Spatial radiative feedbacks from internal variability using multiple regression

Article

Accepted Version

Bloch-Johnson, J., Rugenstein, M. and Abbot, D. S. (2020) Spatial radiative feedbacks from internal variability using multiple regression. Journal of Climate, 33 (10). pp. 4121-4140. ISSN 1520-0442 doi: https://doi.org/10.1175/JCLI-D-19-0396.1 Available at http://centaur.reading.ac.uk/90824/

To link to this article DOI: http://dx.doi.org/10.1175/JCLI-D-19-0396.1

Publisher: American Meteorological Society

Spatial radiative feedbacks from internal variability using multiple regression

Jonah Bloch-Johnson*

NCAS-Climate, University of Reading, Reading, UK

Maria Rugenstein

Max Planck Institute for Meteorology, Hamburg, Germany

Dorian S. Abbot

Department of the Geophysical Sciences, University of Chicago, Chicago, Illinois, USA

**Corresponding author address:*

E-mail: j.bloch-johnson@reading.ac.uk

ABSTRACT

The sensitivity of the climate to CO_2 forcing depends on spatially-varying radiative feedbacks which act both locally and nonlocally. We assess whether a method employing multiple regression can be used to estimate local and nonlocal radiative feedbacks from internal variability. We test this method on millennial-length simulations performed with six coupled atmosphere-ocean general circulation models (AOGCMs). Given the spatial pattern of warming, the method does quite well at recreating the top-of-atmosphere flux response for most regions of the Earth, except over the Southern Ocean where it consistently overestimates the change, leading to an overestimate of the sensitivity. For five of the six models, the method finds that local feedbacks are positive due to cloud processes, balanced by negative nonlocal shortwave cloud feedbacks associated with regions of tropical convection. For four of these models, the magnitude of both are comparable to the Planck feedback, so that changes in the ratio between them could lead to large changes in climate sensitivity. The positive local feedback explains why observational studies that estimate spatial feedbacks using only local regressions predict an unstable climate. The method implies that sensitivity in these AOGCMs increases over time due to a reduction in the share of warming occurring in tropical convecting regions and the resulting weakening of associated shortwave cloud and longwave clear-sky feedbacks. Our results provide a step towards an observational estimate of time-varying climate sensitivity by demonstrating that many aspects of spatial feedbacks appear to be the same between internal variability and the forced response.

1. Introduction

Forecasting global warming is one of climate science's key challenges. As the atmospheric carbon dioxide concentration increases, the planet's radiation of energy to space becomes less than its absorption of sunlight (Arrhenius 1896). This energy imbalance, the *radiative forcing*, warms the surface, setting off processes (*radiative feedbacks*) that close the imbalance, restoring the system to a new steady state. We call the global average of the radiative feedbacks the *climate feedback* (also called the climate feedback parameter, Charney et al. (1979), or the thermal damping rate, Dessler (2012)). The total warming in response to a given increase in CO_2 is thus determined by the resulting radiative forcing and the climate feedback (Charney et al. 1979). The rate of warming also involves the thermal inertia of the surface, mostly due to oceanic heat uptake (Gregory et al. 2002). Uncertainty in the climate feedback contributes the most to uncertainty in future warming (Otto et al. 2013; Lewis and Curry 2015; Lutsko and Popp 2019), in part because of the inverse relationship between feedback and sensitivity (Roe and Baker 2007).

Directly simulating radiative feedbacks is difficult primarily because cloud feedbacks depend on small-scale processes (Wetherald and Manabe 1988). Alternatively, the climate feedback can be inferred from observations, either by solving for it using the observed warming, observed deep ocean heat uptake, and simulated radiative forcing (Gregory et al. 2002; Otto et al. 2013), or by analyzing how the planet's energy imbalance changes as the surface temperatures varies month-to-month or year-to-year (Forster and Gregory 2006; Murphy et al. 2009; Dessler 2010; Cox et al. 2018; Lutsko and Takahashi 2018; Jiménez-de-la Cuesta and Mauritsen 2019; Libardoni et al. 2019). These observational methods often assume that the climate feedback is constant, but many studies have shown that it typically changes with time in simulations (e.g., Murphy 1995; Watterson 2000; Senior and Mitchell 2000; Armour et al. 2012; Jonko et al. 2012; Andrews et al.

2015). While the temperature dependence of feedbacks can cause this to occur under sufficient (and likely strong) warming (Meraner et al. 2013; Bloch-Johnson et al. 2015), the change occurs even after relatively small amounts of warming (e.g., Armour et al. 2012; Andrews et al. 2015; Rugenstein et al. 2016). Since warming in different regions sets off radiative feedbacks of different strengths, the inconstancy of the climate feedback is likely caused by the change in the spatial pattern of warming with time (Winton et al. 2010; Armour et al. 2012). Since the temperature pattern associated with internal variability differs from the forced response, we should expect the climate feedback associated with each to differ (Dessler 2012; Colman and Hanson 2017), and in fact the climate feedback appears to vary across the historical record (Gregory and Andrews 2016; Fueglistaler 2019). The climate feedback may vary between historical and future warming (Zhou et al. 2016; Armour 2017; Proistosescu and Huybers 2017; Andrews et al. 2018), although the importance of this effect may be modest (Lewis and Curry 2018).

Recent modelling work has explored a new framework in which the climate feedback is a linear combination of radiative feedbacks associated with different regions of the surface, weighted by the temperature change in each region (Zhou et al. 2017; Dong et al. 2019). This assumes that the spatial radiative feedbacks themselves are constant, with only the map of surface temperature change evolving. This paper explores a corollary: since internal variability creates an ever-changing pattern of surface temperature and top-of-atmosphere radiative imbalance, a sufficiently long record of this variability should exhibit the behavior of these spatial radiative feedbacks. In this paper, we propose and evaluate a multiple regression (MR) method to estimate the spatial radiative feedbacks of six atmosphere-ocean general circulation models from control simulations, which we compare to existing methods for estimating feedbacks from internal variability (Section 2). We do so in spite of the known bias in regression methods related to stochastic variation in top-of-atmosphere fluxes (Spencer and Braswell 2008, 2011; Choi et al. 2014; Proistosescu et al.

2018). We test the method by convolving the estimated spatial feedbacks with warming patterns from forced simulations performed with the respective models (Section 3), assessing the method's accuracy in recreating aspects of the forced response. We discuss insights the MR method provides into climate dynamics, such as the competing nature of local and nonlocal cloud feedbacks (Section 4) and summarize our findings (Section 5).

2. Illustrating the MR method with a conceptual model

In this section, we present a method for predicting spatial feedbacks from records of unforced variability using multiple regression. We first set up a conceptual climate model designed to illustrate the method and capture some features of the complex climate models discussed in Section 3. This conceptual model has two regions of equal area. In each, the change in surface temperature (T_i) is proportional to the net energy gain of that region, which is the sum of the net downwards top-of-atmosphere (TOA) radiative flux (N_i), the net gain from horizontal energy transport from the atmosphere and ocean combined ($-H$ in region 1, H in region 2), and additional random forcing ($F_{surf,i}$):

$$c_1 \frac{dT_1}{dt} = N_1 - H + F_{surf,1} \quad (1)$$

$$c_2 \frac{dT_2}{dt} = N_2 + H + F_{surf,2} \quad (2)$$

where c_i is the surface thermal inertia associated with region i. This model can be re-expressed in terms of anomalies relative to an initial equilibrium state, so that we consider T_i', N_i', H', and $F_{surf,i}'$ instead of T_i, N_i, H, and $F_{surf,i}$. We assume that heat transport is proportional to the temperature gradient between the two regions:

$$H' = \gamma(T_1' - T_2') \quad (3)$$

Changes in a region's top-of-atmosphere radiative fluxes are caused by radiative feedbacks ($\lambda_{i,j}$, which represents the influence of surface temperature in region j on the net TOA flux in region i), radiative forcing due to changes in a forcing agent such as an increase in CO_2 ($F_{CO_2,i}$), and radiative forcing due to random atmospheric fluctuations that occur independently of surface temperature ($F_{TOA,i}$):

$$N_1' = \lambda_{1,1}T_1' + \lambda_{1,2}T_2' + F_{CO_2,1} + F_{TOA,1} \tag{4}$$

$$N_2' = \lambda_{2,1}T_1' + \lambda_{2,2}T_2' + F_{CO_2,2} + F_{TOA,2} \tag{5}$$

$\lambda_{1,1}$ and $\lambda_{2,2}$ are local radiative feedbacks, while $\lambda_{1,2}$ and $\lambda_{2,1}$ are nonlocal radiative feedbacks (where our sign convention ensures that a negative λ implies a negative, stabilizing feedback).

Nonlocal radiative feedbacks (Rugenstein et al. 2016; Zhou et al. 2017; Po-Chedley et al. 2018; Dong et al. 2019) are changes in a region's top-of-atmosphere flux that occur due to changes in surface temperature elsewhere, independent of local surface temperature changes. For example, in Figure 1, regions 1 and 2 represent the convecting and subsiding branches of an overturning cell respectively. Surface warming in region 1 propagates vertically, warming region 1's free troposphere, and then horizontally into the free troposphere of region 2, increasing H'. Region 2 now has a warmer troposphere, which radiates more, decreasing N_2'. The resulting horizontal advection may also increase the humidity of region 2's free troposphere, increasing N_2'. Assuming region 2 has a subsidence-induced boundary layer inversion, its low cloud cover could also increase, causing a further decrease in N_2'. All of these changes in N_2' occur independently of any changes in T_2', and conspire to make $\lambda_{2,1}$ positive or negative.

We note that an increase in H' will also increase T_2' directly (Eq. 1; Feldl and Roe 2013b). While this latter effect is connected to nonlocal radiative feedbacks in that both occur due to horizontal fluxes of heat and moisture, the two effects are different, and can disagree in the sign of the resulting surface warming, as demonstrated by the above example. While the influence of H' on surface temperature is important for understanding the evolution of the spatial pattern of warming, in this paper we are focused only on the influence of surface temperature on TOA radiative fluxes, and so we focus on nonlocal radiative feedbacks.

Suppose that region 1 has a weak positive local feedback $\lambda_{1,1} = 0.5$ Wm^{-2}K^{-1} (red solid line, Figure 2b), and a stronger negative nonlocal feedback, so that $\lambda_{2,1} = -2$ Wm^{-2}K^{-1} (light blue solid line, Figure 2b). We also assume that the surface temperature of the subsiding region 2 has no net effect on TOA fluxes, so that $\lambda_{1,2} = \lambda_{2,2} = 0$ Wm^{-2}K^{-1} (orange and gray solid lines in Figure 2b). We assume that region 2's thermal inertia is much larger than region 1's, representing more ocean heat uptake in this region (see Appendix for details).

We define the global climate feedback λ to be the dependence of the globally averaged net TOA flux on the globally averaged surface temperature, that is

$$\lambda(t) = \frac{\partial \overline{N}}{\partial \overline{T}}(t) = \sum \left(\Lambda \frac{d\vec{T}}{dt}(t) \right) / \frac{d\overline{T}}{dt}(t) \tag{6}$$

where $\vec{T} = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}$, $\Lambda = \begin{bmatrix} \lambda_{1,1} & \lambda_{2,1} \\ \lambda_{1,2} & \lambda_{2,2} \end{bmatrix}$, and a bar over a vector indicates the global average of that vector. We do not have to use an anomaly for $\overline{N}$ because $\overline{N}$ is 0 in equilibrium. Note that even though the spatial feedbacks Λ are constant, the global feedback λ can change with time because of the evolving spatial pattern of warming $\frac{d\vec{T}}{dt}(t)$.

We perform two 5000-year experiments: a "control" experiment, where all variations in $\vec{T}'_{control}(t)$ and $\vec{N}'_{control}(t)$ are due to random forcing at the surface ($\vec{F}'_{surf}(t) = \begin{bmatrix} F'_{surf,1}(t) \\ F'_{surf,2}(t) \end{bmatrix}$) and TOA ($\vec{F}_{TOA}(t) = \begin{bmatrix} F'_{TOA,1}(t) \\ F'_{TOA,2}(t) \end{bmatrix}$), and an "abrupt4x" experiment in which the time series $\vec{T}'_{abrupt4x}(t)$ and $\vec{N}'_{abrupt4x}(t)$ also respond to an initial step forcing akin to a quadrupling of CO_2 concentration ($F_{CO_2,1} = F_{CO_2,2} = 8$ Wm^{-2}).

For the abrupt4x simulation, the climate feedback $\lambda = \frac{\partial \overline{N}}{\partial \overline{T}'}$ changes significantly around year 20. We therefore define two forced feedbacks, $\lambda_{4x,early}$ and $\lambda_{4x,late}$, which are the slopes of the linear regressions of $\overline{N}_{abrupt4x}(t)$ against $\overline{T}'_{abrupt4x}(t)$ taken over years 1 to 20 and years 21 to 5000 respectively (Figure 2c). Before these regressions are taken, we average each annual time series (gray dots) over roughly exponentially increasing time periods (colored dots). $\Delta\lambda_{4x} \equiv \lambda_{4x,late} - \lambda_{4x,early}$ is the change in feedback between the periods.

We seek a method to predict $\lambda_{4x,early}$, $\lambda_{4x,late}$, and $\Delta\lambda_{4x}$ given $\vec{T}'_{control}(t)$ and $\vec{N}'_{control}(t)$ (internal variability), and $\vec{T}'_{abrupt4x}(t)$ (the spatial pattern of warming). The simplest method would be to regress annual averages of $\overline{N}_{control}(t)$ against $\overline{T}_{control}(t)$ to get the resulting regression slope $\lambda_{control}$ (the slope of the blue line in Figure 2a), and to assume that $\lambda_{4x,early} = \lambda_{4x,late} = \lambda_{control}$ (Forster and Gregory 2006; Murphy et al. 2009; Dessler 2010). We call this the "global" method because it uses information about changes in global surface temperature only.

The radiative feedbacks associated with temperature change induced by random forcing (i.e., $\vec{F}_{surf}$ and $\vec{F}_{TOA}$) differ from those induced by uniform greenhouse forcing ($\vec{F}_{CO_2}$) (Dessler 2012; Colman and Hanson 2017; Proistosescu et al. 2018). Our conceptual model illustrates how this can arise from spatial variation. Since the thermal inertia in region 2 is larger, most of the temperature variability occurs in region 1, so that $\lambda_{control}$ is weighted towards the feedbacks associated with this region ($\lambda_{control} \approx \lambda_{1,1} + \lambda_{2,1}$). The spatial pattern of warming in the forced response is initially dominated by region 1 as well, once more because it has the lowest thermal inertia. As a result, the global method predicts $\lambda_{4x,early}$ well (see Figure 2c and d). However, the global method always predicts $\Delta\lambda_{4x} = 0$, as it assumes a constant λ. Since warming moves to region 2 over time and $\lambda_{1,2} + \lambda_{2,2} > \lambda_{1,1} + \lambda_{2,1}$, $\Delta\lambda_{4x}$ is positive. As a result, the global method underpredicts the warming of the abrupt4x simulation by about 1.5 K (Figure 2c). To address this shortcoming, we need a method that accounts for the spatial variation of feedbacks.

The "local" method is a commonly used method (Boer and Yu (2003b), Crook et al. (2011), the "local" method in Feldl and Roe (2013a), Brown et al. (2015), and Trenberth et al. (2015)) for estimating spatial feedbacks. In this method, we construct $\vec{\lambda}_{local} = \begin{bmatrix} \lambda_{1,local} \\ \lambda_{2,local} \end{bmatrix}$ where $\lambda_{i,local}$ is the result of regressing $N'_{i,control}(t)$ against $T'_{i,control}(t)$. Taking the dot product of $\vec{\lambda}_{local}$ with $\vec{T}'_{abrupt4x}(t)$ then provides an estimate of $\vec{N}'_{abrupt4x}(t)$ which we can use to estimate $\lambda_{4x,early}$, $\lambda_{4x,late}$, and $\Delta\lambda_{4x}$.

This method assumes all radiative feedbacks are local, while allowing for the nonlocal effects of heat transport (Feldl and Roe 2013b). However, if there are nonlocal radiative feedbacks, then the local method can miss or conflate their effects. In region 1, estimates of $\lambda_{1,local}$ tend toward $\lambda_{1,1} = 0.5$ Wm^{-2}K^{-1} (dotted red line, Figure 2b), missing the negative nonlocal feedback $\lambda_{2,1}$. Since the early period is dominated by warming in region 1, the local method overestimates $\lambda_{4x,early}$ (where "overestimates" implies the estimate of $\lambda_{4x,early}$ is more positive than the true value, even if both are negative, resulting in an overestimate of the sensitivity). On the other hand, T'_2 tends to be positively correlated with T'_1, due to heat transport, while T'_1 tends to be anti-correlated with N'_2 because $\lambda_{2,1}$ is negative. As a result, the local method predicts that $\lambda_{2,local}$ is negative (dotted orange line, Figure 2b), even though T'_2 has no net influence on N. Since T'_2 contributes more to warming over time, the local method incorrectly predicts a more negative feedback (Figure 2c and d). Similar discrepancies can occur when local feedbacks are used to diagnose feedbacks in GCMs, which may explain instances when the local method fails to predict feedback changes properly (Rose et al. 2014). We need a method that includes nonlocal feedbacks while accounting for correlation between temperature in different regions.

We propose a multiple regression ("MR") method, which estimates the local and nonlocal feedbacks associated with N'_i (that is, the influence of T'_1 and T'_2 on N'_i) by regressing $N'_{i,control}(t)$ against both regions simultaneously:

$$N'_{i,control}(t) = \lambda_{i,1,MR} T'_{1,control}(t) + \lambda_{i,2,MR} T'_{2,control}(t) + F_{TOA,i} \tag{7}$$

In least squares multiple regression, $\lambda_{i,j,MR}$ is the same as the slope of the regression of $N'_{i,control}(t)^*$ against $T'_{j,control}(t)^*$, where the star indicates that each time series is the residual after regressing against the surface temperatures in all non-j regions (see Appendix). This removes the effect of correlations between surface temperature in different regions giving spurious feedbacks, as with $\lambda_{2,local}$ above. Multiple regression has been used to estimate other surface temperature-dependent feedbacks from internal variability, though not radiative feedbacks (Liu et al. 2008; Li et al. 2012; Li and Forest 2014; Liu et al. 2018). The dashed lines in Figure 2b show that, given sufficient time, the MR method predicts the local and nonlocal feedbacks in each region, so that when we multiply the full matrix of estimated spatial feedbacks $\Lambda_{MR} = \begin{bmatrix} \lambda_{1,1,MR} & \lambda_{1,2,MR} \\ \lambda_{2,1,MR} & \lambda_{2,2,MR} \end{bmatrix}$ by $\vec{T}'_{abrupt,4x}(t)$ to estimate $\vec{N}_{abrupt4x}(t)$, the resulting estimates $\lambda_{4x,early}$, $\lambda_{4x,late}$, and $\Delta\lambda_{4x}$ are accurate (Figure 2c and d). Therefore, for this example, the MR method is able to account for the difference in climate feedback between internal variability and the forced response.

Random fluctuations in N influence T via planetary energy gain at the same time that T influences N via radiative feedbacks. As a result, T will tend to lag N with a positive correlation, while N will lag T with a negative correlation, so that regressions taken without a lag will be biased towards 0 (Spencer and Braswell 2008, 2011; Choi et al. 2014; Proistosescu et al. 2018). This issue does not occur for random forcing at the surface, which only affects N indirectly through radiative feedbacks. Therefore, the more stochastic forcing that occurs at TOA ($\vec{F}_{TOA}$) as opposed to the surface ($\vec{F}_{surf}$), the more the regression of N vs. T will overestimate the true radiative feedback. For the example in Figure 2, $F_{surf,1}$ and $F_{surf,2}$ are white noise with variance 20 $\mathrm{W^2m^{-4}}$, while $F_{TOA,1}$ and $F_{TOA,2}$ are white noise with variance 5 $\mathrm{W^2m^{-4}}$. Figure S1 shows a case where these variances are 10 and 15 W^2m^{-4} respectively, with the result that all three regression methods overestimate $\lambda_{4x,early}$ and $\lambda_{4x,late}$, while underestimating $\Delta\lambda_{4x}$. In other words, given sufficient random TOA forcing, regression estimates of spatial feedbacks will be biased. We consider this bias in discussing our results in the next section.

It should be mentioned that Proistosescu et al. (2018) model ENSO variability as a distinct additional mechanism by which N and T mutually influence each other, which similarly leads to overestimates of λ from regression-based methods. As part of their model, they assume that T influences N with a lag of about three months. Since this is beyond the time scale of most atmospheric processes, we assume that this feedback propagates in part through the ocean, so that the atmospheric component may still operate through the same spatial feedbacks that operate under other forms of variability and under the forced response (e.g., it could occur due to a "tropical atmospheric bridge" mechanism; Klein et al. 1999).

3. Using the MR method on AOGCMs

To test the methods discussed above on atmosphere-ocean general circulation models (AOGCMs), we use simulations from LongRunMIP, an archive of fully coupled millennial-length simulations of complex climate models (Rugenstein et al. 2019). We chose the six models with millennial-length control and abrupt4x simulations for which we have monthly output. Details of these models and simulations are given in Table S1.

We alter the three methods from Section 2 to reflect the more complex nature of AOGCMs:

- CO_2 forcing can lead to atmospheric changes that are independent of surface warming. These "adjustments" to forcing occur mostly within the first year (e.g., Gregory and Webb 2008). We remove this year from our analysis, redefining our early period to be years 2 to 20.

- For AOGCMs, there are more than two regions with distinct behaviors. Dividing our models into n regions, equation 7 becomes

$$N_i'(t) = \lambda_{i,1,MR} T'_{1,control}(t) + \lambda_{i,2,MR} T'_{2,control}(t) + \ldots + \lambda_{i,n,MR} T'_{n,control}(t) + F_{TOA,i}, \tag{8}$$

giving a system of n equations

$$\vec{N}'(t) = \Lambda \vec{T}'(t) + \vec{F}_{TOA} \tag{9}$$

where Λ is a matrix of feedbacks $\lambda_{i,j}$. Each equation in this system has $n-1$ degrees of freedom, so n must be smaller than the length of the control simulation, and preferably much smaller given the significant spatial correlation of surface temperature. For simplicity, we divide the surface equally in latitude and longitude, although this may miss features of the climate system. Since our control simulations last at least 1000 years (Table S1), we use a 15° by 15° grid, giving 288 regions (Figure 3).

- Circulations, and therefore radiative feedbacks, change with season. Thus, we compute feedbacks for each season individually, first by averaging all monthly time series into seasonal time series (where the seasons are DJF, MAM, JJA, SON), and then performing a separate regression for each season (e.g. all DJF values of $\vec{N}'_{control}(t)$ against all DFJ values of $\vec{T}'_{control}(t)$) creating a set of four feedbacks. We multiply each month of $\vec{T}'_{4x}(t)$ by the relevant seasonal feedback, and take the annual average to estimate $\vec{N}'_{4x}(t)$. We compare seasonal averages to other approaches in Tables S2 and S3. While seasonal averaging tends to reduce the error in the MR method, the qualitative behavior of the different methods is not affected by the choice of time averaging.

Figures 3 and 4 show $\overline{N}$ vs. $\overline{T}'$ of the control and abrupt4x simulations of the six models respectively. Figure 4 also shows $\overline{N}$ estimated using the three methods, assuming that each estimate starts with the true value of $\overline{N}$ at year 2. The solid lines in Figure 4 are local regressions of $\overline{N}$ against $\overline{T}'$ performed using LOESS (LOcally Estimated Scatterplot Smoothing; Cleveland and Devlin 1988, see Appendix for more detail). We can use the slopes of these lines mapped against the time series of $\overline{T}$ to estimate feedbacks as a function of time (lines in Figure 5).

Though there is a range of feedback values between models, all six forced simulations have a feedback that gets less negative with time (black lines), consistent with past results for similar models (Andrews et al. 2015). The MR method (green lines) matches or overestimates the feedback value, with this error tending to decrease with time. This error can range from $\sim$1 $\mathrm{Wm^{-2}K^{-1}}$ for the early years of CESM104 and GISSE2R (that is, at least half of the feedback strength itself) to roughly 0 for HadCM3L. The MR method correctly predicts that the feedback gets less negative with time, although for some of the models it underestimates the magnitude of the change.

The global method (blue) overestimates the early feedback. Since the global method is agnostic about the pattern of surface warming, the predicted feedback is mostly constant except for small differences due to changes in the seasonal distribution of warming and in seasonal feedbacks (e.g, the early years of HadCM3L). As a result, as the true feedback increases with time, it becomes more positive than the global estimate for half the models. For some models, this allows the global method to more accurately forecast the equilibrium warming than the other methods, albeit due to compensating errors in the early and later periods (i.e., CESM104 and MPIESM12 in Figure 4).

The local method (orange) predicts a positive feedback for all models except GISSE2R, implying a climate unstable to external forcing, and does not predict the increase in feedback with time seen in all models.

The dots in Figure 5 represent estimates of $\lambda_{4x,early}$ and $\lambda_{4x,late}$ (feedbacks before and after year 20; see Appendix for details). We visualize the estimates of these feedbacks and their difference using a scatter plot (black dots in Figure 6), as in Figure 2d. The global and MR methods perform similarly for $\lambda_{4x,early}$ and $\lambda_{4x,late}$, while the MR method gets closer to accurately predicting $\Delta\lambda_{4x}$, consistent with the discussion around Figure 4 and reflected by the root mean square errors in Table 1 (for feedback values for all models and components, see Tables S7 and S8).

$\vec{N}'$ and λ can be expressed as the sum of shortwave (SW) and longwave (LW) terms, which can be separated in turn into clear-sky (fluxes recalculated as if no clouds were present) and cloud terms (the residual of total and clear-sky terms; cloud feedbacks defined this way may include changes in cloud masking rather than in clouds themselves (Soden et al. 2004)).

Examining these component individually shows that the error in $\lambda_{4x,early}$ in the MR and global methods is due primarily to SW cloud feedbacks (red markers in Figures 6a and b). Both the MR and global methods have smaller errors in $\lambda_{4x,late}$ (Figures 6d and e), but for the MR method this is caused by a reduction in the error in SW cloud, while for the global method this is due to offsetting errors in the SW and LW cloud feedbacks (see also Table 1). Cloud feedbacks are similarly the cause of the local method's large overestimation, while the local method outperforms the other methods at predicting the primarily local SW clear feedback (Table 1). Note that the global method has a relatively small error for the LW clear feedback, consistent with Lutsko and Takahashi (2018). The increase in feedback with time ($\Delta\lambda_{4x}$) and the variation in this increase between models is driven by the SW cloud feedback (Figures 6g, h, and i). The MR method has the smallest error in estimating $\Delta\lambda_{4x}$, with this error tending to be an underestimate. Figures S2-5 show feedback time series plots for all component fluxes.

All methods examined contain some degree of error. We can find the geographic source of these errors by looking at the true and estimated normalized change in $\vec{N}'_{4x}$ (multi-model mean in Figure 7; errors in the multi-model mean and for individual models in Figures S6-S8), calculated by taking the finite difference in $\vec{N}'_{4x}(t)$ between the first and last part of the indicated time period, where each part contains similar amounts of warming (see Appendix). The difference is normalized by the global temperature change, allowing intermodel comparison. For the global method, we make this estimate by regressing $\vec{N}'_{control}(t)$ against $\overline{T}'_{control}$(t) (the "global" method in Feldl and Roe (2013a) and the "local contribution" in Boer and Yu (2003a,b); Crook et al. (2011); Zelinka et al. (2012); Andrews et al. (2015)) and using this as the predicted normalized change in $\vec{N}'_{4x}$.

The MR method does quite well at recreating the multi-model spatial pattern of TOA flux change, both for net and component fluxes (Figures S9-S12), with the exception of regions south of 30°S and the north Atlantic. The MR method also overestimates the change in these regions in individual models (Figures S6-S8). The error in these regions has contributions from all component fluxes, foremost the SW cloud feedback (for multi-model mean component flux errors, see Figures S13-17). For all periods, models, and fluxes except for SW clear-sky (which is primarily a local feedback), the MR method outperforms the other two methods when scored by the area-weighted root mean square error (Table 2; for comparison with annual or monthly approaches, see Table S3; for values for individual models, see Table S4; for details on the error metric, see Appendix). Specifically, the global method has large compensating errors, especially in the tropics, and the local method overestimates the change almost everywhere (Figures S6-S8).

There are several potential explanations for the MR method's overestimate for TOA fluxes south of 30°S and over the north Atlantic. These may be regions where there is significantly more stochastic forcing at TOA than at the surface, resulting in a similar overestimation to that discussed in Section 2 and shown in Figure S1. Alternatively, the spatial feedbacks that influence $\vec{N}'$ in these regions may be nonlinear, either in that they change in value as the world warms (e.g., a reduction in the strength of the SW clear feedback once sea ice melts), or the effect of warming in different regions combines nonlinearly, as might occur in response to circulation changes such as a shift in the mid-latitude jet; or surface fluxes may influence $\vec{N}'$ there independently of surface warming. Further research is needed to diagnose this error.

In spite of this overestimate, the MR method can be used to explain the multi-model forced TOA flux response for roughly three quarters of the Earth using feedbacks estimated from internal variability (see Table S5 and S6, which show the same error metrics as Tables 1 and 2, using only TOA fluxes north of 30°S). We now discuss the spatial feedbacks estimated by the MR method, as well as some of their implications.

4. Discussion

We first test if the spatial feedbacks estimated using the MR method exhibit behavior broadly consistent with physically modelled feedbacks. The i^{th} column of Λ represents the change in $\vec{N}'$ from warming in region i. Zhou et al. (2017) performed fixed-SST experiments with the CAM5 model where the temperature in region i was perturbed. The top row of Figure 8 shows spatial cloud feedbacks for three representative regions calculated using this approach. The bottom row shows the multi-model and multi-season mean response for warming in similar regions estimated by the MR method. For both approaches, warming in the extratropics or in regions of tropical subsidence produces cloud feedbacks that are mostly local and positive, while warming in tropical convecting regions has significant nonlocal feedbacks which are mostly negative. Since the models, region sizes, and degree of perturbation differ, the details and magnitudes of the feedbacks differ. Further, the fixed-SST method allows land temperatures to evolve freely, so that regions that have significant nonlocal effects, like tropical convecting regions, can cause large changes in TOA fluxes over land (Figure 8b). The MR method is able to estimate land feedbacks directly, so that TOA flux changes due to land warming are not included in these tropical convecting feedbacks (Figure 8e). See also Figure 4 in Dong et al. (2019).

The top left panel of Figure 9 shows a map of the multi-model and multi-month mean spatial feedbacks estimated by the MR method: the change in $\overline{N}$ caused by warming in each region divided by that region's fractional area (so that smaller, polar regions do not have artificially smaller feedbacks). Spatial feedbacks are strongly negative in regions of tropical convection (e.g., Indonesia and Central America) and are mostly positive over the tropical oceans in regions of atmospheric subsidence as well as much of the extratropical oceans, in keeping with the examples from Figure 8. These strongly negative feedbacks are robust when feedbacks are recalculated using just the first or second half of the control simulations (Figures S18-22), although outside these regions there is some noise, with the sign of roughly a third of net feedback cells differing between the first and second halves. The variation in the spatial pattern is largely determined by the SW cloud feedback (bottom left panel, Figure 9; for all flux components, see Figures S19-S22).

a. Local and nonlocal feedbacks

The MR method allows us to split spatial feedbacks into local (the diagonal elements of Λ, giving the influence of warming on TOA fluxes directly overhead) and nonlocal components (the off-diagonal elements of Λ), and to calculate the local and nonlocal components of the map of spatial feedbacks (middle and right columns of Figure 9 respectively). We note that the devision between local and nonlocal feedbacks depends on grid resolution, with local feedbacks in coarser grids incorporating more nonlocal processes. For the grid considered in this paper, the local feedback is positive almost everywhere, due to cloud feedbacks (Figures S21 and S22): in the tropics and in subtropical subsiding regions, local warming reduces lower tropospheric stability, leading to a loss of low clouds and a positive SW cloud feedback (Klein and Hartmann 1993; Wood and Bretherton 2006; Zhou et al. 2017; Dong et al. 2019). This result holds for each AOGCM except for GISSE2R, which lacks a positive local SW cloud feedback (Figure S24, Table S8). For most models, there is a partially compensating negative local LW cloud feedback in tropical convecting regions, possibly due to an iris effect (Lindzen et al. 2001; Mauritsen and Stevens 2015). Outside of the tropics, there is a positive local LW cloud feedback, possibly associated with an increase in middle and high cloudiness as convection increases (Zelinka et al. 2012).

Positive local feedbacks provide an explanation for observational studies that use the local method to predict spatial feedbacks, finding that they are positive over much of the Earth and in the global mean (Brown et al. 2015; Trenberth et al. 2015). For example, the multi-model mean feedbacks estimated using the local method (top middle panel, Figure S23) resemble the feedbacks in the upper right panel of Figure 10 from Trenberth et al. (2015). While local method feedbacks can differ from the local component of MR method feedbacks due to correlation between temperature in different regions as discussed in Section 2, the observational studies provide evidence that real world local feedbacks are substantially positive. If we use the MR method to estimate the local components of $\lambda_{4x,early}$ and $\lambda_{4x,late}$ (Table S8), we get positive values for all models except GISSE2R. For these models, the mean estimated local feedback is 3.37 $\mathrm{Wm}^{-2}\mathrm{K}^{-1}$ for the early period and 3.13 $\mathrm{Wm}^{-2}\mathrm{K}^{-1}$ for the late period (Tables S8).

The MR method implies that in the absence of negative nonlocal feedbacks, five out of six of these AOGCMs would be unstable to radiative forcing, even accounting for the dominant stabilizing Planck feedback. The MR method predicts that there are strongly negative nonlocal feedbacks coming from regions of tropical convection (upper right panel, Figure 9), largely due to the SW cloud feedback (lower right panel). This is consistent with tropical convecting regions behaving similarly to region 1 of the conceptual model from Section 2: surface warming in the convecting tropics propagates throughout the tropical free troposphere, increasing the temperature aloft while leaving surface temperatures alone. This increases the lower tropospheric stability, and thus low cloud cover (a negative SW cloud feedback), as well as the troposphere's outgoing longwave radiation (a negative LW clear feedback) (Rose and Rayborn 2016; Andrews and Webb 2017; Ceppi and Gregory 2017; Klein et al. 2017; Zhou et al. 2017; Dong et al. 2019). Note that incorporating these nonlocal interactions changes both local and total values of the LW clear feedback, giving different values than studies that analyze this feedback purely locally (e.g., Koll and Cronin 2018).

For the five models with positive local components, the average nonlocal component of the abrupt4x feedbacks is -4.21 Wm^{-2}K^{-1} for the early period and -3.69 Wm^{-2}K^{-1} for the late period (Table S8). so that the net forced climate feedback is a small residual between competing local and nonlocal feedbacks, with local and nonlocal feedbacks strongly anti-correlated between different models (Table S8; the correlation coefficient for early period non-GISSE2R local vs. nonlocal feedbacks is -0.96, and for late is -0.98). A modest shift in the relative strength of these feedbacks (for example, due to a shift in circulation) could lead to large changes in climate sensitivity; an increase in the local feedback of only a third would be enough to make these AOGCMs unstable (local and nonlocal feedbacks differ by $\sim$1 Wm^{-2}K^{-1}, which is on average roughly a third of the magnitude of the local feedback for the non-GISSE2R models). Additional research is needed to understand what mechanisms cause the anti-correlation between local and nonlocal feedback strength, and whether we expect this cancellation to hold in different climate states. Given that the local/nonlocal cancellation does not hold in all contexts – for example, the nonlocal feedback's seasonal cycle has a larger amplitude and is more latitudinally constrained than the local feedback's seasonal cycle (Figure S25) – it is unlikely that this cancellation is purely a statistical artifact. Our findings have bearing for exoplanet research, as they suggest that it may be harder to have a cloudy atmosphere with a stable climate than previously thought (Leconte et al. 2013), potentially reducing the chance of finding habitable worlds.

b. The cause of the increase in climate feedback over time

For all six models, the change in feedback with time ($\Delta\lambda_{4x}$) is positive, primarily because of the SW cloud feedback, and secondarily the LW clear feedback (Figure 4 and Table S7). The MR method gets the correct sign of $\Delta\lambda_{4x}$ but underestimates this increase for each model, once more primarily due to the SW cloud feedback (Table S8).

We can estimate how much the change in the spatial pattern of warming with time (Figure 10a) contributes to $\Delta\lambda_{4x}$ by multiplying this change by the MR estimate of the spatial pattern of feedbacks for each flux component (Figure 9, Figures S18-S22). The resulting maps show the contribution of the change in warming pattern to the change in feedback (Figures 10b-f).

The MR method identifies two main latitude bands that contribute to the increase in feedback with time: the tropics, whose convecting regions increase the SW cloud and LW clear feedbacks (less warming in these regions reduces the role of the strongly negative nonlocal feedbacks discussed above, consistent with Andrews and Webb 2017; Ceppi and Gregory 2017; Dong et al. 2019; Fueglistaler 2019); and the Southern Ocean, which increases the SW clear feedback (due to the delayed warming in this region leading to the delayed melting of sea ice). The MR method estimates that the LW clear sky and SW cloud feedback have offsetting negative contributions in the Southern Ocean. While the LW clear sky offset is consistent with the total change in the LW clear feedback being small, and with the LW clear TOA flux change getting more negative in the Southern Ocean due to a more strongly negative local feedback (zonal figures in the top row of Figure S19), the change in the SW cloud TOA flux is too negative in this region (lower left panel of Figure S17), suggesting that the SW cloud negative contribution is an error, and is likely the reason for the MR method's underestimate of $\Delta\lambda_{4x}$.

While the exact evolution of temperature patterns in the tropics in AOGCMs may be incorrect due to cold-tongue biases (Seager et al. 2019), our findings match with Dong et al. (2019), in that as long as the feedbacks in tropical convecting regions are far more negative than anywhere else, the delayed warming in regions of ocean heat uptake will ensure an increase in sensitivity over time. Observational evidence suggests that $\overline{N}$ depends on tropical midtropospheric temperatures

(Dessler et al. 2018; Ceppi and Gregory 2019; Fueglistaler 2019), supporting our argument that a reduction in the share of surface warming occurring in the tropical convecting regions which set these temperatures likely influences the Earth's sensitivity.

5. Conclusions

The global climate feedback, one of the key parameters in determining future climate change, is inconstant in part because radiative feedbacks vary spatially. The MR method estimates these spatial feedbacks from records of its internal variability, and improves upon existing methods for doing so by incorporating both local and nonlocal radiative responses to surface warming. For the six atmosphere-ocean general circulation models studied, the spatial feedbacks estimated by the MR method applied to the pattern of surface warming recreate the spatial pattern of top-of-atmosphere flux response to forcing more accurately than existing methods, as well as providing better estimates of the change in feedback with time. The method consistently overestimates the change in TOA flux over the Southern Ocean and north Atlantic, and so overestimates the sensitivity. The method finds that that there are significant negative nonlocal feedbacks associated with regions of tropical convection, and that the reduction in the share of warming that occurs in these regions over time contributes to an increase in the global feedback with time in these models, consistent with recent studies (Andrews and Webb 2017; Ceppi and Gregory 2017; Dong et al. 2019; Fueglistaler 2019).

The MR method finds that five of the six AOGCMs have strongly positive local cloud feedbacks countered by strongly negative nonlocal cloud feedbacks. These positive local feedbacks may explain why studies that use local regressions to estimate spatial feedbacks from observed internal variability find that they are on average positive (Brown et al. 2015; Trenberth et al. 2015). While the AOGCMs exhibit an anti-correlation between local and nonlocal feedbacks, a small relative shift in the balance between these feedbacks could cause large changes in sensitivity, and such shifts may be relevant for paleoclimate or future warming. Given the large magnitudes associated with these local and nonlocal cloud feedbacks, it may be harder for cloudy exoplanets to have stable atmospheres, reducing the chances of finding habitable worlds.

Spatial feedbacks estimated from observations could potentially improve warming forecasts and serve as emerging constraints on AOGCMs. The success of the MR method for most fluxes and regions of the Earth (with the important exception of Southern Ocean cloud feedbacks) suggests that many of the spatial feedbacks at work under global warming are observable under internal variability. Challenges remain to applying the MR method to observations. We would need to reduce the information necessary to fit our statistical model to be less than the length of the satellite record; to remove changes in forcing from records of top-of-atmosphere fluxes; and to account for systematic biases in the observations themselves. We would also need to account for regions of the Earth and states of the climate where the MR method is biased, such as for Southern Ocean cloud feedbacks. Furthermore, since spatial feedbacks are just one link in the coupled energy balance of the climate, we would need complementary theory to complete the forecast of future warming, particularly its spatial pattern. Still, our results suggest that the processes that will determine the sensitivity in both the near and far future may be observable today.

Acknowledgments. This paper was inspired by conversations with Kyle Armour, Cristian Proistosescu, Daniel Koll, and Elizabeth Moyer and was improved further by discussions with B. B. Cael, Yue Dong, Jonathan Gregory, Malte Jansen, Thorsten Mauritsen, Brian Rose, Hansi Singh, M Wu, and Chen Zhou. The research would not have been possible without the efforts of the contributors to the LongRunMIP project. The authors thank Chen Zhou for permission to use the top row in Figure 8. We acknowledge support from the National Science Foundation under NSF award number 1623064. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 786427, project "Couplet"). Our work was completed with resources provided by the University of Chicago Research Computing Center, with special thanks to Hossein Pourreza. We also thank three anonymous reviewers for their insightful comments and our editor, Timothy Delsole, for his guidance throughout the publication process.

APPENDIX

Data and methods

a. Data/code access

For LongRunMIP data access, visit `http://www.longrunmip.org/`. This paper's code is available at `https://github.com/jsbj/spatial`.

b. Matrix and vector notation

Note that in the main body of the text, time is treated as continuous, so that time-series are written as functions (e.g., $\vec{T}(t)$ is the evolving spatial pattern of warming). Since the Appendix documents the calculations we have employed, it treats time as discrete, and so time is instead treated as an additional dimension (e.g., $\mathbf{T}$ is the evolving spatial pattern of warming). Therefore, a vector in the main body of the text refers to a spatial pattern, while a vector in the Appendix refers to a time-series of a scalar value (such as a global average).

c. Conceptual model

The conceptual model is a system of stochastic differential equations:

$$
\begin{aligned}
c_1 \frac{dT_1'}{dt} &= N_1' - H' + F_{surf,1} \\
c_2 \frac{dT_2'}{dt} &= N_2' + H' + F_{surf,2}
\end{aligned}
$$

where $H' = \gamma(T_1' - T_2')$ and

$$N_1' = \lambda_{1,1}T_1' + \lambda_{1,2}T_2' + F_{CO_2,1} + F_{TOA,1} \tag{A1}$$

$$N_2' = \lambda_{2,1}T_1' + \lambda_{2,2}T_2' + F_{CO_2,2} + F_{TOA,2} \tag{A2}$$

The thermal inertia c_i is defined as $m_i \rho c_p$, where ρ and c_p are the density and specific heat of ocean water respectively, and m_i is an equivalent mixed layer depth; m_1 is 50m, and m_2 is 1000m. $F_{CO_2,1} = F_{CO_2,2}$ are both 0 Wm^{-2} (8 Wm^{-2}) for the control (abrupt4x) simulation. $\lambda_{1,1} = 0.5$ Wm^{-2}K^{-1}, $\lambda_{2,1} = -2$ Wm^{-2}K^{-1}, $\lambda_{1,2} = \lambda_{2,2} = 0$Wm^{-2}K^{-1}, and $\gamma = 2$ Wm^{-2}K^{-1}. The terms $\vec{F}_{surf}$ and $\vec{F}_{TOA}$ are white noise processes. In the example shown in Figure 2, the variance of $F_{surf,1}$ and $F_{surf,2}$ is 40 Wm^{-2} and the variance of $F_{TOA,1}$ and $F_{TOA,2}$ is 5 Wm^{-2}, while for the example in Figure S1, the variance of $F_{surf,1}$ and $F_{surf,2}$ is 10 Wm^{-2} and the variance of $F_{TOA,1}$ and $F_{TOA,2}$ is 15 Wm^{-2}.

d. The multiple regression method

Suppose that we have a time series of surface temperatures and TOA radiative fluxes of the Earth, real or simulated, where the surface of the Earth is regridded into n_{grid} (288) regions, and where we have n_{time} years of monthly observations. For each season s ($1 \leq s \leq 4$), we can define an $n_{time} \times n_{grid}$ matrix $\mathbf{T}_m$, where the element in row i and column j, $T_{i,j,s}$, is the surface temperature in region j during season s of year i. We can also define a matrix of anomalies, $\mathbf{T}'_s$, where

$$
\begin{aligned}
\mathbf{T}'_s \;=\; & \begin{bmatrix} T_{1,1,s} & T_{1,2,s} & \cdots & T_{1,n_{grid},s} \\ T_{2,1,s} & T_{2,2,s} & \cdots & T_{2,n_{grid},s} \\ \vdots & \vdots & \ddots & \vdots \\ T_{n_{time},1,s} & T_{n_{time},2,s} & \cdots & T_{n_{time},n_{grid},s} \end{bmatrix} \\
& - \frac{1}{n_{time}} \begin{bmatrix} \sum_{i=1}^{n_{time}} T_{i,1,s} & \sum_{i=1}^{n_{time}} T_{i,2,s} & \cdots & \sum_{i=1}^{n_{time}} T_{i,n_{grid},s} \\ \sum_{i=1}^{n_{time}} T_{i,1,s} & \sum_{i=1}^{n_{time}} T_{i,2,s} & \cdots & \sum_{i=1}^{n_{time}} T_{i,n_{grid},s} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n_{time}} T_{i,1,s} & \sum_{i=1}^{n_{time}} T_{i,2,s} & \cdots & \sum_{i=1}^{n_{time}} T_{i,n_{grid},s} \end{bmatrix}
\end{aligned}
$$

To estimate the spatial feedbacks associated with a TOA radiative flux of type f (where f is either net, $LW\ clear$, $SW\ clear$, $LW\ cloud$, or $SW\ cloud$) and season s, we first define an $n_{time} \times n_{grid}$ matrix of anomalies $\mathbf{R}'_{f,s}$, which is analogous to $\mathbf{T}'_s$ above (N from the main body of the text is R_{net}). We can fit the statistical model defined in Equation 9 using least squares to solve for seasonal spatial feedbacks ($\Lambda_{f,s}$):

$$
\Lambda_{f,s} = \begin{bmatrix} \lambda_{f,1,1} & \lambda_{f,1,2} & \cdots & \lambda_{f,1,n_{grid}} \\ \lambda_{f,2,1} & \lambda_{f,2,2} & \cdots & \lambda_{f,2,n_{grid}} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{f,n_{grid},1} & \lambda_{f,n_{grid},2} & \cdots & \lambda_{f,n_{grid},n_{grid}} \end{bmatrix} = (\mathbf{T}_s'^T \mathbf{T}_s')^{-1} \mathbf{T}_s'^T \mathbf{R}'_{f,s} \qquad \text{(A3)}
$$

Seasonal feedbacks are used in Section 3, but Section 2 uses an annual version, in which case instead of a set of four seasonal feedback matrices, only one feedback matrix estimated using the above Equation d, with the difference that the time series are annual averages. The “monthly” approach in Section 1.2.1 of the SI is the same as the seasonal approach in Equation d, except instead of a four regressions, twelve are performed, with all time series being monthly averages sampled every twelve months. The "all months" approach instead performs only one regression, just like the annual approach, except that monthly average time series are used instead of annual averages (the logic being that even though months may have different properties, there may be an advantage in maximizing the data available to fit a regression).

e. Estimating the forced response

1) FORCED FEEDBACKS

Suppose that we have a $n_{time,abrupt4x}$-year long abrupt4x simulation of a GCM for which we have spatial feedbacks estimated from a control run. We then define an early period (years 2 to 20) and a late period (years 21 to $n_{time,abrupt4x}$). The true feedbacks $\lambda_{f,p}$ for the abrupt4x simulation during each period p (where p is *early* or *late*) are defined as the slope of the least squares fit of the linear regression of the time series of globally averaged TOA flux anomalies of type f from the abrupt4x simulation ($\vec{R}'_{f,abrupt4x}$), against the globally averaged surface temperature anomalies from the abrupt4x simulation $\vec{T}'_{abrupt4x}$:

$$\lambda_{abrupt4x,f,p} = \frac{\{\vec{T}'_{abrupt4x}\}_p \cdot \{\vec{R}'_{f,abrupt4x}\}_p}{||\{\vec{T}'_{abrupt4x}\}_p||^2} \tag{A4}$$

where the curly brackets denote that the time series are averaged over exponentially longer periods, with annual averages for the first decade increasing to centennial averages by the simulation's end, and the p subscript denotes whether values from before or after year 20 are used. $\vec{R}'_{f,abrupt4x}$ and $\vec{T}'_{abrupt4x}$ are vectors with as many entries as years in the abrupt4x simulation (1000 years).

We can make estimates of these feedbacks using the MR method by first estimating the abrupt4x simulation's TOA radiative flux of type f for each month of the year m by multiplying the surface temperature time series of that abrupt4x simulation for that month, $\mathbf{T}'_{m,abrupt4x}$ (a $n_{time,abrupt4x} \times n_{grid}$ matrix) by the spatial feedbacks for that month's season:

$$\hat{\mathbf{R}}'_{f,m,abrupt4x} = \mathbf{T}'_{m,abrupt4x}\Lambda_{f,s(m)} \tag{A5}$$

We use months instead of seasonal averages because our seasons do not start in January, and this approach allows us to have annual averages that start in January. These monthly time series $\hat{\mathbf{R}}'_{f,m,abrupt4x}$ can then be turned into annual averages $\hat{\mathbf{R}}'_{f,abrupt4x}$, and then global averages $\hat{\vec{R}}'_{f,abrupt4x}$, allowing us to estimate the feedbacks for period p by performing the same least squares fit as above:

$$\hat{\lambda}_{abrupt4x,f,p} = \frac{\{\vec{T}'_{abrupt4x,p}\} \cdot \{\hat{\vec{R}}'_{f,abrupt4x,p}\}}{\|\{\vec{T}'_{abrupt4x,p}\}\|^2} \tag{A6}$$

2) Spatial patterns of TOA flux change

We quantify the normalized spatial pattern of TOA radiative flux change of flux type f across a period p by taking a finite difference approach, taking the mean value of $\vec{R}'_{f,abrupt4x}$ during two parts of the period and subtracting the first part from the second (where the divisions for the *early* period are years 2-6 and 7-20, and the divisions for the *late* period are 21-170 and 171-$n_{time,abrupt4x}$, with both divisions chosen to allow for substantial warming in each period), and then dividing this by the average change in the globally averaged surface temperature between these two periods:

$$\Delta\vec{R}'_{f,abrupt4x,p} = \frac{\left(\sum_{i=t_{mid,p}+1}^{t_{end,p}} \begin{bmatrix} R'_{f,abrupt4x,i,1} \\ R'_{f,abrupt4x,i,2} \\ \vdots \\ R'_{f,abrupt4x,n_{grid}} \end{bmatrix} - \sum_{i=t_{start,p}}^{t_{mid,p}} \begin{bmatrix} R'_{f,abrupt4x,i,1} \\ R'_{f,abrupt4x,i,2} \\ \vdots \\ R'_{f,abrupt4x,n_{grid}} \end{bmatrix}\right)}{\left(\sum_{i=t_{mid,p}+1}^{t_{end,p}} T_{abrupt4x,i} - \sum_{i=t_{start,p}}^{t_{mid,p}} T_{abrupt4x,i}\right)} \tag{A7}$$

where $t_{start,p}$ and $t_{end,p}$ are the first and last years in period p, respectively, where $t_{mid,p}$ is 6 for the early period and 170 for late period, where $R'_{f,abrupt4x,i,j}$ is the element in the i^{th} row and j^{th} column of $\mathbf{R'_{f,abrupt4x}}$, and where $T_{abrupt4x,i}$ is the i^{th} element in $\vec{T}_{abrupt4x}$. Finite difference is used instead of regressing values against a global average because the presence of local and nonlocal feedbacks causes nonlinear relationships between $N'_i(t)$ and $T'_i(t)$ (or $\overline{T}'(t)$), which would lead to biased estimates of change from a linear regression.

f. Errors

We calculate two types of errors: feedback errors (Tables 1 and S2), and spatial errors (Tables 2 and S3). We add a subscript g to our feedbacks and spatial patterns of TOA flux change to signify that they belong to the GCM g, where g is one of CCSM3, CESM104, GISSE2R, HadCM3L, IPSLCM5A, and MPIESM12. The feedback error is given by the root mean square error:

$$\varepsilon_{feedback,f,p} = \sqrt{\frac{1}{n_{GCMs}} \sum_{g \in GCMs} (\hat{\lambda}_{f,abrupt4x,p,g} - \lambda_{f,abrupt4x,p,g})^2} \tag{A8}$$

where n_{GCMs} is 6, the number of AOGCMs. The spatial error is measured by taking the area-weighted root mean square error of the spatial estimate

$$\varepsilon_{spatial,f,p} = \sqrt{\frac{\sum_{i=1}^{n_{grid}} (\Delta\hat{\vec{\tilde{R}}}'_{f,abrupt4x,p,i} - \Delta\vec{\tilde{R}}'_{f,abrupt4x,p,i})^2 a_i}{\sum_{i=1}^{n_{grid}} a_i}} \tag{A9}$$

where a_i is the area of the i^{th} grid cell. For the spatial errors in the main body of the paper, this is taken on multi-model mean values of $\Delta\hat{\vec{\tilde{R}}}'_{f,abrupt4x,p,i}$ and $\Delta\vec{\tilde{R}}'_{f,abrupt4x,p,i}$. For the same calculation for individual models (Table S4 and Figures S6-S8 in the supplementary materials), values for each model are used instead.

g. *Other methods to calculate feedbacks*

We consider two other methods for deriving spatial feedbacks, estimating abrupt4x feedbacks, and estimating spatial patterns of TOA flux change:

1) The global method

The seasonal version of the "global" method used in the main body of the paper is estimated using the least squares fit on this regression:

$$\lambda_{global,f,s} = \frac{\vec{T}_s' \cdot \vec{R}_{f,s}'}{||\vec{T}_s'||^2} \tag{A10}$$

where $\vec{T}_s$ and $\vec{R}_{f,s}$ are globally and seasonally averaged time series of control simulation surface temperature and TOA flux f respectively, sampled every fourth seasonal value so that all elements of the time series are from season s. The four seasonal feedbacks are used to recreate estimates of the global averaged time series $\vec{R}_{f,abrupt4x}$, which in turn is used, as above, to estimate abrupt4x feedbacks. Once more, different averaging of the control time series and groupings of regression equations can be used to make the annual, monthly, and all months versions of this method featured in Tables S3 and S4.

The normalized spatial pattern of TOA flux change can be found by first estimating the "local contribution" (Boer and Yu 2003a,b; Crook et al. 2011; Zelinka et al. 2012; Andrews et al. 2015), using Equation 1, but replacing the time series vector $\vec{R}_{f,s}'$ with the spatial time series matrix $\mathbf{R}_{f,s}'$ from above, and replacing the single feedback $\lambda_{global,f,s}$ with the spatial vector of feedbacks, $\vec{\lambda}_{global,f}$.

2) The local method

The "local" method assumes the statistical model

$$R_i'(t) = \lambda_{local,i} T_i'(t) + \varepsilon(t) \text{ for each region } i \tag{A11}$$

Spatial feedbacks are estimated using least squares:

$$\vec{\lambda}_{local,f} = \begin{bmatrix} \lambda_{local,f,1} \\ \lambda_{local,f,2} \\ \vdots \\ \lambda_{local,f,n_{grid}} \end{bmatrix} = \begin{bmatrix} \frac{\vec{T}_1'\cdot\vec{R}_{f,1}'}{\|\vec{T}_1'\|^2} \\ \frac{\vec{T}_2'\cdot\vec{R}_{f,2}'}{\|\vec{T}_2'\|^2} \\ \vdots \\ \frac{\vec{T}_{n_{grid}}'\cdot\vec{R}_{f,n_{grid}}'}{\|\vec{T}_{n_{grid}}'\|^2} \end{bmatrix} \tag{A12}$$

where $\vec{T}_i'$ and $\vec{R}_{f,i}'$ are the i^{th} rows of $\mathbf{T}'$ and $\mathbf{R}_f'$ respectively. We can then generate estimates of $\mathbf{R}_{f,abrupt4x}'$ as above. We apply these estimates to Equations A6 and A7 to estimate forced global feedbacks and spatial patterns of TOA flux change.

h. Local regression

We use LOESS (LOcally Estimated Scatterplot Smoothing; Cleveland and Devlin 1988) to take local regression of scatterplots of $\overline{N}$ vs $\overline{T}'$. LOESS uses a weighted regression of a certain number of nearest neighbors, in our case 30. Full details can be found in the code for this paper listed above and in the LocallyWeightedRegression.jl Julia package (`https://github.com/juliohm/LocallyWeightedRegression.jl`).

References

Andrews, T., J. M. Gregory, and M. J. Webb, 2015: The Dependence of Radiative Forcing and Feedback on Evolving Patterns of Surface Temperature Change in Climate Models. *Journal of Climate*, **28 (4)**, 1630–1648.

Andrews, T., and M. J. Webb, 2017: The Dependence of Global Cloud and Lapse Rate Feedbacks on the Spatial Structure of Tropical Pacific Warming. *Journal of Climate*, **31 (2)**, 641–654.

Andrews, T., and Coauthors, 2018: Accounting for Changing Temperature Patterns Increases Historical Estimates of Climate Sensitivity. *Geophysical Research Letters*, **45 (16)**, 8490–8499, doi:10.1029/2018GL078887, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL078887.

Armour, K. C., 2017: Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks. *Nature Climate Change*, **7 (5)**, 331.

Armour, K. C., C. M. Bitz, and G. H. Roe, 2012: Time-Varying Climate Sensitivity from Regional Feedbacks. *Journal of Climate*, **26 (13)**, 4518–4534.

Arrhenius, P. S., 1896: XXXI. On the influence of carbonic acid in the air upon the temperature of the ground. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **41 (251)**, 237–276, doi:10.1080/14786449608620846.

Bloch-Johnson, J., R. T. Pierrehumbert, and D. S. Abbot, 2015: Feedback temperature dependence determines the risk of high warming. *Geophysical Research Letters*, **42 (12)**, 2015GL064 240.

Boer, G., and B. Yu, 2003a: Climate sensitivity and response. *Climate Dynamics*, **20 (4)**, 415–429.

Boer, G. J., and B. Yu, 2003b: Climate sensitivity and climate state. *Climate Dynamics*, **21 (2)**, 167–176.

Brown, P. T., W. Li, J. H. Jiang, and H. Su, 2015: Unforced Surface Air Temperature Variability and Its Contrasting Relationship with the Anomalous TOA Energy Flux at Local and Global Spatial Scales. *Journal of Climate*, **29 (3)**, 925–940.

Ceppi, P., and J. M. Gregory, 2017: Relationship of tropospheric stability to climate sensitivity and Earth's observed radiation budget. *Proceedings of the National Academy of Sciences*, **114 (50)**, 13 126–13 131.

Ceppi, P., and J. M. Gregory, 2019: A refined model for the Earths global energy balance. *Climate Dynamics*, **53 (7)**, 4781–4797, doi:10.1007/s00382-019-04825-x, URL https://doi.org/10.1007/s00382-019-04825-x.

Charney, J. G., and Coauthors, 1979: *Carbon Dioxide and Climate: A Scientific Assessment*. doi:10.17226/12181.

Choi, Y.-S., H. Cho, C.-H. Ho, R. S. Lindzen, S. K. Park, and X. Yu, 2014: Influence of non-feedback variations of radiation on the determination of climate feedback. *Theoretical and Applied Climatology*, **115 (1)**, 355–364, doi:10.1007/s00704-013-0998-6, URL https://doi.org/10.1007/s00704-013-0998-6.

Cleveland, W. S., and S. J. Devlin, 1988: Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association*, **83 (403)**, 596–610, doi:10.2307/2289282, URL https://www.jstor.org/stable/2289282.

Colman, R., and L. Hanson, 2017: On the relative strength of radiative feedbacks under climate variability and change. *Climate Dynamics*, **49 (5-6)**, 2115–2129.

Cox, P. M., C. Huntingford, and M. S. Williamson, 2018: Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature*, **553 (7688)**, 319–322, doi:10.1038/nature25450, URL https://www.nature.com/articles/nature25450.

Crook, J. A., P. M. Forster, and N. Stuber, 2011: Spatial Patterns of Modeled Climate Feedback and Contributions to Temperature Response and Polar Amplification. *Journal of Climate*, **24 (14)**, 3575–3592.

Dessler, A. E., 2010: A Determination of the Cloud Feedback from Climate Variations over the Past Decade. *Science*, **330 (6010)**, 1523–1527.

Dessler, A. E., 2012: Observations of Climate Feedbacks over 2000-10 and Comparisons to Climate Models. *Journal of Climate*, **26 (1)**, 333–342.

Dessler, A. E., T. Mauritsen, and B. Stevens, 2018: The influence of internal variability on Earth's energy balance framework and implications for estimating climate sensitivity. *Atmospheric Chemistry and Physics*, **18 (7)**, 5147–5155, doi:https://doi.org/10.5194/acp-18-5147-2018, URL https://www.atmos-chem-phys.net/18/5147/2018/.

Dong, Y., C. Proistosescu, K. C. Armour, D. S. Battisti, Y. Dong, C. Proistosescu, K. C. Armour, and D. S. Battisti, 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*, doi:10.1175/JCLI-D-18-0843.1, URL https://journals.ametsoc.org/doi/abs/10.1175/JCLI-D-18-0843.1.

Feldl, N., and G. H. Roe, 2013a: Four perspectives on climate feedbacks. *Geophysical Research Letters*, **40 (15)**, 4007–4011.

Feldl, N., and G. H. Roe, 2013b: The Nonlinear and Nonlocal Nature of Climate Feedbacks. *Journal of Climate*, **26 (21)**, 8289–8304, doi:10.1175/JCLI-D-12-00631.1.

Forster, P. M. F., and J. M. Gregory, 2006: The Climate Sensitivity and Its Components Diagnosed from Earth Radiation Budget Data. *Journal of Climate*, **19 (1)**, 39–52.

Fueglistaler, S., 2019: Observational Evidence for Two Modes of Coupling Between Sea Surface Temperatures, Tropospheric Temperature Profile, and Shortwave Cloud Radiative Effect in the Tropics. *Geophysical Research Letters*, **46 (16)**, 9890–9898, doi:10.1029/2019GL083990, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL083990.

Gregory, J., and M. Webb, 2008: Tropospheric Adjustment Induces a Cloud Component in CO2 Forcing. *Journal of Climate*, **21 (1)**, 58–71, doi:10.1175/2007JCLI1834.1.

Gregory, J. M., and T. Andrews, 2016: Variation in climate sensitivity and feedback parameters during the historical period. *Geophysical Research Letters*, **43 (8)**, 3911–3920, doi:10.1002/2016GL068406, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL068406.

Gregory, J. M., R. J. Stouffer, S. C. B. Raper, P. A. Stott, and N. A. Rayner, 2002: An Observationally Based Estimate of the Climate Sensitivity. *Journal of Climate*, **15 (22)**, 3117–3121, doi:10.1175/1520-0442(2002)015⟨3117:AOBEOT⟩2.0.CO;2.

Jiménez-de-la Cuesta, D., and T. Mauritsen, 2019: Emergent constraints on Earths transient and equilibrium response to doubled CO 2 from post-1970s global warming. *Nature Geoscience*, **12 (11)**, 902–905, doi:10.1038/s41561-019-0463-y, URL https://www.nature.com/articles/s41561-019-0463-y.

Jonko, A. K., K. M. Shell, B. M. Sanderson, and G. Danabasoglu, 2012: Climate Feedbacks in CCSM3 under Changing CO2 Forcing. Part II: Variation of Climate Feedbacks and Sensitivity with Forcing. *Journal of Climate*, **26 (9)**, 2784–2795, doi:10.1175/JCLI-D-12-00479.1.

Klein, S. A., A. Hall, J. R. Norris, and R. Pincus, 2017: Low-Cloud Feedbacks from Cloud-Controlling Factors: A Review. *Surveys in Geophysics*, **38 (6)**, 1307–1329.

Klein, S. A., and D. L. Hartmann, 1993: The Seasonal Cycle of Low Stratiform Clouds. *Journal of Climate*, **6 (8)**, 1587–1606.

Klein, S. A., B. J. Soden, and N.-C. Lau, 1999: Remote Sea Surface Temperature Variations during ENSO: Evidence for a Tropical Atmospheric Bridge. *Journal of Climate*, **12 (4)**, 917–932, doi:

10.1175/1520-0442(1999)012⟨0917:RSSTVD⟩2.0.CO;2, URL https://journals.ametsoc.org/doi/full/10.1175/1520-0442%281999%29012%3C0917%3ARSSTVD%3E2.0.CO%3B2.

Koll, D. D. B., and T. W. Cronin, 2018: Earth's outgoing longwave radiation linear due to H2o greenhouse effect. *Proceedings of the National Academy of Sciences*, **115 (41)**, 10 293–10 298, doi:10.1073/pnas.1809868115, URL https://www.pnas.org/content/115/41/10293.

Leconte, J., F. Forget, B. Charnay, R. Wordsworth, and A. Pottier, 2013: Increased insolation threshold for runaway greenhouse processes on Earth-like planets. *Nature*, **504 (7479)**, 268–271, doi:10.1038/nature12827.

Lewis, N., and J. Curry, 2018: The Impact of Recent Forcing and Ocean Heat Uptake Data on Estimates of Climate Sensitivity. *Journal of Climate*, **31 (15)**, 6051–6071, doi:10.1175/JCLI-D-17-0667.1.

Lewis, N., and J. A. Curry, 2015: The implications for climate sensitivity of AR5 forcing and heat uptake estimates. *Climate Dynamics*, **45 (3-4)**, 1009–1023.

Li, W., and C. E. Forest, 2014: Estimating the Sensitivity of the Atmospheric Teleconnection Patterns to SST Anomalies Using a Linear Statistical Method. *Journal of Climate*, **27 (24)**, 9065–9081, doi:10.1175/JCLI-D-14-00231.1, URL https://journals.ametsoc.org/doi/full/10.1175/JCLI-D-14-00231.1.

Li, W., C. E. Forest, and J. Barsugli, 2012: Comparing two methods to estimate the sensitivity of regional climate simulations to tropical SST anomalies. *Journal of Geophysical Research: Atmospheres*, **117 (D20)**, doi:10.1029/2011JD017186, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JD017186.

Libardoni, A. G., C. E. Forest, A. P. Sokolov, and E. Monier, 2019: Underestimating Internal Variability Leads to Narrow Estimates of Climate System Properties. *Geophysical Research Letters*, **46 (16)**, 10 000–10 007, doi:10.1029/2019GL082442, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL082442.

Lindzen, R. S., M.-D. Chou, and A. Y. Hou, 2001: Does the Earth Have an Adaptive Infrared Iris? *Bulletin of the American Meteorological Society*, **82 (3)**, 417–432, doi:10.1175/1520-0477(2001)082⟨0417:DTEHAA⟩2.3.CO;2, URL https://journals.ametsoc.org/doi/abs/10.1175/1520-0477%282001%29082%3C0417%3ADTEHAA%3E2.3.CO%3B2.

Liu, F., J. Lu, O. Garuba, L. R. Leung, Y. Luo, and X. Wan, 2018: Sensitivity of Surface Temperature to Oceanic Forcing via q-Flux Greens Function Experiments. Part I: Linear Response Function. *Journal of Climate*, **31 (9)**, 3625–3641, doi:10.1175/JCLI-D-17-0462.1, URL https://journals.ametsoc.org/doi/10.1175/JCLI-D-17-0462.1.

Liu, Z., N. Wen, and Y. Liu, 2008: On the Assessment of Nonlocal Climate Feedback. Part I: The Generalized Equilibrium Feedback Assessment. *Journal of Climate*, **21 (1)**, 134–148, doi:10.1175/2007JCLI1826.1, URL http://journals.ametsoc.org/doi/abs/10.1175/2007JCLI1826.1.

Lutsko, N. J., and M. Popp, 2019: Probing the Sources of Uncertainty in Transient Warming on Different Timescales. *Geophysical Research Letters*, **46 (20)**, 11 367–11 377, doi:10.1029/2019GL084018, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL084018.

Lutsko, N. J., and K. Takahashi, 2018: What Can the Internal Variability of CMIP5 Models Tell Us about Their Climate Sensitivity? *Journal of Climate*, **31 (13)**, 5051–5069, doi:10.1175/JCLI-D-17-0736.1.

Mauritsen, T., and B. Stevens, 2015: Missing iris effect as a possible cause of muted hydrological change and high climate sensitivity in models. *Nature Geoscience*, **8 (5)**, 346–351, doi:10.1038/ngeo2414, URL https://www.nature.com/articles/ngeo2414.

Meraner, K., T. Mauritsen, and A. Voigt, 2013: Robust increase in equilibrium climate sensitivity under global warming. *Geophysical Research Letters*, **40 (22)**, 5944–5948.

Murphy, D. M., S. Solomon, R. W. Portmann, K. H. Rosenlof, P. M. Forster, and T. Wong, 2009: An observationally based energy balance for the Earth since 1950. *Journal of Geophysical Research: Atmospheres*, **114 (D17)**.

Murphy, J. M., 1995: Transient Response of the Hadley Centre Coupled Ocean-Atmosphere Model to Increasing Carbon Dioxide. Part 1: Control Climate and Flux Adjustment. *Journal of Climate*, **8 (1)**, 36–56.

Otto, A., and Coauthors, 2013: Energy budget constraints on climate response. *Nature Geoscience*, **6 (6)**, 415.

Po-Chedley, S., K. C. Armour, C. M. Bitz, M. D. Zelinka, B. D. Santer, and Q. Fu, 2018: Sources of intermodel spread in the lapse rate and water vapor feedbacks. *Journal of Climate*.

Proistosescu, C., A. Donohoe, K. C. Armour, G. H. Roe, M. F. Stuecker, and C. M. Bitz, 2018: Radiative Feedbacks From Stochastic Variability in Surface Temperature and Radiative Imbalance. *Geophysical Research Letters*, **45 (10)**, 5082–5094, doi:10.1029/2018GL077678.

Proistosescu, C., and P. J. Huybers, 2017: Slow climate mode reconciles historical and model-based estimates of climate sensitivity. *Science Advances*, **3 (7)**, e1602 821, doi:10.1126/sciadv.1602821.

Roe, G. H., and M. B. Baker, 2007: Why Is Climate Sensitivity So Unpredictable? *Science*, **318 (5850)**, 629–632, doi:10.1126/science.1144735.

Rose, B. E. J., K. C. Armour, D. S. Battisti, N. Feldl, and D. D. B. Koll, 2014: The dependence of transient climate sensitivity and radiative feedbacks on the spatial pattern of ocean heat uptake. *Geophysical Research Letters*, **41 (3)**, 1071–1078.

Rose, B. E. J., and L. Rayborn, 2016: The Effects of Ocean Heat Uptake on Transient Climate Sensitivity. *Current Climate Change Reports*, **2 (4)**, 190–201.

Rugenstein, M., and Coauthors, 2019: LongRunMIP - motivation and design for a large collection of millennial-length AO-GCM simulations. *Bulletin of the American Meteorological Society*, doi:10.1175/BAMS-D-19-0068.1, URL https://journals.ametsoc.org/doi/abs/10.1175/BAMS-D-19-0068.1.

Rugenstein, M. A. A., K. Caldeira, and R. Knutti, 2016: Dependence of global radiative feedbacks on evolving patterns of surface heat fluxes. *Geophysical Research Letters*, **43 (18)**, 9877–9885, doi:10.1002/2016GL070907.

Seager, R., M. Cane, N. Henderson, D.-E. Lee, R. Abernathey, and H. Zhang, 2019: Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. *Nature Climate Change*, **9 (7)**, 517–522, doi:10.1038/s41558-019-0505-x, URL https://www.nature.com/articles/s41558-019-0505-x.

Senior, C. A., and J. F. B. Mitchell, 2000: The time-dependence of climate sensitivity. *Geophysical Research Letters*, **27 (17)**, 2685–2688.

Soden, B. J., A. J. Broccoli, and R. S. Hemler, 2004: On the Use of Cloud Forcing to Estimate Cloud Feedback. *Journal of Climate*, **17 (19)**, 3661–3665.

Spencer, R. W., and W. D. Braswell, 2008: Potential Biases in Feedback Diagnosis from Observational Data: A Simple Model Demonstration. *Journal of Climate*, **21 (21)**, 5624–5628, doi:10.1175/2008JCLI2253.1, URL https://journals.ametsoc.org/doi/full/10.1175/2008JCLI2253.1.

Spencer, R. W., and W. D. Braswell, 2011: On the Misdiagnosis of Surface Temperature Feedbacks from Variations in Earth's Radiant Energy Balance. *Remote Sensing*, **3 (8)**, 1603–1613, doi:10.3390/rs3081603, URL https://www.mdpi.com/2072-4292/3/8/1603.

Trenberth, K. E., Y. Zhang, J. T. Fasullo, and S. Taguchi, 2015: Climate variability and relationships between top-of-atmosphere radiation and temperatures on Earth. *Journal of Geophysical Research: Atmospheres*, **120 (9)**, 3642–3659, doi:10.1002/2014JD022887.

Watterson, I. G., 2000: Interpretation of Simulated Global Warming Using a Simple Model. *Journal of Climate*, **13 (1)**, 202–215.

Wetherald, R. T., and S. Manabe, 1988: Cloud Feedback Processes in a General Circulation Model. *Journal of the Atmospheric Sciences*, **45 (8)**, 1397–1416, doi:10.1175/1520-0469(1988)045⟨1397:CFPIAG⟩2.0.CO;2.

Winton, M., K. Takahashi, and I. M. Held, 2010: Importance of Ocean Heat Uptake Efficacy to Transient Climate Change. *Journal of Climate*, **23 (9)**, 2333–2344, doi:10.1175/2009JCLI3139.1.

Wood, R., and C. S. Bretherton, 2006: On the Relationship between Stratiform Low Cloud Cover and Lower-Tropospheric Stability. *Journal of Climate*, **19 (24)**, 6425–6432.

Zelinka, M. D., S. A. Klein, and D. L. Hartmann, 2012: Computing and Partitioning Cloud Feedbacks Using Cloud Property Histograms. Part I: Cloud Radiative Kernels. *Journal of Climate*, **25 (11)**, 3715–3735.

Zhou, C., M. D. Zelinka, and S. A. Klein, 2016: Impact of decadal cloud variations on the Earth's energy budget. *Nature Geoscience*, **9 (12)**, 871–874, doi:10.1038/ngeo2828, URL https://www.nature.com/articles/ngeo2828.

Zhou, C., M. D. Zelinka, and S. A. Klein, 2017: Analyzing the dependence of global cloud feedback on the spatial pattern of sea surface temperature change with a Green's function approach. *Journal of Advances in Modeling Earth Systems*, **9 (5)**, 2174–2189.

LIST OF TABLES

TABLE 1. *Feedback errors.* Root mean square errors of estimates of abrupt4x feedbacks ($\lambda_{4x,early}$, $\lambda_{4x,late}$) and their change with time ($\Delta\lambda_{4x}$), for net TOA fluxes and each component flux (in $\mathrm{Wm^{-2}K^{-1}}$) and for the seasonal versions of the three methods presented in Section 2 (see Appendix for details). For annual and monthly values, see Table S2, and for fluxes north of 30°S, see Table S5.

	net			*LW clear*			*SW clear*			*LW cloud*			*SW cloud*		
	MR	global	local	MR	global	local	MR	global	local	MR	global	local	MR	global	local
early	0.69	0.74	2.54	0.08	0.12	0.63	0.18	0.48	1.21	0.13	0.23	0.02	0.45	0.55	1.19
late	0.29	0.26	1.87	0.15	0.21	0.47	0.13	0.52	1.09	0.31	0.35	0.17	0.2	0.6	0.65
change	0.44	0.73	0.78	0.12	0.17	0.22	0.08	0.11	0.18	0.19	0.13	0.17	0.39	0.57	0.64

TABLE 2. *Spatial errors.* The model-mean area-weighted root mean square error of estimates of the warming-normalized change in TOA fluxes during the early and late periods of the abrupt4x simulations, and the change in pattern between these period (see Appendix for details). All values have units of $\mathrm{Wm^{-2}K^{-1}}$. For annual and monthly versions in addition to seasonal, see Table S2, for individual models see Table S4, and for fluxes north of 30°S, see Table S6.

	net			*LW clear*			*SW clear*			*LW cloud*			*SW cloud*		
	MR	global	local	MR	global	local	MR	global	local	MR	global	local	MR	global	local
early	1.02	3.41	2.77	0.33	2.29	1.02	0.7	5.23	2.15	0.82	1.09	0.71	1.05	5.15	2.25
late	0.8	3.08	1.86	0.34	2.27	0.75	0.52	5.13	1.67	1.28	1.21	1.03	1.09	5.1	1.85
change	0.74	1.14	1.26	0.27	0.91	0.42	0.54	0.72	0.83	0.84	0.79	0.79	1.01	1.27	1.37

LIST OF FIGURES

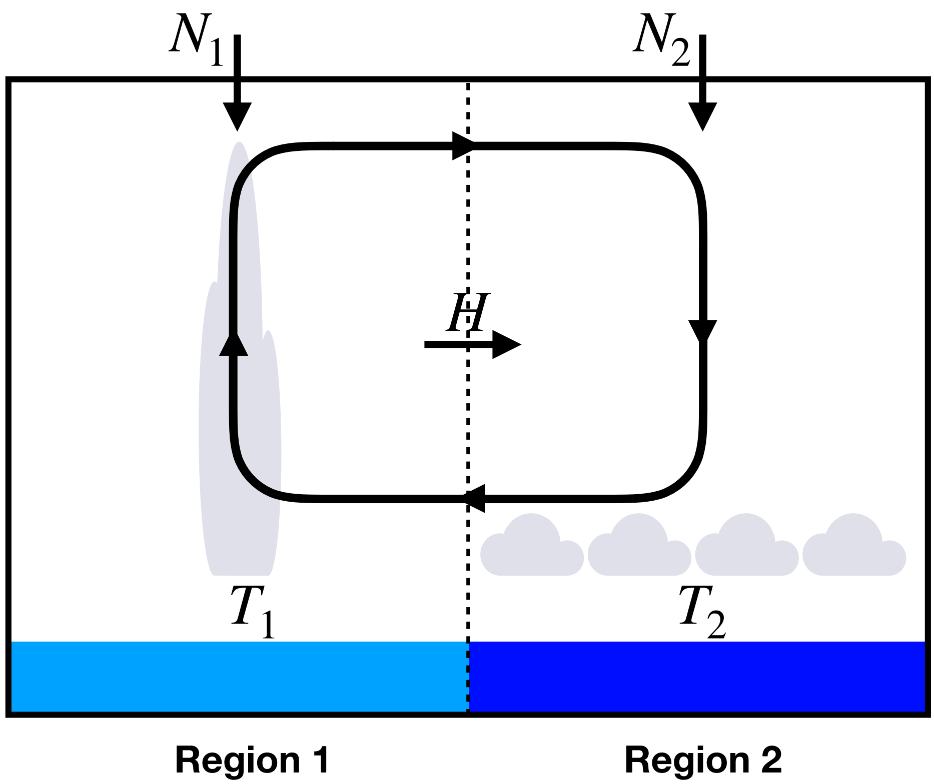

FIG. 1. A schematic representation of the conceptual model used in Section 2, consisting of an overturning cell with a convecting (1) and a subsiding (2) region. Warming of the surface temperature T_1 has nonlocal effects: it increases the horizontal heat transport H, and it changes properties of the atmosphere aloft in region 2 that affect its net top-of-atmosphere radiative flux, N_2, for instance by warming its free troposphere, increasing its lower tropospheric stability, and therefore increasing its low cloud cover. The dependence of N_2 on T_1 (holding T_2 fixed) is an example of a nonlocal radiative feedback.

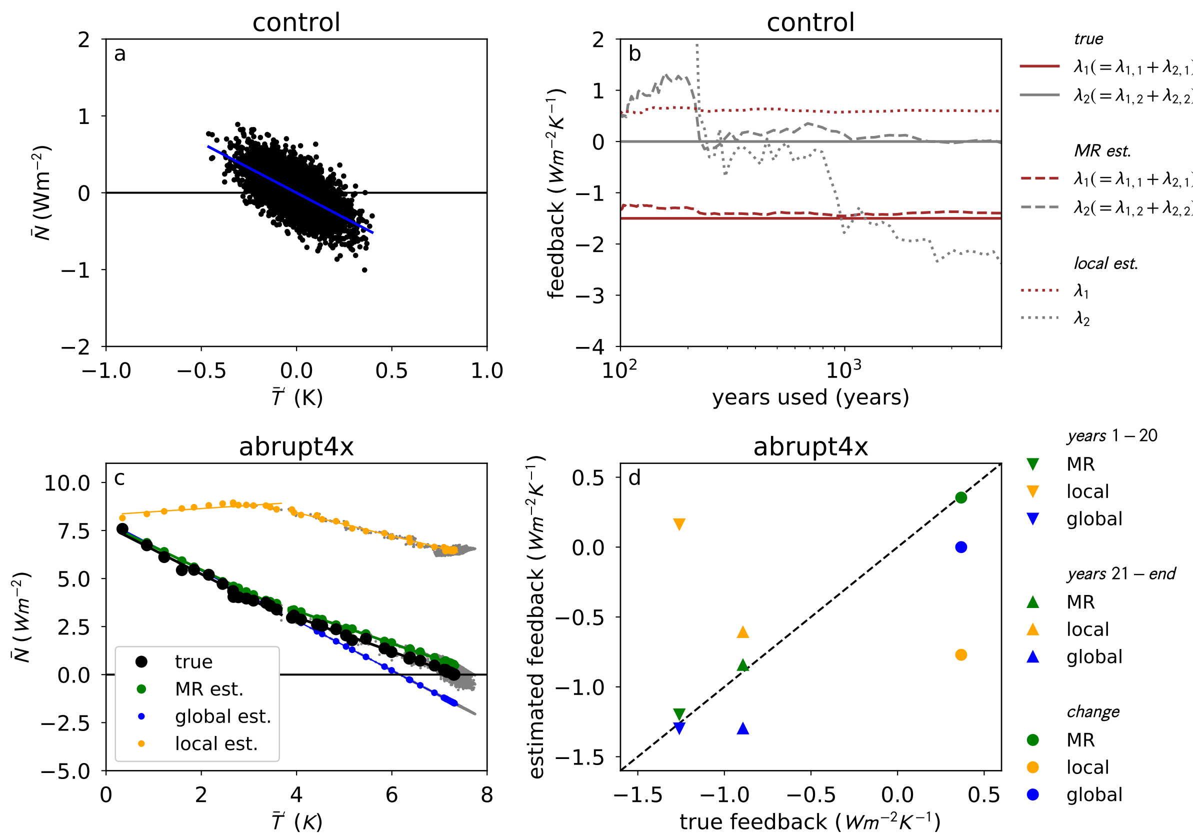

FIG. 2. Two experiments are performed with the conceptual model in Equation 1: an unforced "control" simulation (panels a,b) and a forced "abrupt4x" simulation (panels c,d). Values of $\overline{N}$ vs. $\overline{T}'$ from each experiment are given by the black dots in panels a and c, representing annual averages for the control simulation and exponentially increasing averages for the abrupt4x simulation. The global method assumes that the slope of the regression in panel a (blue line) gives the slope of the black dots in the lower left panel, underestimating the increase in this slope over time (blue lines and markers, panels c,d). The local method regresses N_i' against T_i' to estimate λ_i for both regions (dotted lines, panel b), which leads to an overestimate of the combined feedback associated with region 1 ($\lambda_1 = \lambda_{1,1} + \lambda_{2,1}$, dotted red line in panel b), and therefore an overestimate of the feedback early on (orange lines and markers, panels c,d). The MR method, given sufficient years to regress over, correctly estimates all spatial feedbacks (dashed lines, panel b), accurately predicting the feedbacks and its change with time (green lines and markers, panels c,d).

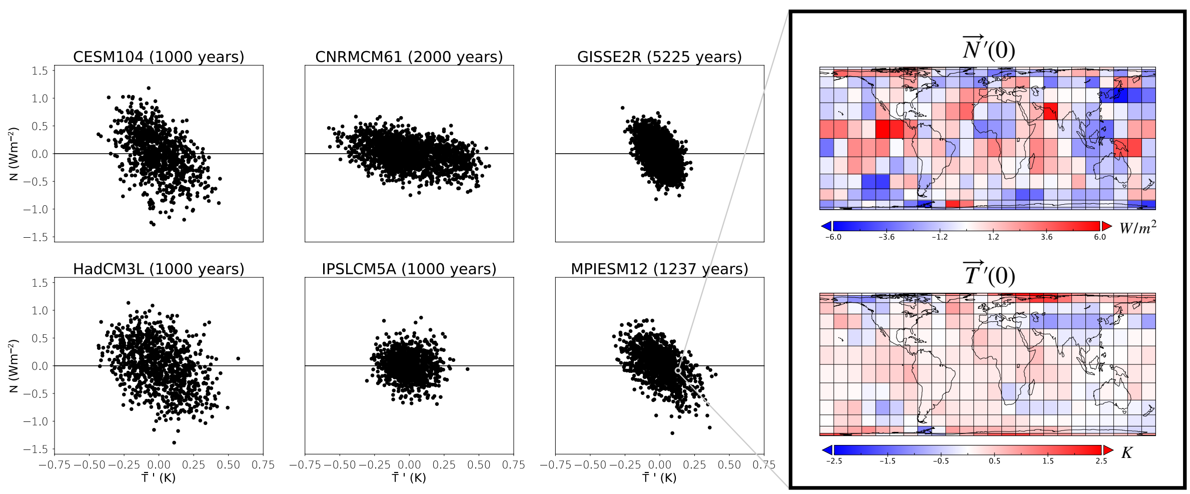

FIG. 3. Plots of $\overline{N}$ vs. $\overline{T}'$ for control simulations of six coupled atmosphere-ocean general circulation models (see Table S1 for details). We use the simulations to estimate spatial feedbacks using the global, local, and MR methods. We regrid simulations to $15^\circ \times 15^\circ$ grids, giving 288 regions.

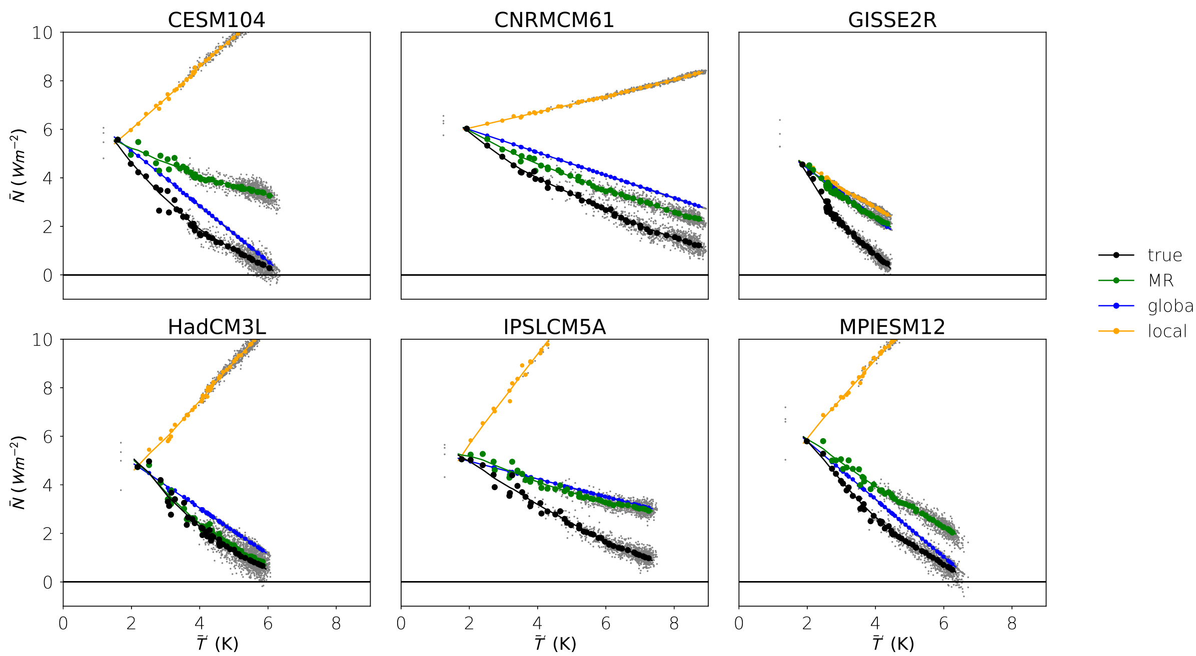

FIG. 4. $\overline{N}$ vs. $\overline{T}'$ for abrupt4x simulations of the same six GCMs from Figure 3 (black dots). Colored dots show estimates of $\overline{N}_{abrupt4x}(t)$ made using the spatial feedbacks inferred from each model's control simulation and its spatial pattern of warming ($\vec{T}'_{abrupt4x}(t)$) using the three methods described in the text; year one is not included in any method. Larger dots represent averages taken over exponentially increasing periods, except gray dots, which show all years. Solid lines show local regressions using LOESS. Global estimates for GISSE2R does not appear because it is nearly identical with MR estimates.

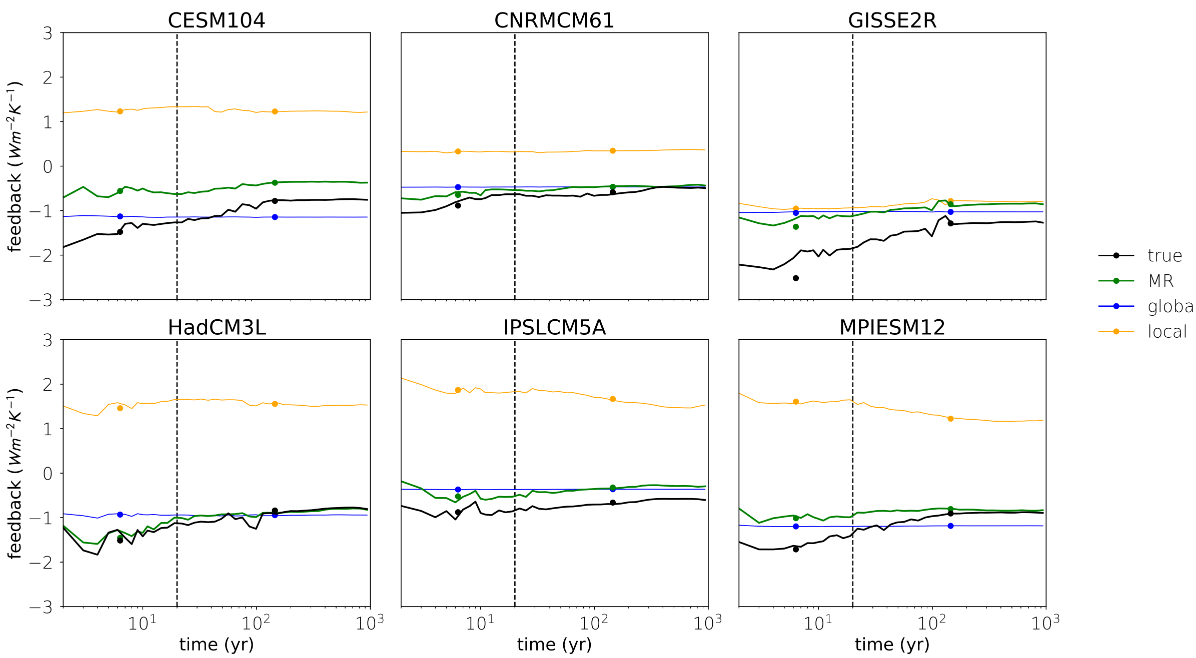

FIG. 5. True and estimated abrupt4x feedbacks as a function of time calculated using slopes of the local regression from Figure 4 (solid lines). Vertical dotted lines show the division between the early (2-20 years) and late (21-end) periods. Dots show true and estimated values of $\lambda_{4x,early}$ and $\lambda_{4x,late}$. Feedbacks get more positive over time for all models. The MR and global methods initially overestimate feedbacks. The MR estimate increases with time as well, while the global method predicts a roughly constant feedback. The local method greatly overestimates the true feedback for all models except GISSE2R. Figures S2-5 give the same plot for component fluxes.

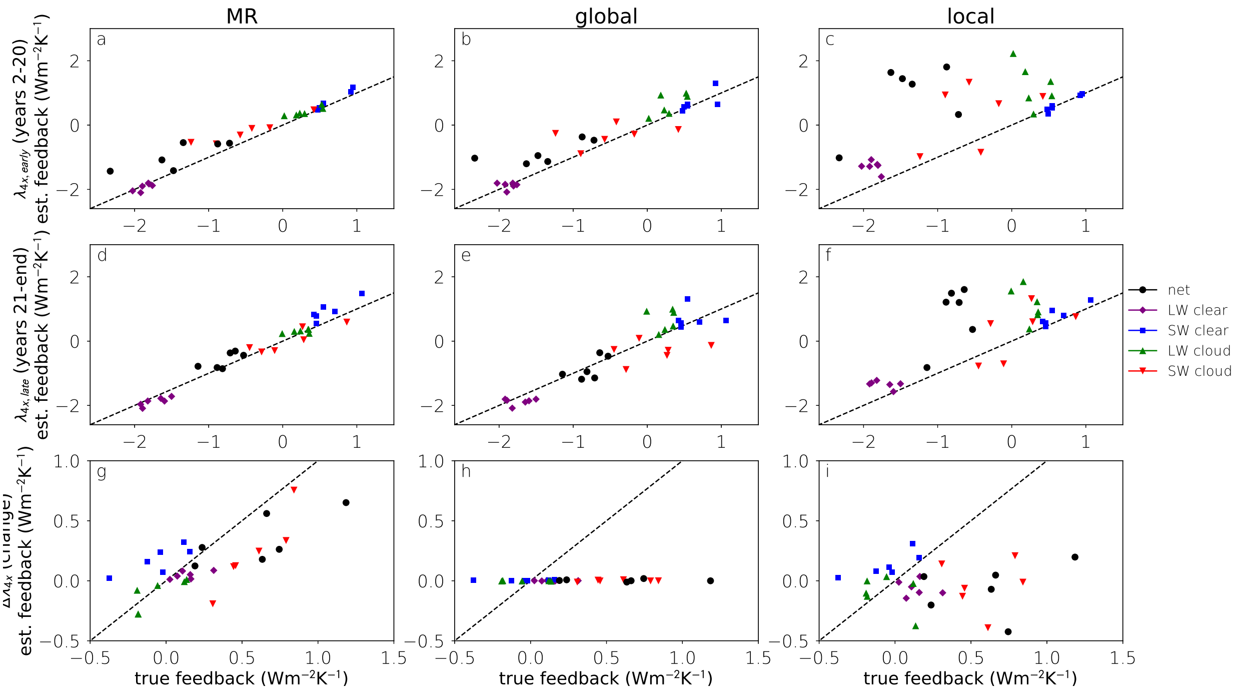

FIG. 6. True vs. estimated feedbacks for the early (panels a, b, and c) and late (panels d, e, and f) periods and the change between them (panels g, h, and i). Black dots give values for the net feedback, while colored markers give values of the component feedbacks, which sum to the net feedback. The MR and global methods overestimate the early feedback due to SW cloud (red) feedbacks. The MR estimate of the late period has a small error across all components (panel d), while the global estimate has a smaller net error due to offsetting errors between LW and SW cloud feedbacks (panel e). As in Figure 5, the MR method is able to capture some of the change in feedback, while the global method does not. The local method greatly overestimates the net feedback, primarily due to cloud feedbacks. Numerical values of the feedbacks are given in Table S7 and S8.

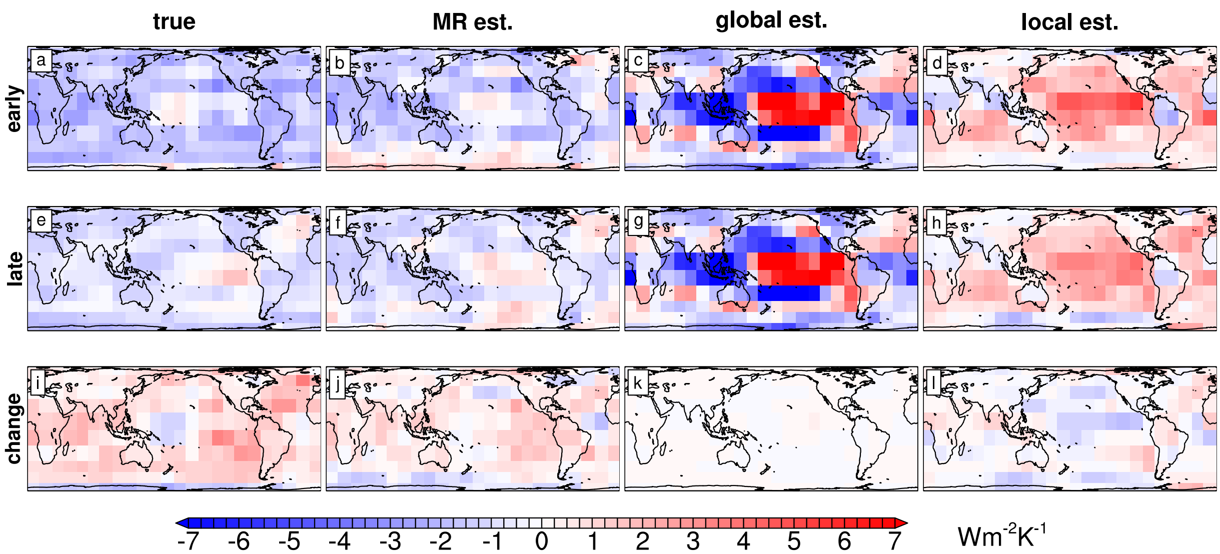

FIG. 7. Multi-model mean spatial pattern of net TOA flux change associated with the early (top row) and late (middle row) periods and the change between them (bottom row), calculated by taking the finite difference across each period. Changes are normalized by the total warming in each period, giving units of $\mathrm{Wm^{-2}K^{-1}}$. The MR method is close to the true pattern except for overestimates south of 30°S and during the early period in the North Atlantic. This holds for individual flux components as well (Figures S9-S17). The global and local methods both have substantial errors over most of the globe. Figures S6-S8 show errors (estimates - true values) for the multi-model mean and individual models.

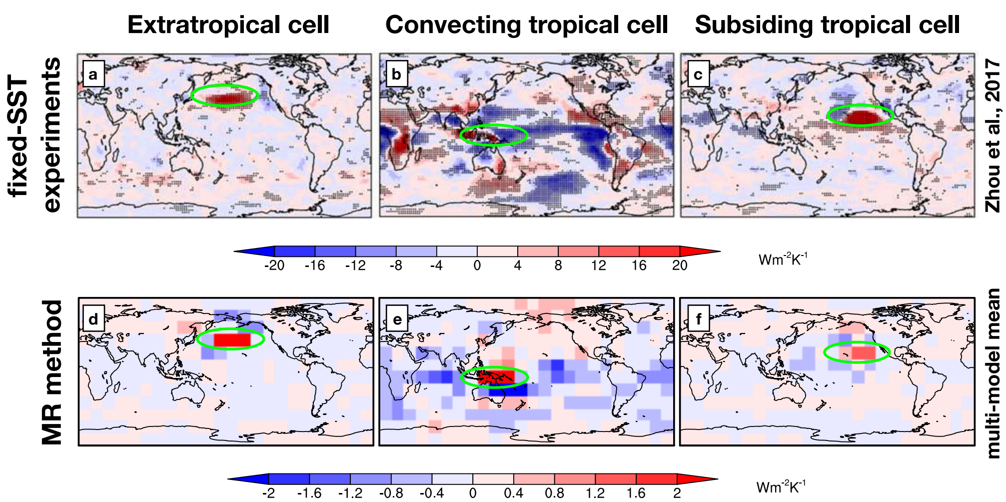

FIG. 8. Net cloud feedbacks associated with warming in regions circled in green estimated for CAM5 by Zhou et al. (2017) using fixed-SST experiments (panels a, b, and c) or as a multi-model and multi-season mean using the MR method (panels d, e, and f). For perturbations outside of tropical convecting regions (panels a, c, d, and f), the effects are mostly local and positive, while perturbations in tropical convecting regions have significant negative nonlocal effects in many regions of the Earth (panels b, e). Note that fixed-SST experiments allow some land warming in response to these perturbations (panel b), while the MR method is agnostic about whether the surface is land or ocean, and so does not include resulting land warming (panel e).

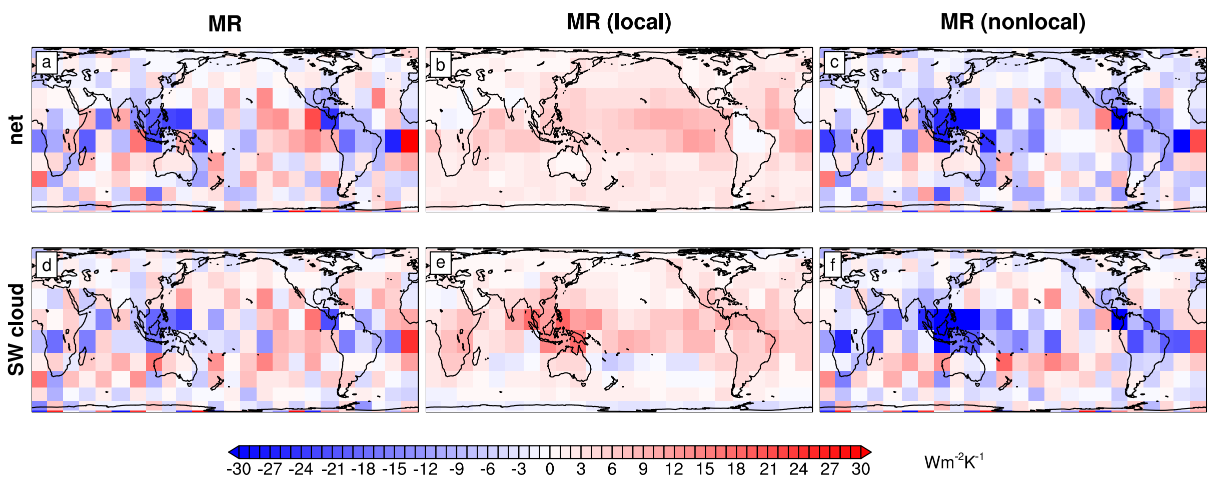

FIG. 9. Multi-model and multi-season mean spatial feedbacks estimated by the MR method. Panel a shows the estimated change in $\overline{N}$ caused by warming a degree in each cell as weighted by the cell's area. This is the sum of local changes in $\vec{N}$ (panel b), which are almost uniformly positive, and nonlocal changes (panel c), which are usually negative, especially in regions of tropical convection. The competing positive local and negative nonlocal components are primarily due to the SW cloud feedback (panels d, e, and f). For maps of all flux components and assessments of uncertainty, see Figures S18-S22. For spatial feedbacks of all methods, see Figure S23. Compare with estimates of spatial feedbacks for CAM4 in Figure 5c of Dong et al. (2019).

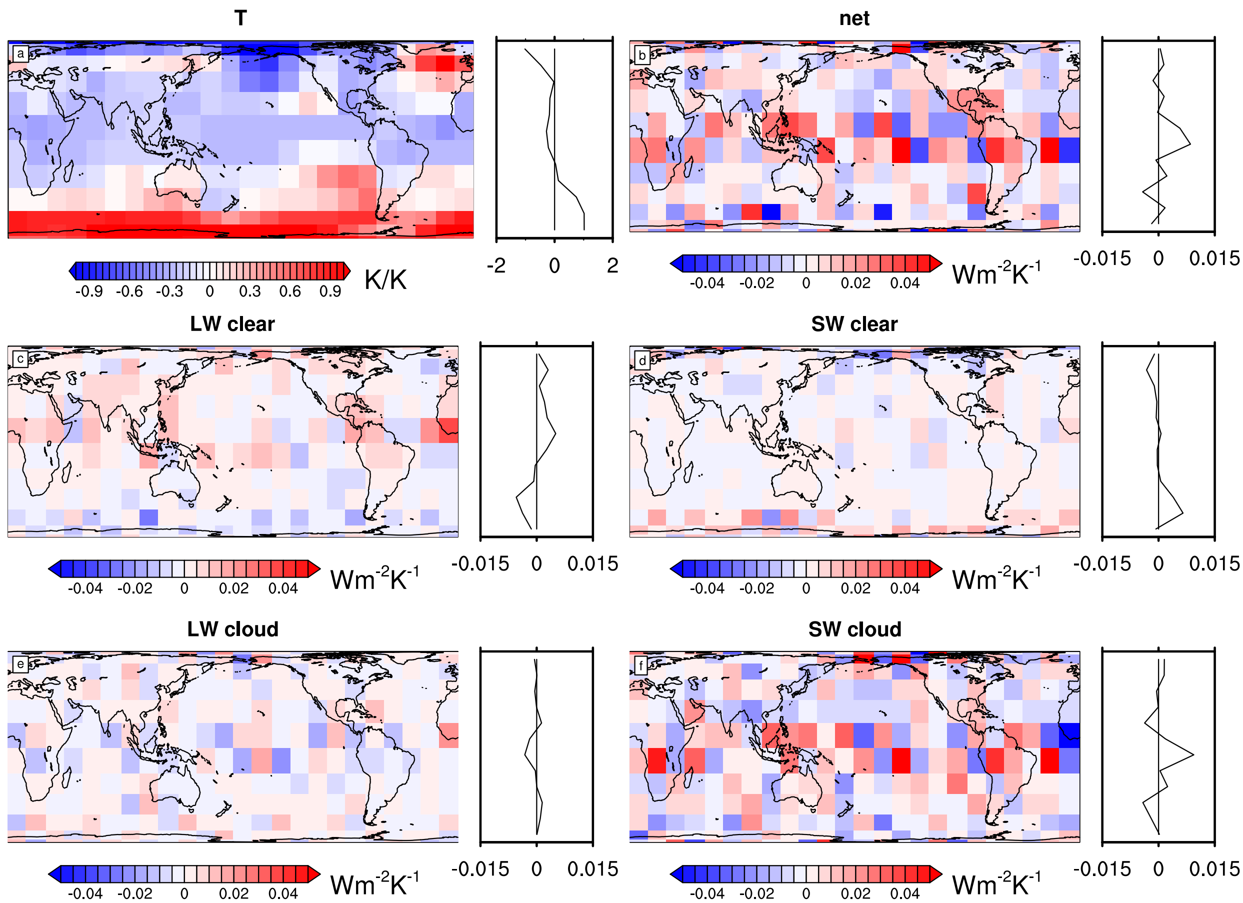

FIG. 10. Panel a shows the multi-model mean change in the pattern of warming between the abrupt4x early and late period, showing a shift towards regions of deep ocean heat uptake. Multiplying this pattern by MR-estimated spatial feedbacks gives an estimate of each grid cell's contribution to the change in feedback with time, $\Delta\lambda_{4x}$ (panels b-f). Although the resulting patterns are patchy, there are positive contributions from tropical convecting regions via the SW cloud and LW clear feedbacks, and from regions of Southern Ocean sea ice in the SW clear feedback, as shown by the accompanying zonal averages. The LW clear feedback has a compensating negative term from the Southern Ocean, so that its total estimated contribution to $\Delta\lambda_{4x}$ is smaller than the SW cloud feedback's (e.g., Figure S2 vs. Figure S5).