.

The above two considerations inform our choice to use global, aggregate components of Earth's energy budget as our predictor fields and to use change in GMSAT as our predictand. These choices allow us to avoid the evocation of *ceteris paribus* assumptions that tend to be implicit in studies that focus on a subset of feedbacks or a particular geographic region.

Predictand definitions used. We define ΔT predictands to be the modelled change in GMSAT between the period 1986–2005 and various periods in the future under different emissions scenarios. Predictands also have 0.61 °C added to them so that they represent change with respect to the preindustrial period⁴⁵. Eight sets of ΔT predictands were used in this study, corresponding to two ' T_{final} ' time periods (the mean over the years 2046–2065 and the mean over the years 2081–2100) and four Representative Concentration Pathways⁵² (RCP 2.6, RCP 4.5, RCP 6.0 and RCP 8.5): $\Delta T_{2055\text{-RCP}2.6}$, $\Delta T_{2055\text{-RCP}4.5}$, $\Delta T_{2055\text{-RCP}6.0}$, $\Delta T_{2055\text{-RCP}8.5}$, $\Delta T_{2090\text{-RCP}2.6}$, $\Delta T_{2090\text{-RCP}4.5}$, $\Delta T_{2090\text{-RCP}6.0}$, $\Delta T_{2090\text{-RCP}8.5}$.

Preprocessing and predictor fields. All model and observed data were bilinearly interpolated onto the same spatial grid (37 × 72) for ease of intercomparison. This grid was coarser than the observational and modelled native grids so as to emphasize spatially broad features. Predictor fields consisted of aggregate aspects of Earth's energy budget: the full global spatial patterns of the climatology and climatological variability of OLR, OSR and net downward radiation ($\downarrow N$). The climatological predictor field is simply the local time-mean of these variables from the period 2001–2015. The magnitude of the seasonal cycle predictor field is the local standard deviation (σ) of the mean seasonal cycle. The magnitude of the monthly variability predictor field is the local σ of the monthly time series with the mean seasonal cycle removed.

Normalization of predictor fields under simultaneous PLS regression. The nine energy-budget predictor fields have different across-model variances, partly

owing to differences in the fundamental nature of each variable itself. Thus, when forming a predictor matrix that simultaneously includes all the predictor fields, it is necessary to normalize the variance of the predictor fields such that each field has the same across-model variance. This is done by subtracting the global model mean and dividing by the global model-mean σ for each predictor field separately before forming the simultaneous predictor matrix $[X]$.

Use of $\downarrow N$ in addition to OSR and OLR. For the simultaneous predictor field (using all nine energy-budget predictors at once) it may not be clear why the inclusion of $\downarrow N$ provides additional information over OSR and OLR since $\downarrow N = \text{ISR} - \text{OSR} - \text{OLR}$ (where ISR is incoming shortwave/solar radiation). One reason is that ISR is not necessarily the same for all models so that variation is accounted for in $\downarrow N$. More importantly, however, $\downarrow N$ provides information on how OSR and OLR covary over the course of the seasonal cycle and in their month-to-month variability. This becomes apparent when one considers that highly correlated OSR and OLR variability will be associated with large-magnitude variability in $\downarrow N$ but highly anti-correlated OSR and OLR variability will be associated with low-magnitude variability in $\downarrow N$. Thus, the variability of $\downarrow N$ is not strictly implied from the variability in OSR and OLR and so $\downarrow N$ variability provides additional information as a predictor.

CERES satellite observations. For the observationally informed projections, we used the CERES Energy Balanced and Filled (EBAF4.0) product²¹. We use the CERES product exclusively because it has lower observational uncertainties compared to those associated with atmospheric reanalysis products or previous satellite-based products⁵³. In particular, CERES has observational uncertainties that are approximately 2–3 times smaller than the Earth Radiation Budget Experiment (ERBE)²¹. Atmospheric reanalysis products have errors that are sufficiently large to preclude their use as a standard with which to compare climate models. For example, biases of up to about 15 W m⁻² in global mean $\downarrow N$ are a feature of some atmospheric reanalyses; see ref. 54 for further discussion.

Models used. We used atmosphere–ocean–land global climate models that participated in Phase 5 of the Coupled Modelled Intercomparison Project (CMIP5)⁵⁵ (Supplementary Table 1). We used only a single realization for each model so that independence across the ensemble was maximized. Furthermore, since the CERES observations inevitably contain an element of unforced internal variability, it would be inappropriate to take averages of multiple ensemble members to the predictand relative to observations. For RCP 2.6 there were 24 models available, for RCP 4.5 there were 35 models available, for RCP 6.0 there were 18 models available and for RCP 8.5 there were 37 models available. The analysis of equilibrium climate sensitivity and feedback magnitude was done on 24 and 22 models respectively. Supplementary Table 1 provides model specifications and documents which models were available for each experiment.

PLS regression. PLS regression is applicable to partial correlation problems analogously to the more widely used multiple linear regression. In multiple linear regression, coefficients \mathbf{b} are found such that the mean squared residuals \mathbf{r} are minimized in the system:

$$\mathbf{y} = [\mathbf{X}]\mathbf{b} + \mathbf{r} \quad (2)$$

For the present application, \mathbf{y} is a vector that contains predictands (ΔT for each model at a given time and under a given emissions scenarios) and the matrix $[\mathbf{X}]$ contains the global spatial field of the observable predictor variable (where the rows correspond to models and the columns correspond to locations). Because of the high degree of spatial autocorrelation in the predictor fields used here, the columns in $[\mathbf{X}]$ will be highly collinear, and thus $[\mathbf{X}]$ will be well below full rank. This precludes the application of multiple linear regression to the problem. However, PLS offers a solution to this problem by creating linear combinations of the columns in $[\mathbf{X}]$ (PLS components) that represent a large portion of the variability of $[\mathbf{X}]$. The procedure is similar to principal component analysis (often used in climate science) but instead of seeking components that explain the maximum variability in $[\mathbf{X}]$ itself, PLS seeks components in $[\mathbf{X}]$ that explain the covariability between $[\mathbf{X}]$ and \mathbf{y} . Ultimately, PLS is akin to the multiple linear regression procedure performed on a matrix $[\mathbf{Z}]$ containing a relatively low number of PLS components that represent most of the variability in $[\mathbf{X}]$:

$$\mathbf{y} = [\mathbf{Z}]\boldsymbol{\beta} + \mathbf{r} \quad (3)$$

PLS can also be used on the nine predictor fields simultaneously (where the columns in $[\mathbf{X}]$ span every predictor field as well as every global location) to account for redundancies across the predictors.

We carry out PLS regression using the MATLAB function 'plsregress' (<https://www.mathworks.com/help/stats/plsregress.html>). This function performs PLS regression using the SIMPLS algorithm²². A summary of the SIMPLS procedure is as follows:

(1) Both the predictor matrix $[X]$ and the predictand vector \mathbf{y} are centred (that is, they have their column means subtracted) to obtain $[X]_0$ and \mathbf{y}_0 . Note that in our application, $[X]$ is an $n \times p$ matrix and \mathbf{y} is a $n \times 1$ vector where n is the number of models and p is the number of global gridded locations (or the number of global gridded locations multiplied by the number of predictor variables used if multiple predictor fields are used simultaneously).

(2) The cross-product vector is calculated as

$$\mathbf{s} = [X]_0^T \mathbf{y}_0 \quad (4)$$

(3) Singular value decomposition of \mathbf{s} is performed to obtain the dominant eigen-vector of $\mathbf{s}\mathbf{s}^T$, which is referred to as the \mathbf{y} -block-factor²² weight \mathbf{q} .

(4) The $[X]$ -block-factor weights are calculated as

$$\mathbf{r} = \mathbf{s}\mathbf{q} \quad (5)$$

(5) The $[X]$ -block-factor scores are calculated as

$$\mathbf{t} = [X]\mathbf{q} \quad (6)$$

(6) The $[X]$ -block-factor scores are centred and normalized as

$$\mathbf{t} = \mathbf{t} - \bar{\mathbf{t}} \quad (7)$$

$$\mathbf{t} = \frac{\mathbf{t}}{\sqrt{\mathbf{t}^T \mathbf{t}}} \quad (8)$$

(7) The $[X]$ -block-factor weights are altered appropriately as

$$\mathbf{r} = \frac{\mathbf{r}}{\sqrt{\mathbf{t}^T \mathbf{t}}} \quad (9)$$

(8) The $[X]$ -block-factor loadings are computed as

$$[P] = [X]^T \mathbf{t} \quad (10)$$

Note that $[P]$ is what is shown in Fig. 3 and Extended Data Figs 6, 9 and 10a–c. $[P]$ is a quantification of how much the original predictor field at each location relates to the PLS factor \mathbf{t} .

(9) The \mathbf{Y} -block-factor loadings are computed as

$$\mathbf{q} = \mathbf{y}_0^T \mathbf{t} \quad (11)$$

(10) The \mathbf{Y} -block-factor scores are computed as

$$\mathbf{u} = \mathbf{y}_0 \mathbf{q} \quad (12)$$

(11) If the current PLS component is the second PLS component (or greater), the X -block-factor loadings \mathbf{v} are made to be orthogonal to previous loadings \mathbf{p} and the factor scores \mathbf{u} are made to be orthogonal to previous scores as

$$\mathbf{v} = \mathbf{p} \quad (13)$$

$$\mathbf{v} = \mathbf{v} - [V]([V]^T \mathbf{p}) \quad (14)$$

$$\mathbf{u} = \mathbf{u} - [T]([T]^T \mathbf{u}) \quad (15)$$

(12) The loadings are normalized

$$\mathbf{v} = \frac{\mathbf{v}}{\sqrt{\mathbf{v}^T \mathbf{v}}} \quad (16)$$

(13) The cross-product vector is deflated

$$\mathbf{s} = \mathbf{s} - \mathbf{v}(\mathbf{v}^T \mathbf{s}) \quad (17)$$

(14) The vectors associated with the calculation of the current PLS component (\mathbf{r} , \mathbf{t} , \mathbf{p} , \mathbf{q} , \mathbf{u} and \mathbf{v}) are stored in their appropriate locations in $[R]$, $[T]$, $[P]$, \mathbf{Q} , $[U]$ and $[V]$, respectively.

(15) The PLS regression coefficients matrix is calculated as

$$[B] = [R]\mathbf{Q}^T \quad (18)$$

(16) If desired, a prediction $\hat{\mathbf{y}}_0$ is made based on an observed or held-one-out row vector \mathbf{x}_0^*

$$\hat{\mathbf{y}}_0 = \mathbf{x}_0^* [B] \quad (19)$$

To make a prediction for a non-centred \mathbf{y} value, from a non-centred \mathbf{x}^* vector, the original units must be reintroduced by adding an intercept term in the first column of $[B]$ and adding 1 to the (1,1) location of \mathbf{x}^* .

(17) Steps (3)–(16) are repeated until the number of desired PLS components is reached.

Observationally informed projections. The spread ratios and ΔT predictions ratios (Fig. 1) were calculated by first obtaining a central estimate of the observationally informed ΔT prediction by using the predictor field(s) associated with the CERES satellite data as \mathbf{x}_0^* and making a prediction using equation (19). The spread about the observationally informed prediction comes from hold-one-out cross-validation (where equation (19) is applied to each model as it plays the role of ‘test model’). See Extended Data Fig. 3 and Supplementary Video 1.

Fourfold cross-validation. Similar to the hold-one-out cross-validation described in the main text, we also test the robustness of our statistical approach using fourfold cross-validation (Extended Data Fig. 4b). In this method, models are randomly separated into a training group that contains approximately 75% of the models and a test group that contains approximately 25% of the models. PLS is performed on the training group only, and the resulting regression coefficients are used to predict the ΔT values for each model j in the test group of m total models. Equation (1) then becomes:

$$\frac{\sqrt{\sum_{j=1}^m (\widehat{\Delta T}_j - \Delta T_j)^2}}{m} \bigg/ \frac{\sqrt{\sum_{i=1}^n (\Delta T_i - \bar{\Delta T})^2}}{n} \quad (20)$$

In equation (20), $\widehat{\Delta T}_j$ is the PLS prediction of ΔT for the j th test model (informed using only the $n - m$ training models) and ΔT_j is the actual ΔT for the j th test model. We perform fourfold cross-validation 1,000 times with random assignment of models to the training and test group and the reported spread ratios are averages over these 1,000 trials. Fourfold cross-validation is a more rigorous test of robustness than hold-one-out cross-validation because a higher proportion of models are held out as test models and fewer models are used as training models. Spread ratios are generally below 1 even under fourfold cross-validation (Extended Data Fig. 4b), providing further evidence that these nine predictor fields provide useful information that can constrain GMSAT response.

No cross-validation. When cross-validation is not used (that is, when the model test group and training group are identical), spread ratios approach zero as the number of PLS components is increased (Extended Data Fig. 4a). This result shows that PLS is able to find coefficients that can fit $[X]$ to \mathbf{y} arbitrarily well, provided enough PLS components are used and thus it demonstrates the necessity of cross-validation.

Testing for systematic bias in the statistical procedure. In conjunction with cross-validation, we perform three additional experiments designed to expose any systematic biases in our methodology. These three experiments involve supplying the statistical procedure with data that should not produce any constraint on the magnitude of ΔT or the size of the ΔT spread. In one experiment, we substituted the described energy-budget predictor fields with global surface air temperature annual anomalies for each model. Since annual surface air temperature anomaly fields are dominated by chaotic unforced variability, the across-model relationship of these patterns for any given year is unlikely to be related to the across-model relationship in ΔT . In this experiment we use Berkeley Earth Surface Temperature⁴³ data as the observations. We ran this experiment on annual anomalies for every year between 1861 and 2005, each time targeting the $\Delta T_{2090\text{-RCP8.5}}$ predictand. In a second experiment, we substituted the original $\Delta T_{2090\text{-RCP8.5}}$ predictand vector with versions of the vector that have had its values randomly scrambled. Thus, these scrambled $\Delta T_{2090\text{-RCP8.5}}$ predictands have the same statistical moments as the original but any across-model relationship between predictors and predictands should be eliminated on average. Finally, in a third experiment, we used both the surface air temperature anomaly fields and the scrambled $\Delta T_{2090\text{-RCP8.5}}$ vectors as the predictors and predictands, respectively.

The results of the three experiments are shown in Extended Data Fig. 4c and d. In contrast to results between the energy-budget predictor fields and the various ΔT predictands (Fig. 1), the three experiments all result in spread ratios above 1 and ΔT prediction ratios centred on 1. These results indicate that the spread ratios below 1 and the ΔT prediction ratios above 1 in Fig. 1 are a result of real underlying relationships between the predictors and predictands and are not an artefact of the statistical procedure itself.

Production of Fig. 2. The time series in Fig. 2 are filtered with a 20-year LOWESS (locally weighted scatterplot smoothing) function⁴⁴ to suppress model spread due to unforced variability. The shaded area represents the 2σ model spread and by definition has zero width at the centre of the baseline time period (1986–2005). The constrained spread is linearly extrapolated to grow to its appropriate constrained ΔT_{2090} size by 2090 and is constant thereafter. The black line and grey shading correspond to the historical experiment and the yellow line corresponds to observed GMSAT from the Berkeley Earth Surface Temperature dataset⁴³. In Fig. 2e, the whiskers span the entire distribution range;

the blue box spans the 25th to the 75th percentiles and the red line indicates the median.

Feedback magnitude predictands. In addition to targeting the various ΔT predictands, we also target the magnitude of six radiative fast feedbacks: the short-wave cloud, longwave cloud, water vapour, surface albedo, lapse rate and Planck feedbacks³⁰, as well as the net feedback (Extended Data Figs 7d and 8). These feedback strengths were calculated from instantaneous $4 \times \text{CO}_2$ experiments in ref. 30. We use the standard feedback magnitudes presented in table 1 of ref. 30, not those calculated relative to constant relative humidity.

PLS loadings associated with the magnitude of monthly variability. The first PLS loading patterns of the magnitude of monthly variability (Extended Data Fig. 10) indicate that more sensitive models are characterized by more tropical and subtropical monthly variability in the OSR, OLR and $\downarrow N$. Observations, however, project negatively onto these patterns, resulting in ΔT prediction ratios below 1 for these PLS components (leftmost points of dashed lines in Fig. 1a). Since the first PLS component by itself suggests a ΔT ratio below 1, but this is unrepresentative of the eventual ΔT ratio above 1 that is found when more PLS components are used (for OSR and OLR, Fig. 1a), we focus here on the PLS loading patterns associated with the second PLS component (Extended Data Fig. 6ag–ai). The second PLS component loading patterns are most similar to the PLS loading patterns associated with the magnitude of the global water vapour feedback (in terms of root-mean-square error, RMSE; see Extended Data Fig. 9cg–ci). In particular, a larger water vapour feedback is associated with less monthly variability in OSR and OLR in the tropics (Extended Data Fig. 9cg and 9ch).

In the tropics, unforced monthly fluctuations in surface temperature are associated with decreases in OSR and increases in OLR (Extended Data Fig. 10e). An enhanced water vapour feedback would damp the decrease in OSR by maintaining sufficient water vapour for the presence of clouds and would damp the increase in OLR by increasing the greenhouse effect of the atmosphere as surface temperatures increase⁵⁶. Thus, observations of monthly variability in OLR and OSR imply a water vapour feedback that is greater than the model mean. This, however, is not sufficient to imply that the model-mean water vapour feedback is above the raw model mean (Extended Data Fig. 7d).

Explained variance between the PLS component and predictand. The explained variance values in Fig. 3 and Extended Data Figs 6 and 10 quantify how well variation

in the given PLS component explains variation in the predictand vector. These explained variance values are calculated as:

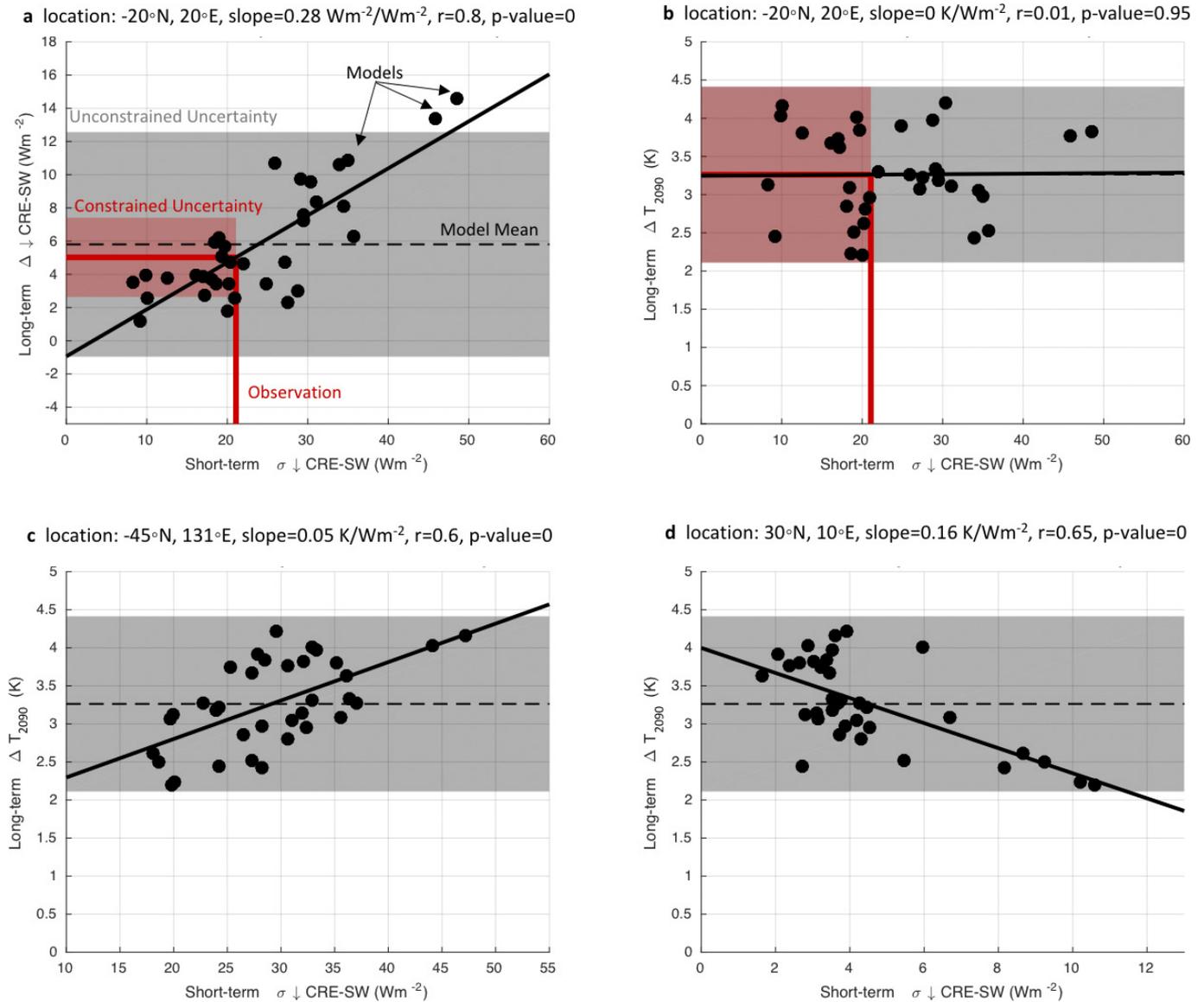
$$\frac{\sum |\mathbf{q}|^2}{\sum |\mathbf{y}_0|^2} \quad (21)$$

where \mathbf{y}_0 is the centred predictand vector and \mathbf{q} is the Y -block-factor weighting vector (equation (11)).

Code availability. The primary code used to make the observationally informed projections is available in the Supplementary Information.

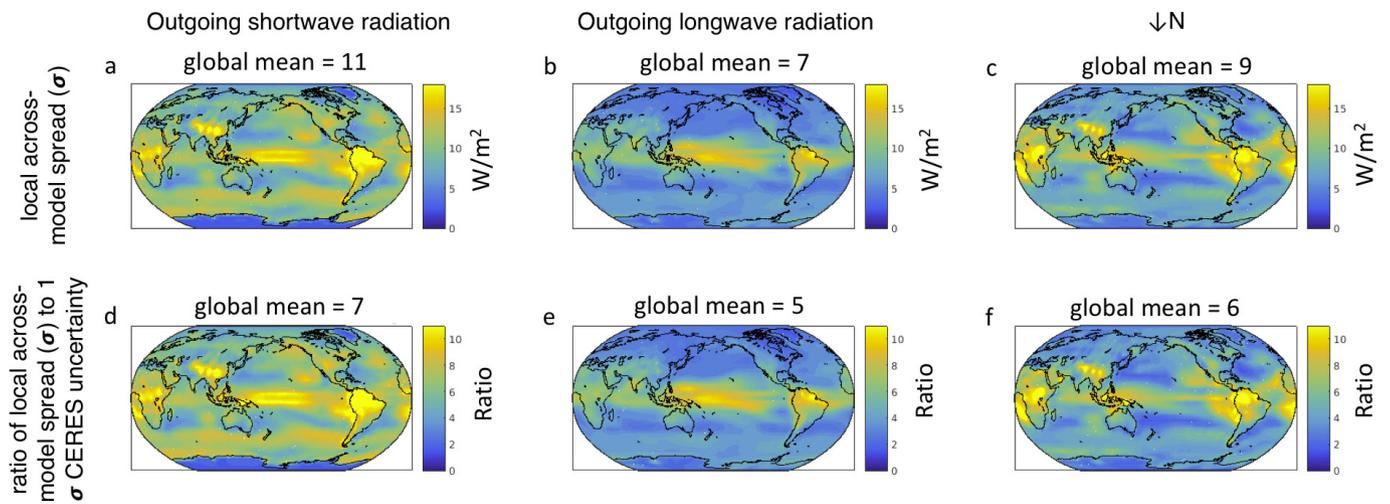
Data availability. The CMIP5 data used for this study can be accessed at <http://pcmdi9.llnl.gov/>. The CERES satellite observations can be accessed at https://ceres.larc.nasa.gov/order_data.php. Other data and material requests are available from the corresponding author.

46. Ramanathan, V. *et al.* Cloud-radiative forcing and climate: results from the Earth Radiation Budget experiment. *Science* **243**, 57–63 (1989).
47. Armour, K. C., Bitz, C. M. & Roe, G. H. Time-varying climate sensitivity from regional feedbacks. *J. Clim.* **26**, 4518–4534 (2013).
48. Rugenstein, M. A. A., Caldeira, K. & Knutti, R. Dependence of global radiative feedbacks on evolving patterns of surface heat fluxes. *Geophys. Res. Lett.* **43**, 9877–9885 (2016).
49. Roe, G. H., Feldl, N., Armour, K. C., Hwang, Y.-T. & Frierson, D. M. W. The remote impacts of climate feedbacks on regional climate predictability. *Nat. Geosci.* **8**, 135–139 (2015).
50. Feldl, N. & Roe, G. H. The nonlinear and nonlocal nature of climate feedbacks. *J. Clim.* **26**, 8289–8304 (2013).
51. Hwang, Y.-T. & Frierson, D. M. W. Increasing atmospheric poleward energy transport with global warming. *Geophys. Res. Lett.* **37**, L24807 (2010).
52. Meinshausen, M. *et al.* The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim. Change* **109**, 213–241 (2011).
53. Trenberth, K. E., Fasullo, J. T. & Kiehl, J. Earth's Global Energy Budget. *Bull. Am. Meteorol. Soc.* **90**, 311–323 (2009).
54. Trenberth, K. E. & Fasullo, J. T. Regional energy and water cycles: transports from ocean to land. *J. Clim.* **26**, 7837–7851 (2013).
55. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
56. Trenberth, K. E., Zhang, Y., Fasullo, J. T. & Taguchi, S. Climate variability and relationships between top-of-atmosphere radiation and temperatures on Earth. *J. Geophys. Res. Atmos.* **120**, 3642–3659 (2015).



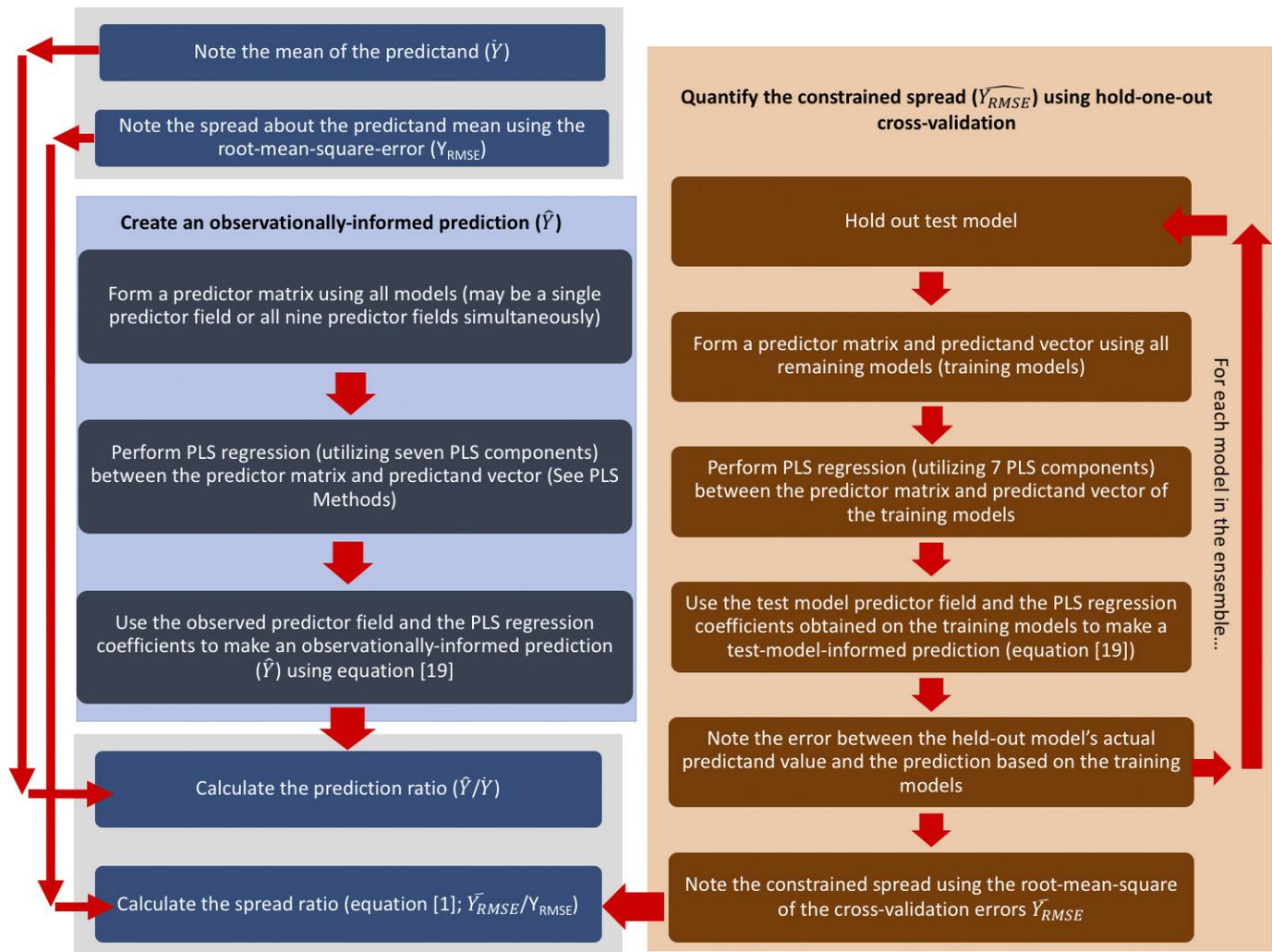
Extended Data Figure 1 | Across-model relationships between short-term variability in the shortwave cloud radiative effect and long-term changes between the present and the end of the twenty-first century. The relationship between predictor and predictand depends both on the parameter chosen as the predictand and on the location used for the predictor. **a**, Relationship (at $20^{\circ}\text{S}, 20^{\circ}\text{E}$) between the standard deviation of the climatological seasonal cycle (σ) in the downward shortwave cloud radiative effect ($\downarrow\text{CRE-SW}$) over the period 2001–2015 and the long-term change (Δ) in the $\downarrow\text{CRE-SW}$ (mean from 2085–2099 minus the mean

from 2001–2015). **b**, As in **a** but showing the relationship with GMSAT change. Grey confidence bounds are $\pm 2\sigma$ for the full model range, while the red confidence bounds are $\pm 2\sigma$ using the linear relationship between the predictor and the predictand. **c**, Relationship (at $45^{\circ}\text{S}, 131^{\circ}\text{E}$) between σ for $\downarrow\text{CRE-SW}$ over the period 2001–2015 and GMSAT change. **d**, As in **c** but for $30^{\circ}\text{N}, 10^{\circ}\text{E}$. The linear regression slope, Pearson’s correlation coefficient r and standard Pearson’s P -value of the correlation coefficient are shown.

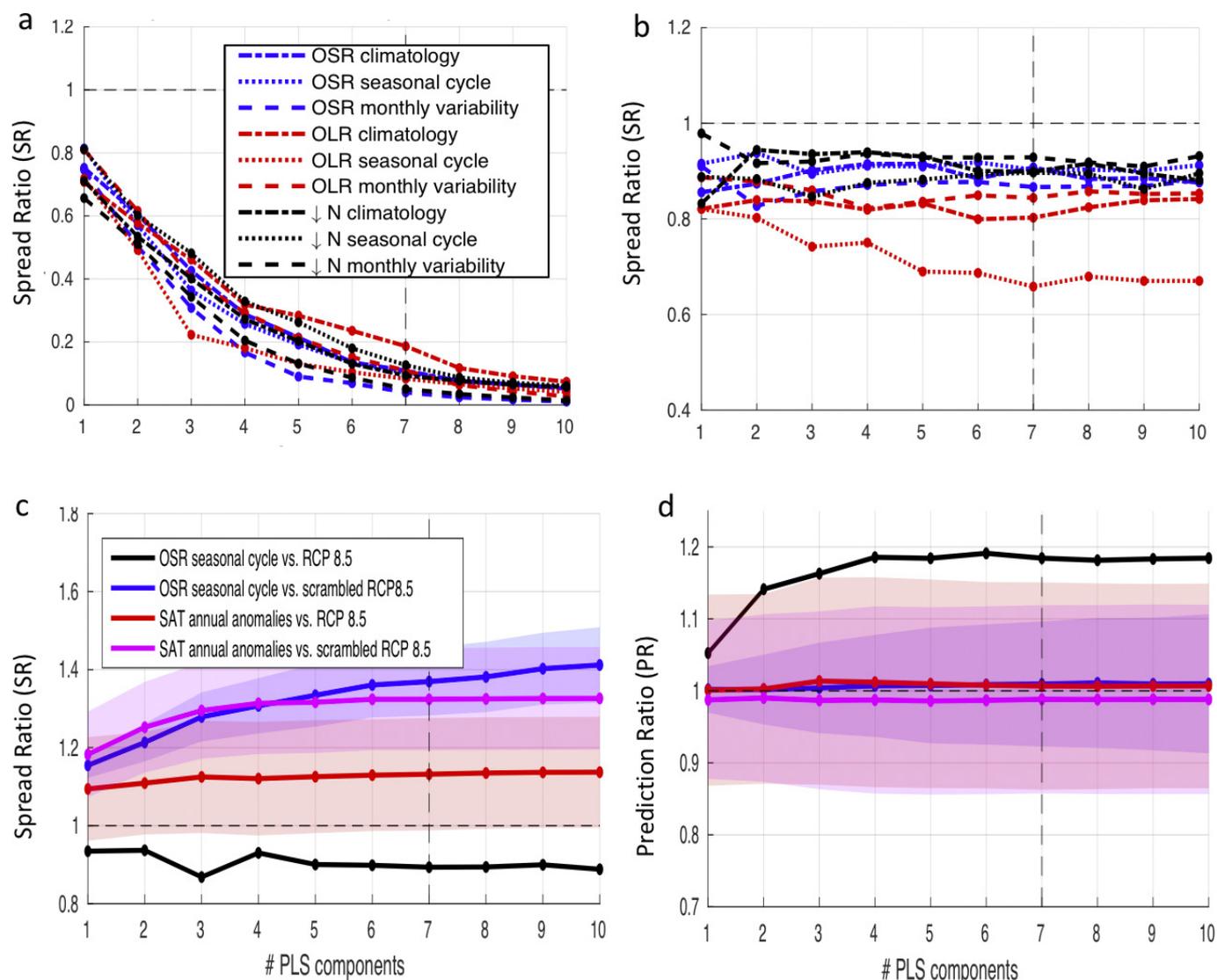


Extended Data Figure 2 | Size of model spread compared to observational uncertainty. a–c, Model spread (σ) of climatological OSR, OLR and $\downarrow N$ (see colour scale) d–f, Ratio of local model spread to CERES

observational uncertainty (see colour scale). The global spatial mean of each map is displayed at the top of each panel.

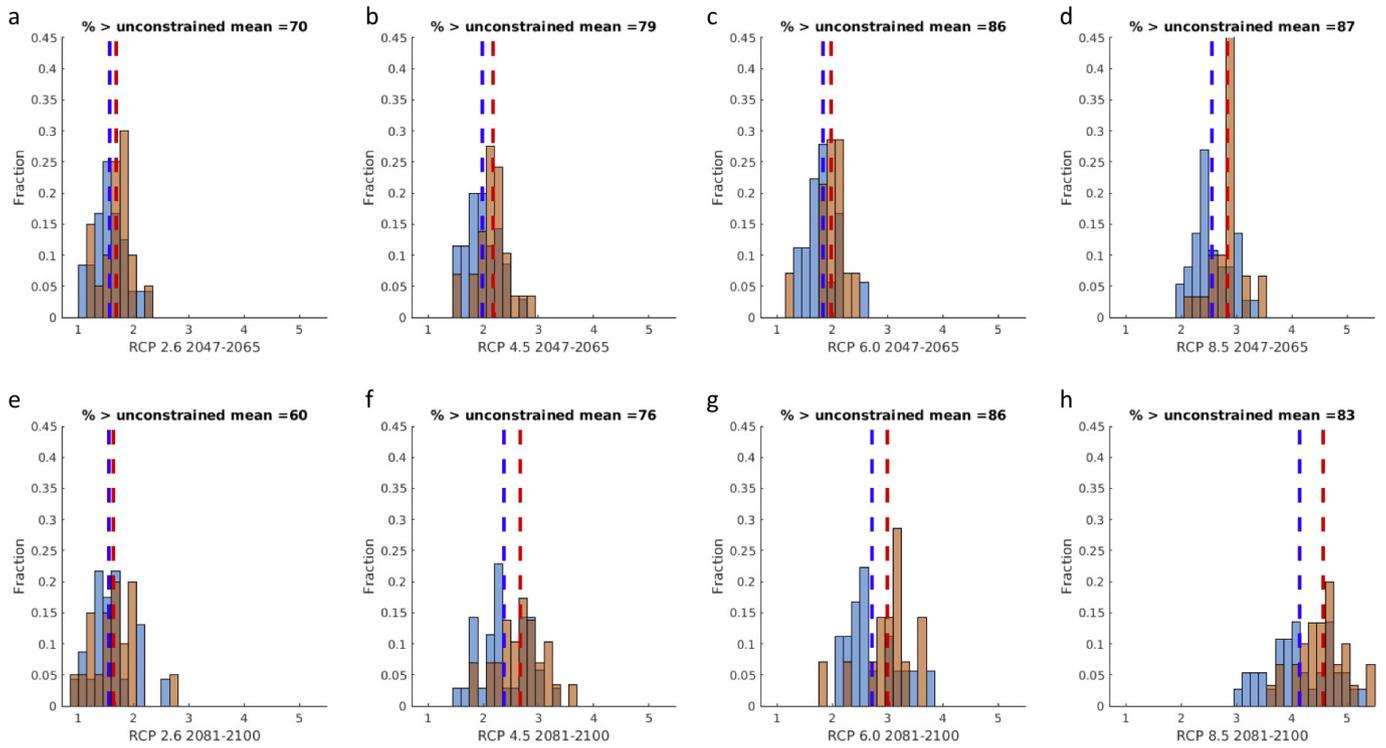


Extended Data Figure 3 | Flow chart summarizing the statistical procedure that is conducted in order to arrive at the prediction ratio and spread ratio. See Supplementary Video 1 for an animation of the procedure.



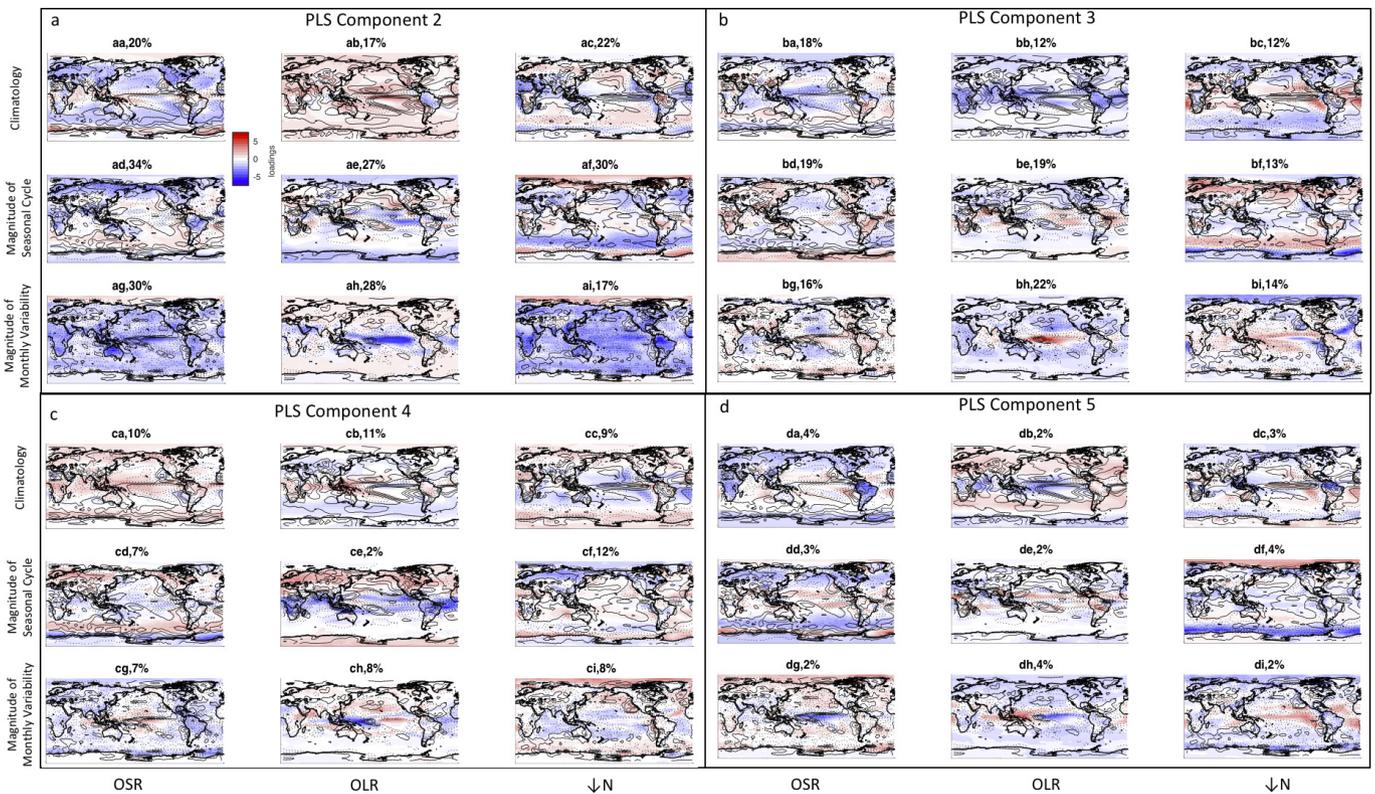
Extended Data Figure 4 | Tests of spread ratio and prediction ratio robustness. **a**, Spread ratios, as a function of the number of PLS components used, for the nine energy-budget predictor fields, each individually targeting the $\Delta T_{2090-RCP8}$ predictand without the use of cross-validation. **b**, Same as **a** but using fourfold cross-validation. **c**, Spread ratios for test data that would not be expected to result in any predictive skill between the predictor and predictand (see Methods) using

hold-one-out cross-validation. The blue and magenta lines correspond to experiments where the predictand vectors have had their values randomly scrambled or reordered. **d**, As in **c** but showing prediction ratios. The 2σ ranges of the test data across all trials are shaded in **c** and **d**. For context, the test data results are compared to one particular predictor + predictand combination from our main results (the OLR predictor field targeting the $\Delta T_{2090-RCP8.5}$ predictand, black line).



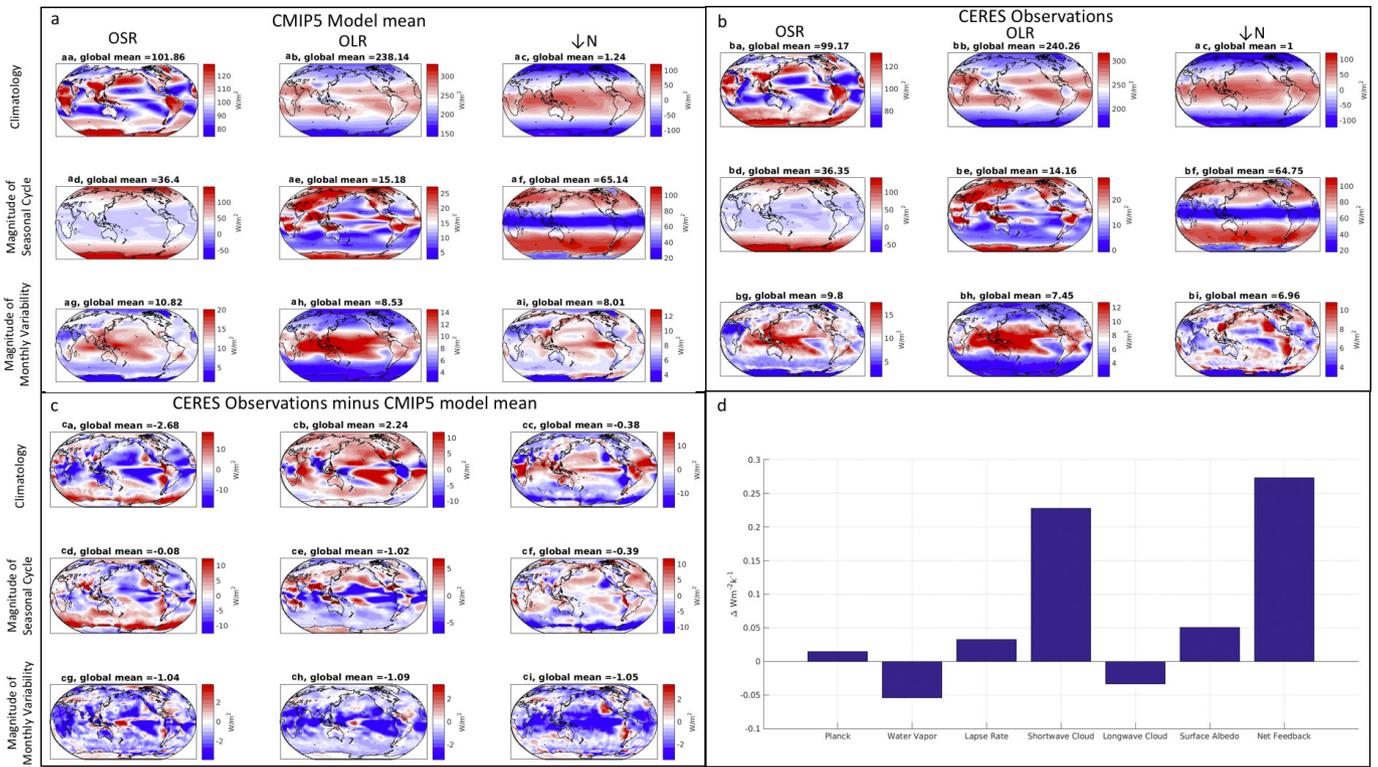
Extended Data Figure 5 | Histograms for the raw, unconstrained and observationally informed projections. a–d, Distributions for mid-century (2046–2065). **e–h,** Distributions for the end of the century (2081–2100). Raw, unconstrained model distributions are shown in blue and observationally informed (using all nine predictor fields simultaneously)

distributions are shown in orange. The blue and red dashed lines indicate distribution means. The percentage of the constrained distribution that is larger than the mean of the unconstrained distribution is displayed in the title of each panel.



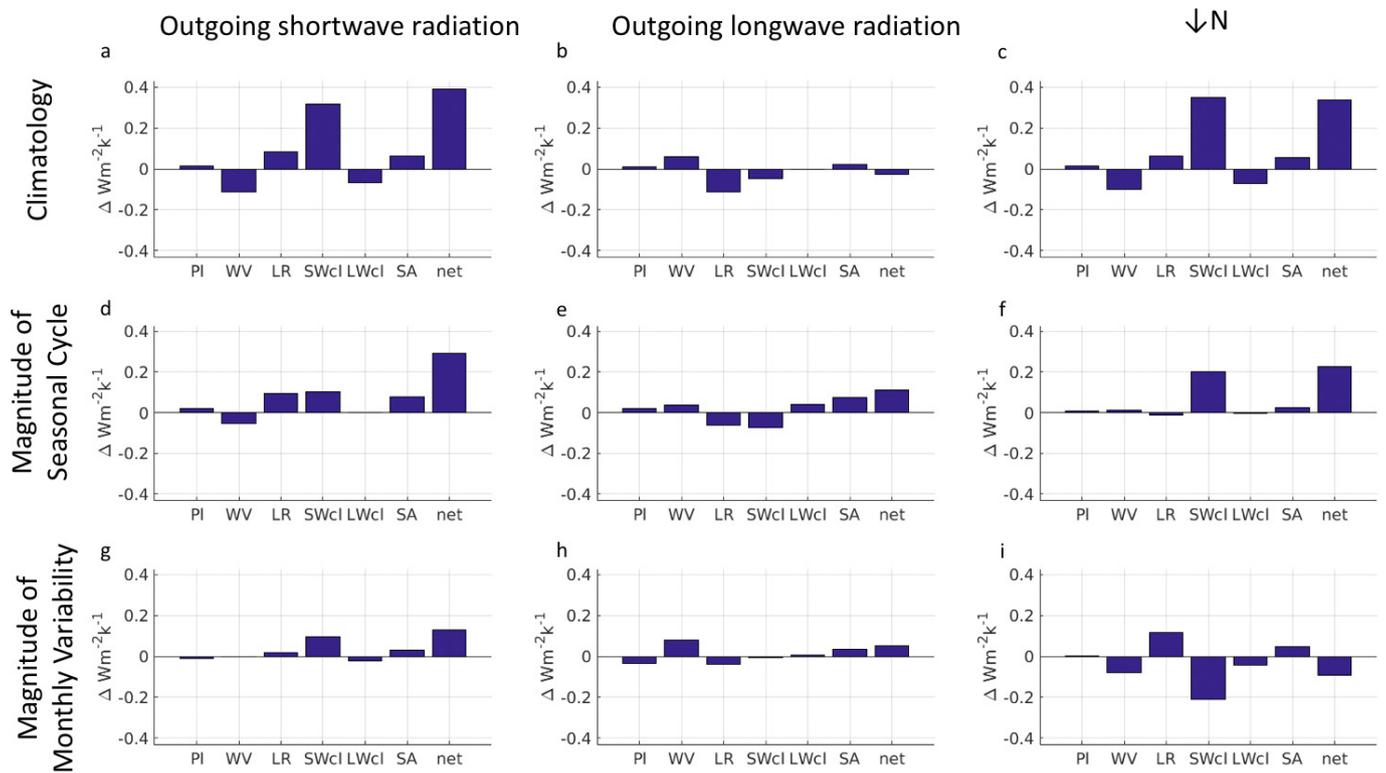
Extended Data Figure 6 | PLS loadings for PLS components 2–5. Panels **aa–ai** correspond to the nine predictor fields' second PLS component, panels **ba–bi** correspond to the third PLS component of the nine predictor fields, panels **ca–ci** correspond to the fourth PLS component of the nine

predictor fields and panels **da–di** correspond to the fifth PLS component of the nine predictor fields. The number on top of each panel is the variance explained in the $\Delta T_{2090-RCp8.5}$ predictand.



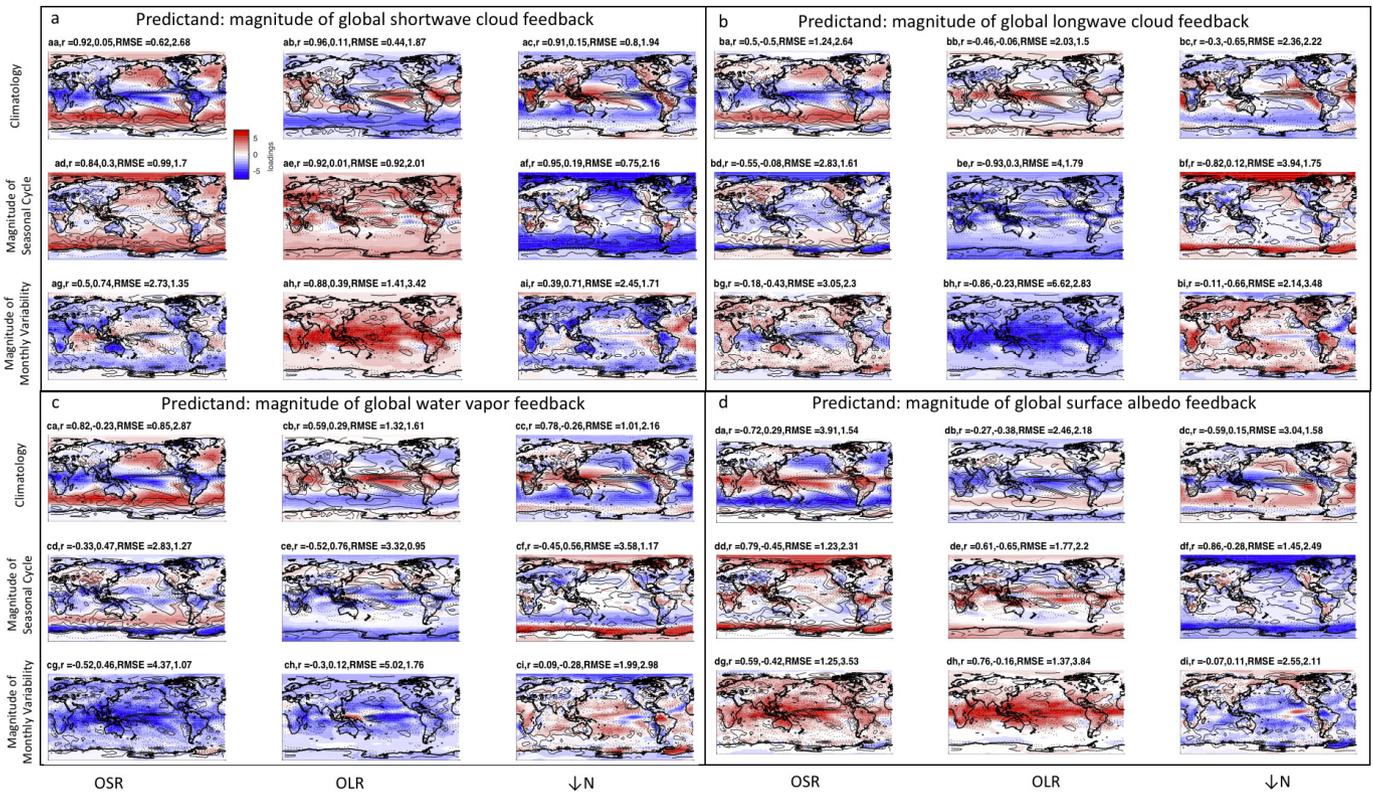
Extended Data Figure 7 | Observed/modelled predictor fields and observationally informed changes in fast feedback magnitude. **aa–ai**, Model-mean value of the nine energy-budget predictor fields calculated over the period 2001–2015. Colour bars are centred on the global mean. **ba–bi**, CERES satellite observations of the nine energy-budget predictor fields calculated over the period 2001–2015. Colour bars are centred on the global mean. **ca–ci**, CERES satellite observations minus the model

mean of the nine energy-budget predictor fields. **d**, Difference between the observationally informed and raw model-mean prediction (analogous to the prediction ratio, but taking the difference rather than the ratio) for the magnitude of six fast feedbacks (Planck, water vapour, lapse rate, shortwave cloud, longwave cloud, surface albedo) and the net feedback reported in ref. 52. Extended Data Fig. 8 shows an analogous figure but using each of the nine predictor fields separately.



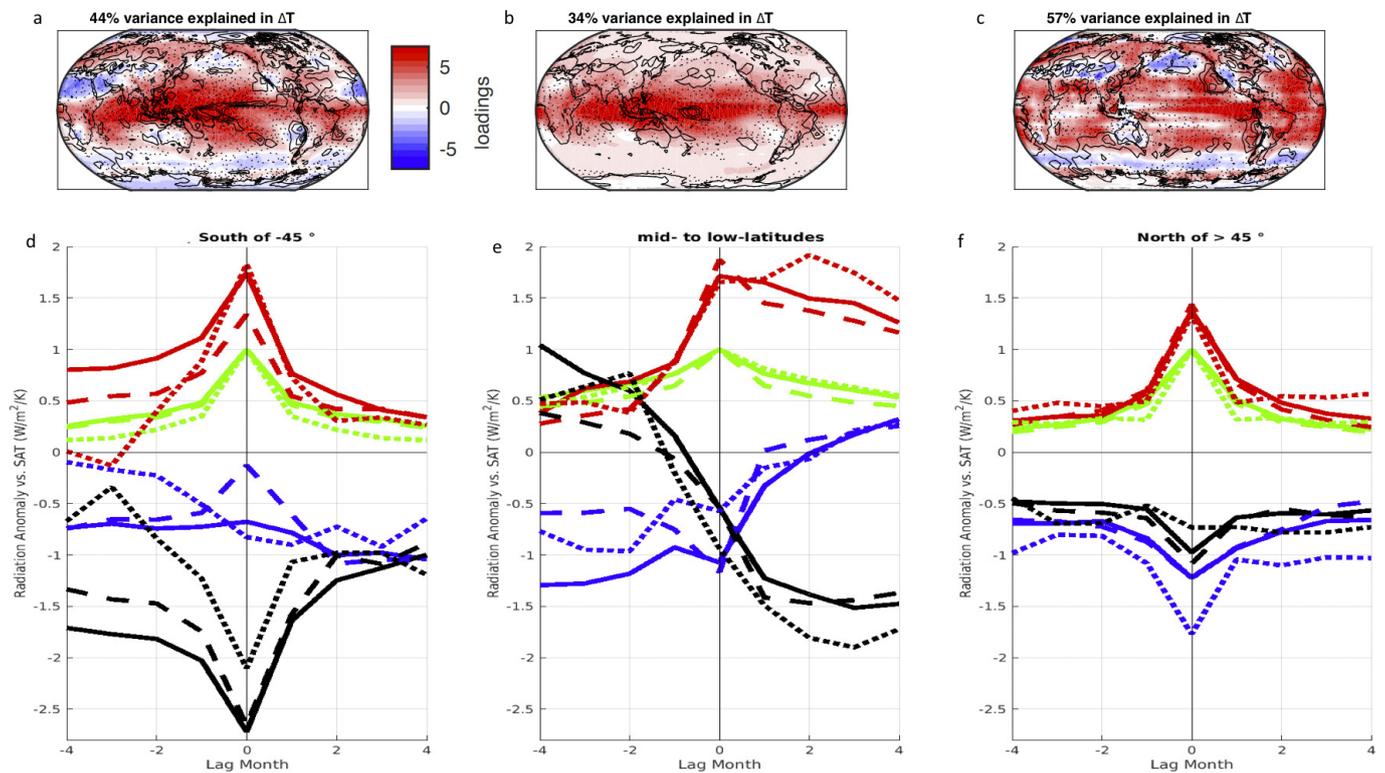
Extended Data Figure 8 | The association of the energy-budget predictor fields with feedback strength. a–i, Difference between the observationally informed and raw model mean for the magnitude of six fast feedbacks (PI, Planck; WV, water vapour; LR, lapse rate; SWcl, shortwave cloud; LWcl, longwave cloud; SA, surface albedo) and the net feedback from ref. 52, corresponding to each of the nine predictor fields individually.

shortwave cloud; LWcl, longwave cloud; SA, surface albedo) and the net feedback from ref. 52, corresponding to each of the nine predictor fields individually.



Extended Data Figure 9 | PLS loadings for the magnitude of different feedbacks. **aa–ai**, Targeting the magnitude of shortwave cloud feedback. **ba–bi**, Targeting the magnitude of the longwave cloud feedback. **ca–ci**, Targeting the magnitude of the water vapour feedback. **da–di**, Targeting the magnitude of the surface albedo feedback. These are the PLS loading patterns (equation (10)) associated with the first PLS component. Each

panel shows the Pearson’s pattern correlation coefficient r as well as the RMSE between the given map and the associated map targeting the ΔT predictand shown in Fig. 3 and Extended Data Fig. 10a–c. The two r numbers and the two RMSE numbers correspond to each panel’s relationship with the first and second PLS loading patterns associated with the ΔT predictand, respectively.



Extended Data Figure 10 | Magnitude of monthly variability relationship to ΔT . a–c, PLS loadings of the first PLS component for the predictor fields associated with the magnitude of the monthly-variability predictor. Positive loadings indicate that models with larger values tend to simulate more twenty-first century global warming and negative loadings indicate that models with smaller values tend to simulate more twenty-first-century global warming (see equation (10) in Methods). d–f, Cross-regression coefficients between monthly time series of

components of the energy budget and surface air temperature separated by latitude bands. Solid lines represent the model mean for the more-sensitive models (models with ΔT above the model median) and dashed lines represent the model mean for the less-sensitive models (models with ΔT below the model median). Negative (positive) values on the x axis indicate variability preceding (following) surface air temperature in time. CERES observations are shown as dotted lines.