

Comparison of Short-Term and Long-Term Radiative Feedbacks and Variability in Twentieth-Century Global Climate Model Simulations

MEGHAN M. DALTON

Oregon Climate Change Research Institute, Oregon State University, Corvallis, Oregon

KAREN M. SHELL

Oregon State University, Corvallis, Oregon

(Manuscript received 31 July 2012, in final form 19 June 2013)

ABSTRACT

The climate sensitivity uncertainty of global climate models (GCMs) is partly due to the spread of individual feedbacks. One approach to constrain long-term climate sensitivity is to use the relatively short observational record, assuming there exists some relationship in feedbacks between short and long records. The present work tests this assumption by regressing short-term feedback metrics, characterized by the 20-yr feedback as well as interannual and intra-annual metrics, against long-term longwave water vapor, longwave atmospheric temperature, and shortwave surface albedo feedbacks calculated from 13 twentieth-century GCM simulations. Estimates of long-term feedbacks derived from reanalysis observations and statistically significant regressions are consistent with but no more constrained than earlier estimates.

For the interannual metric, natural variability contributes to the feedback uncertainty, reducing the ability to estimate the interannual behavior from one 20-yr time slice. For both the interannual and intra-annual metrics, uncertainty in the intermodel relationships between 20-yr metrics and 100-yr feedbacks also contributes to the feedback uncertainty. Because of differences in time scales of feedback processes, relationships between the 20-yr interannual metric and 100-yr water vapor and atmospheric temperature feedbacks are significant for only one feedback calculation method. The intra-annual and surface albedo relationships show more complex behavior, though positive correspondence between Northern Hemisphere surface albedo intra-annual metrics and 100-yr feedbacks is consistent with previous studies. Many relationships between 20-yr metrics and 100-yr feedbacks are sensitive to the specific GCMs included, highlighting that care should be taken when inferring long-term feedbacks from short-term observations.

1. Introduction and background

Earth's radiative energy balance is an important framework for understanding climate change. Any net positive (negative) imbalance ΔR of the global energy flux at the top of the atmosphere (TOA) averaged over a number of years leads to a warming (cooling). Changes in the TOA energy balance are given by

$$\Delta R = \Delta G + \lambda \Delta T, \quad (1)$$

where ΔG is the external radiative forcing; ΔT is the temperature response; and λ is the feedback parameter, which is inversely proportional (and of opposite sign) to

the climate sensitivity (Bony et al. 2006). The range in estimates of equilibrium climate sensitivity to a doubling of CO_2 for atmospheric global climate models (GCMs) from phase 3 of the World Climate Research Programme (WCRP) Coupled Model Intercomparison Project (CMIP3) multimodel dataset is 2.1–4.4 K with a mean value of 3.2 K (Randall et al. 2007). This range is due primarily to uncertainties of the individual feedbacks that make up the total feedback parameter.

There are many ways to decompose λ into individual components (Held and Shell 2012). Here, we decompose the feedback parameter as

$$\lambda = \lambda_{Q_{\text{LW}}} + \lambda_{Q_{\text{SW}}} + \lambda_{\text{TS}} + \lambda_{\text{TA}} + \lambda_{\alpha} + \lambda_{\text{C}} + \epsilon, \quad (2)$$

where $\lambda_{Q_{\text{LW}}}$ and $\lambda_{Q_{\text{SW}}}$ are the longwave and shortwave water vapor feedbacks; λ_{TS} and λ_{TA} are the surface and

Corresponding author address: Meghan M. Dalton, Oregon State University, 104 CEOAS Admin. Bldg., Corvallis, OR 97331.
E-mail: mdalton@coas.oregonstate.edu

atmospheric temperature feedbacks; λ_α is the surface albedo feedback; λ_C is the cloud feedback; and ϵ contains the cross-feedback terms and is assumed to be small, $\approx 10\%$ (Shell et al. 2008; Jonko et al. 2012). The atmospheric temperature feedback consists of the Planck response λ_0 and lapse rate feedback λ_L . Since λ_0 is essentially constant (Soden and Held 2006), model differences in λ_{TA} result from λ_L . Of the two water vapor feedbacks, λ_{QLW} dominates λ_{QSW} . The cloud feedback, while the largest source of uncertainty, requires complex treatment (Sanderson and Shell 2012) beyond the scope of this paper. Thus, we constrained our analysis to three feedbacks: namely, λ_{QLW} , λ_{TA} , and λ_α .

The relatively short satellite observational record can be used to constrain modeled feedback estimates over the same time period (i.e., a few decades). Measurements from the Earth Radiation Budget Experiment (ERBE; Barkstrom 1984) indicate that the observed water vapor feedback associated with El Niño–Southern Oscillation (ENSO) variability (Soden 1997) and the Mt. Pinatubo eruption (Forster and Collins 2004), as well as global-mean radiative damping rates derived from interannual variability (Chung et al. 2010), are consistent with GCM estimates. Dessler (2010, 2013) finds good agreement between feedbacks calculated from a decade of observations and from a GCM control simulation. On the other hand, Hall and Qu (2006) find that most modeled snow albedo seasonal cycle feedback strengths are outside the range of observed estimates, and Flanner et al. (2011) find that models underestimate the observed Northern Hemisphere (NH) albedo feedback as derived from a 30-yr record.

If the spatial structures of short-term and long-term climate variable changes are similar, then feedback parameters derived over short-term periods are likely representative of long-term climate change (Forster and Gregory 2006; Colman and Hanson 2013; Boer and Yu 2003). Interannual variability may provide some information about climate change feedbacks (Colman and Power 2010), and the constraint of long-term climate sensitivity by observed seasonal sensitivity may be justified if the two are governed by similar processes (Knutti et al. 2006; Hall and Qu 2006). However, estimates from short-term observations may not be appropriate for long-term inferences if the fast feedback components do not represent the total climate feedback parameter (Lin et al. 2011). For example, the feedback parameter calculated from observed tropical variability and ENSO can vary from the climate change feedback parameter (Forster and Gregory 2006). The representativeness of the short-term observations may depend based on the particular processes and locations considered.

To what extent can long-term (e.g., 100 yr) feedbacks of the actual climate system be estimated by a short period (e.g., 20 yr) of observations? Since only GCMs have the luxury of long records, several studies have compared GCMs feedbacks over different time scales. Some results are encouraging, such as relationships between NH springtime snow albedo change per temperature change and the April albedo change per temperature change between the twentieth and twenty-second centuries in CMIP3 models (Hall and Qu 2006) and between NH temperature seasonal cycle amplitude and climate sensitivity (Knutti et al. 2006). However, other studies find that feedbacks operating under shorter time scales (e.g., ENSO; unforced variability) overestimate or underestimate those operating under longer time scales (Dessler and Wong 2009; Colman and Power 2010), while others have found no or weak relationships between short- and long-term feedbacks (Dessler 2010, 2013; Colman and Hanson 2013). Furthermore, Armour et al. (2013) suggest that feedbacks operate on different time scales, based on the pattern of surface warming, with the implication that using short-term feedbacks to estimate long-term climate change is not always feasible.

Assuming there is a relationship between short-term and long-term behavior in models, the short-term modeled behavior can be compared with observations to constrain estimates of feedbacks over the longer period (e.g., Knutti et al. 2006; Hall and Qu 2006). Model improvements can focus on better representation of, for example, the seasonal snow cycle (Hall and Qu 2006) to narrow the spread in climate change feedback strength. Note, however, that this framework assumes the modeled relationship between short-term and long-term feedbacks exists in the actual climate.

We test three short-term feedback variability metrics as “proxies” for long-term water vapor, atmospheric temperature, and surface albedo feedbacks. First, we quantify short-term twentieth-century feedbacks and interannual and intra-annual feedback variability in an ensemble of GCM simulations, and then we compare these short-term characteristics with modeled long-term twentieth-century feedbacks. We use an ensemble, rather than a single model, because we are searching for relationships that hold across many models, suggesting some fundamental process within the actual climate system that is adequately captured by most models. We also calculate the uncertainty due to the natural variability internal to each model, indicating how representative one short-term observation is of natural variability calculated from a longer period. Finally, we compare modeled short-term feedback variability with the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim)

TABLE 1. Abbreviation, name, and number of ensemble members (runs) of the 13 coupled atmosphere–ocean global climate models used in this study.

Abbreviation	Expanded model name	Runs
CCSM3 (CCSM)	NCAR Community Climate System Model, version 3	8
CGCM3.1(T47) (CGCMT47)	Canadian Centre for Climate Modelling and Analysis (CCCma) Coupled General Circulation Model, version 3.1 (T47 resolution)	5
CGCM3.1(T63) (CGCMT63)	Canadian Centre for Climate Modelling and Analysis (CCCma) Coupled General Circulation Model, version 3.1 (T63 resolution)	1
ECHAM	Max Planck Institute for Meteorology ECHAM, version 5	3
GFDL CM2.0 (CM20)	Geophysical Fluid Dynamics Laboratory Climate Model, version 2.0	3
GFDL CM2.1 (CM21)	Geophysical Fluid Dynamics Laboratory Climate Model, version 2.1	3
GISS-AOM (GISSAOM)	National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies, Atmosphere–Ocean Model	1
GISS-E2-H (GISSEH)	NASA Goddard Institute for Space Studies Model E2, coupled with the Hybrid Coordinate Ocean Model (HYCOM)	5
GISS-E2-R (GISSER)	NASA Goddard Institute for Space Studies Model E2, coupled with the Russell ocean model	9
INM-CM3.0 (INM)	Institute of Numerical Mathematics Coupled Model, version 3.0	1
IPSL-CM4 (IPSL)	L'Institut Pierre-Simon Laplace Coupled Model, version 4	1
MIROC3.2 (hires) (MIROCHI)	Model for Interdisciplinary Research on Climate, version 3.2 (high resolution)	1
MIROC3.2 (medres) (MIROCMED)	Model for Interdisciplinary Research on Climate, version 3.2 (medium resolution)	3

product and estimate long-term feedbacks based on the significant relationships from the models and our estimates of internal variability.

2. Data and methods

We analyze feedback in twentieth-century simulations from 13 fully coupled atmosphere–ocean GCMs from the CMIP3 archive (Table 1). We use twentieth-century simulations instead of runs with natural variability alone [i.e., preindustrial simulations as in Colman and Hanson (2013) and Dessler (2013)] because the observations do not reflect purely natural variability, but the superimposed externally forced warming as well. The twentieth-century experiment attempts to recreate the observed forcings of the actual climate system and is the most appropriate dataset for comparison with recent observations.

We first analyze the entire 100-yr period 1901–2000 (or 1900–99) as the long-term climate information.

Then we divide the 100-yr period into five sequential, nonoverlapping 20-yr slices and perform the same analysis on each slice for each ensemble member of every model. These 20-yr slices may be thought of as separate realizations of a short-term period analogous to the record length of reliable satellite or reanalysis observations. By analyzing several 20-yr periods, we can determine the inherent variability and uncertainty of using a single short period of observations to constrain future long-term projections.

We calculate three metrics of short-term feedback variability are 1) 20-yr feedback, 2) an interannual metric, and 3) and intra-annual metric for three feedback variables: water vapor, atmospheric temperature, and surface albedo. We also calculate 100-yr feedbacks. Note that the water vapor metrics only consider the longwave (LW) effects. Feedback metric calculation methods are summarized in Table 2 and discussed further in sections 2b–2d. Since Colman and Hanson (2013) recently performed a comparison of short-term and

TABLE 2. Feedback calculation methods and metrics of feedback variability. [For intra-annual feedback variables, the seasonal values are June–August (JJA), December–February (DJF), July–September (JAS), and January–March (JFM).]

Metric	Description
Feedback, M1	Regression of deseasonalized TOA flux against $\overline{T_{as}}$ anomalies
Feedback, M2	Difference between first and last 20-yr averages of deseasonalized TOA flux anomalies divided by difference of $\overline{T_{as}}$ anomalies
Interannual	Standard deviation of deseasonalized, detrended TOA flux anomalies
Intra-annual	Seasonal cycle amplitude of feedback variables
Water vapor	Global: JJA – DJF; NH: JAS – JFM; SH: JFM – JAS
Atmospheric temperature	Global: JJA – DJF; NH: JJA – DJF; SH: [JJA – DJF]
Surface albedo	Global: FM – JA; NH: [June – April]; SH: August – June

long-term longwave feedbacks with a goal similar to ours, we highlight relevant similarities and differences in metric definitions below.

a. TOA flux anomalies

The radiative kernel technique (Soden et al. 2008; Shell et al. 2008) decomposes each feedback into two components: the TOA flux change due to a standard change in the feedback variable at each horizontal location and vertical level (radiative kernel; $K_x = \partial R_x / \partial x$) and the change in the feedback variable in response to a surface air temperature change (climate response; $dx/d\bar{T}_{as}$). Note that we normalize the feedback by the standard anomaly used to compute the kernel (Shell et al. 2008) but omit that notation for simplicity. The feedback strength for variable x is given by

$$\lambda_x = \frac{\Delta R_x}{\Delta T_{as}} \approx \frac{\partial R_x}{\partial x} \left(\frac{dx}{d\bar{T}_{as}} \right) \approx K_x \left(\frac{\Delta x}{\Delta T_{as}} \right). \quad (3)$$

We use the precalculated National Center for Atmospheric Research (NCAR) Community Atmosphere Model (CAM; approximately 2.8° latitude by 2.8° longitude with 17 vertical levels) kernel (Shell et al. 2008), so only the change in the climate variable Δx , given a change in climate ΔT_{as} , is needed to calculate the feedback strength. Limiting our use to only one model's kernel is unlikely to substantially affect our results (Soden et al. 2008). In a similar study to ours, Colman and Hanson (2013) find little dependence of feedback estimates on two different kernels. Additionally, the focus of this paper is to identify the variations in climate responses of models resulting in variations in climate feedbacks, not to obtain specific values for feedbacks. The standard kernel technique uses the differences in mean variables between two climate states. We use a modified technique by instead considering anomalies in specific humidity, atmospheric temperature, and surface albedo from the mean climate, as in Dessler (2013). Because absorption of radiation by water vapor behaves like the natural log of specific humidity (Cess 1974; Thompson and Warren 1982), we use the natural log of the specific humidity as the water vapor variable, as in Soden et al. (2008).

For each variable, we first subtract the average seasonal cycle from each year of the corresponding 20- or 100-yr time series. Then, we multiply these feedback variable anomalies by the radiative kernel at each grid point and level (for atmospheric temperature and water vapor) to produce the TOA flux anomalies ($\Delta R_x = K_x \Delta x$). Finally, we sum the flux anomalies from the surface to the top of the model's atmosphere, resulting in the cumulative TOA radiative effect. Converting feedback variable anomalies

to TOA flux anomalies preserves the spatial pattern of anomalies and accounts for their contribution to the TOA energy balance. In contrast, Colman and Hanson (2013) sum only up to the tropopause. Thus, their feedbacks omit stratospheric effects while ours can be more directly compared to, for example, observed TOA flux changes.

In the clear-sky test (Shell et al. 2008), clear-sky TOA fluxes are compared to the sum of the kernel-derived radiative effects of feedback variables and the CO_2 radiative forcing. We find good agreement for both the CMIP3 climate models and the ERA-Interim dataset. The largest differences occur where the radiative effect of sulfate aerosols is large. This supports the assumption that the cross-feedback term in Eq. (2) is small not only for large climate changes but also for monthly anomalies. Other studies have shown the usefulness of this technique with satellite data (Dessler et al. 2008).

We then average the time series of deseasonalized TOA flux anomalies for each variable over the globe, NH, and Southern Hemisphere (SH). While important features that affect the global energy balance should be discerned globally (Trenberth et al. 2010), there are different and competing processes due to differences in land configuration and atmospheric circulation in each hemisphere that are important to understand separately.

b. Feedback calculations

We calculate feedbacks from the area-averaged TOA flux and surface temperature anomalies in two ways (Table 2). The first method (M1) consists of regressing deseasonalized monthly TOA flux anomalies onto deseasonalized monthly \bar{T}_{as} anomalies, a commonly used technique (e.g., Dessler 2013) similar to that introduced by Gregory et al. (2004). The second method (M2) consists of dividing the difference between the first and last 20-yr averages of the deseasonalized TOA flux anomaly time series by the corresponding difference in the deseasonalized \bar{T}_{as} anomalies, similar to the method of Soden et al. (2008). Colman and Hanson (2013) use this method to calculate transient feedbacks between 10-yr periods, rather than 20-yr periods. However, we find little difference in 100-yr M2 feedback values calculated using 10- or 20-yr averages. Table 3 lists the global 100-yr feedback values for both methods, which are within each other's standard errors.

For comparison, we include the twenty-first-century feedback estimates derived from Soden and Held (2006). The twentieth-century values are close to the twenty-first-century estimates, though markedly smaller for the atmospheric temperature and surface albedo feedbacks. Soden and Held (2006) sum feedbacks from the surface to the tropopause rather than throughout the entire

TABLE 3. Global twentieth-century longwave water vapor, longwave atmospheric temperature, and shortwave surface albedo feedbacks ($\text{W m}^{-2} \text{K}^{-1}$) calculated using both methods, described in Table 2 and in the text. Twenty-first-century feedback estimates derived from Soden and Held (2006) are listed in the bottom row. Values given are the multimodel mean with one standard deviation.

	Water vapor	Atmospheric temperature	Surface albedo
M1	1.69 ± 0.28	-2.73 ± 0.40	0.20 ± 0.13
M2	1.79 ± 0.26	-2.69 ± 0.44	0.20 ± 0.16
Soden and Held	1.8 ± 0.18	-3.2 ± 0.3	0.26 ± 0.08

atmosphere, which may account for the difference in atmospheric temperature feedbacks. As the troposphere warms, the stratosphere cools, thus reducing the outgoing longwave radiation and damping the negative temperature feedback. The surface albedo feedback is small in the twentieth century; while most models exhibit a positive albedo feedback, some models actually have a negative feedback, unlike for the twenty-first century. In addition, Soden and Held (2006) use a different subset of CMIP3 models, resulting in a different ensemble mean.

There is general positive correspondence between M1 and M2 feedbacks for all variables and regions (Fig. 1). For water vapor and atmospheric temperature, M1 feedbacks are generally smaller in magnitude than M2 feedbacks. Thus, the processes omitted by M2 tend to decrease feedbacks (making M2 feedbacks larger in magnitude than M1), indicating nonlinearities in these feedbacks. On the other hand, there is close to a one-to-one correspondence between M1 and M2 surface albedo feedbacks, suggesting that the feedback calculation methodology is less important. In all cases, though, individual models may display anomalous behavior. ECHAM in particular stands out, especially for the SH and globe, as having much larger M1 feedbacks for water vapor and atmospheric temperature and a much larger M2 feedback for surface albedo.

Differences in long-term feedback values between M1 and M2 are due to the nonlinear behavior of some models. Because monthly perturbations of TOA flux anomalies and surface air temperature anomalies are correlated, M1 feedbacks capture more of the inherent variability of feedback behavior throughout the record on month-to-month and year-to-year time scales. In contrast, M2 aggregates 20 yr of these monthly perturbations to capture the overall change with less attention to shorter-scale variability. Thus, M1 incorporates relationships over time scales ranging from years to decades, while M2 includes only relationships that operate

on long-term (roughly 80 yr) time scales. If feedback behavior varies across time scales, then these two calculations will produce different results. In fact, Armour et al. (2013) find that the apparent time variation of feedbacks within a model is actually due to a changing influence of different regions on the global-average temperature change. Similarly, Winton et al. (2010) suggest that the climate system can be interpreted to have a time-varying “efficacy” of ocean heat uptake, which influences the transient response of climate to a forcing. In both these frameworks, the pattern of temperature response is important. To the extent that the temperature change between 1900 and 2000 has a different horizontal structure than interannual temperature anomalies (e.g., those found for ENSO), we expect that M1 and M2 will differ. Additionally, M2 feedbacks may be biased by decadal internal variability if the start and end years for averaging happen to fall during opposite phases of decadal variability. The use of 10-yr averages, as opposed to 20-yr averages, increases the likelihood of this type of bias.

M2 (based on differences in beginning and ending averages) assumes that changes in feedback variables occur only in response to temperature anomalies and thus includes the “fast responses” (Gregory and Webb 2008) excluded by M1 (the radiative kernel-regression technique). However, Andrews and Forster (2008) find that the noncloud fast responses contribute little to differences in feedbacks calculated using their “direct” method (similar to our M2) and their “climate” method (similar to our M1). Colman and McAvaney (2011) also find insignificant or small rapid responses by water vapor, lapse rate, or surface albedo in a GCM. Thus, we do not expect fast responses to contribute to the differences between methods for the feedbacks we consider, especially across the ensemble average.

Long-term feedbacks are calculated by both methods, but short-term feedbacks are only calculated with M1. Since division by small (close to zero) temperature changes over 20-yr results in unrealistically large feedback values, M2 is inadequate for short records.

c. Interannual metric

The interannual metric of feedback variability is the standard deviation of deseasonalized, detrended global, NH, or SH TOA flux anomalies. A linear least squares trend is removed from the 20-yr time periods and, because of nonlinearity, a quadratic trend is removed from the 100-yr periods. We then calculate the standard deviations of detrended TOA flux anomalies in an attempt to summarize the general state of month-to-month variability of a particular time period in a single value (e.g., if the time series behaves wildly with a large range

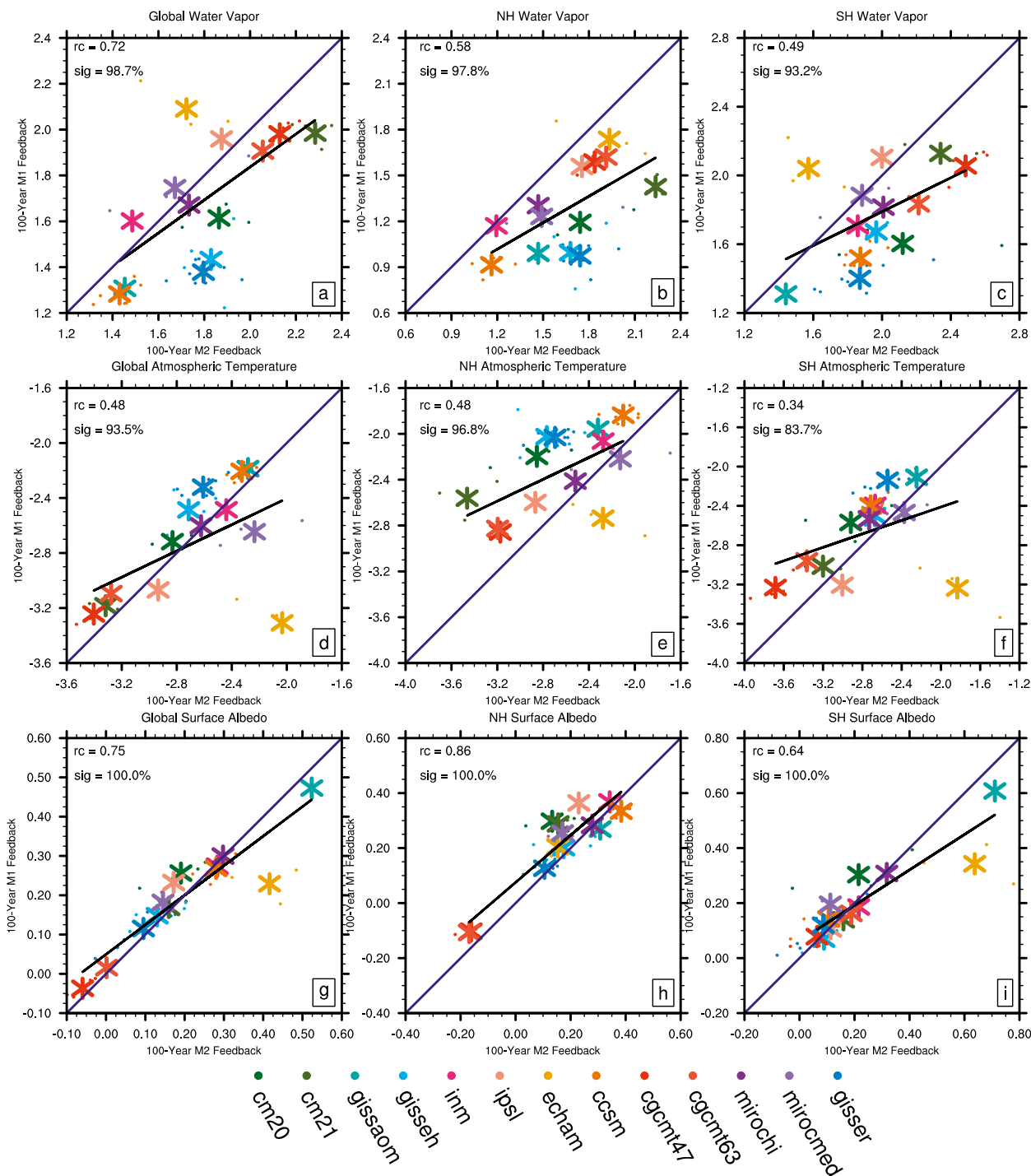


FIG. 1. Comparison of 100-yr feedbacks ($\text{W m}^{-2} \text{K}^{-1}$) calculated with M1 and M2 for all three regions and variables: (left)–(right) the globe, NH, and SH; (top)–(bottom) water vapor, atmospheric temperature, and albedo. Each dot represents one ensemble member. The asterisks indicate the mean metric for each of the 13 models (colors). The blue line indicates the one-to-one correspondence. The regression coefficients (rc) and significances (sig) of the nonzero regression slope using the model-mean values (black line) are listed for each comparison.

100-year Average Seasonal Cycle

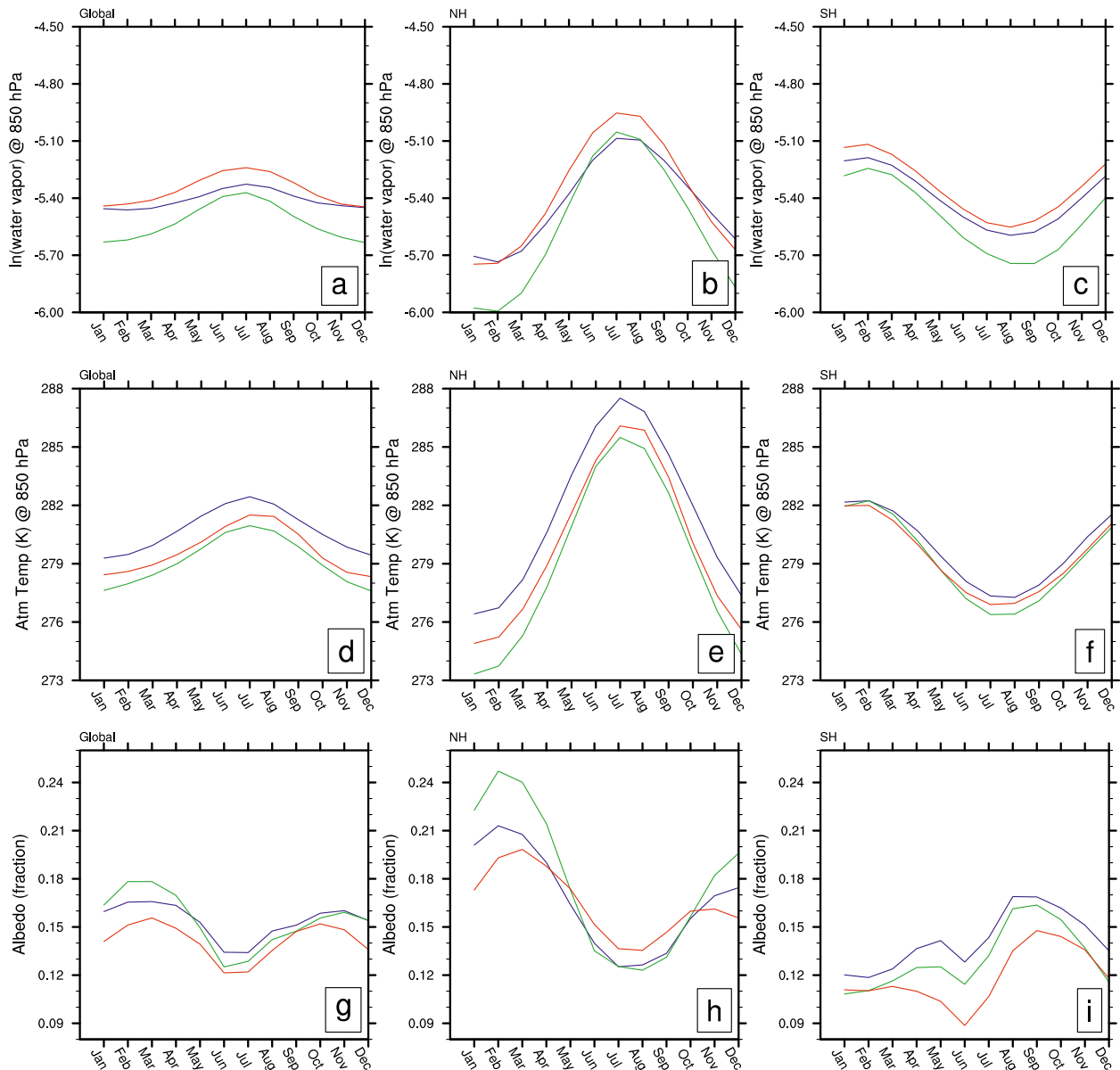


FIG. 2. The 100-yr-average seasonal cycle of (top) natural log of specific humidity at 850 hPa, (middle) atmospheric temperature at 850 hPa, and (bottom) surface albedo for CCSM3 (blue), GFDL CM2.0 (green), and GISS-E2-R (red) for the (left) the globe, (center) NH, and (right) SH.

of values or if it varies more calmly with values generally closer to the zero mean). We use this metric to examine whether large interannual swings in TOA flux anomalies are indicative of larger long-term feedbacks. In contrast, the interannual feedback of Colman and Hanson (2013) is calculated by regressing radiative perturbations between adjacent 1-yr averages onto equivalent surface air temperature perturbations (similar to M1) and thus estimates feedbacks dealing with variability on 2-yr time scales.

d. Intra-annual metric

The intra-annual metric is the amplitude of the seasonal cycle of a feedback variable that is removed before the TOA flux anomaly calculation. We calculate the seasonal cycle by averaging all the Januarys, Februarys, and so on, over the length of each 20-yr slice or 100-yr period. For the atmospheric variables, we use the seasonal cycle at 850 hPa to avoid very large variability due

to localized surface processes while still retaining the large seasonal cycle in the lower troposphere compared with higher altitudes. Figure 2 shows the seasonal cycles averaged over the 100-yr period for three models. We use this metric to examine whether large changes over the seasonal cycle are indicative of large long-term feedbacks: that is, whether the seasonal cycle is representative of climate change (e.g., Hall and Qu 2006; Knutti et al. 2006). Note that this metric is different from the seasonal metric of Colman and Hanson (2013), which is calculated from radiative perturbations over 2-monthly steps.

We define the amplitude differently for each variable and geographic area, with the goal of capturing the maximum seasonal change signal. Seasonal amplitude definitions are listed in Table 2. For water vapor and atmospheric temperature, we use the difference between the summer and winter seasons; for surface albedo, we use the greatest seasonal rate of change. While the NH dominates the global seasonal cycle for water vapor and atmospheric temperature, for the surface albedo, the NH controls the global cycle in the boreal winter and spring, whereas the SH dominates the rest of the year. Thus, we define the global seasonal surface albedo amplitude (Fig. 2g) as the average of the two maximum months [February and March (FM)] minus the average of the two minimum months [July and August (JA)]. We define the NH seasonal amplitude as the change in surface albedo between April and June because this period has the largest (negative) rate of change (Fig. 2h). This is similar to the definition used by Hall and Qu (2006). We define the SH surface albedo seasonal amplitude as the change from June to August, the largest (positive) rate of change (Fig. 2i). Since Hall and Qu (2006) find a relationship by normalizing the percent snow albedo change by the temperature change of the same period, we also divide the albedo seasonal amplitudes by the corresponding surface air temperature amplitudes and compute regressions with this “normalized” seasonal amplitude metric.

e. Ensemble members and 20-yr time slices

Each model has a spread of metric values across ensemble members and 20-yr slices. Ensemble members from a single model are not independent, nor are the five 20-yr slices within a member. The range among 20-yr slices represents the uncertainty due to internal variability on the short time scale, cautioning the use of a single 20-yr period to make inferences about a longer period. To quantify this internal variability, we define the spread for each model as the standard deviation of the 20-yr values from all ensemble members of the model divided by the average of the 20-yr values, expressed

as a percentage. This value is used as an estimate for observational uncertainty when estimating long-term feedback from reanalysis observations.

To examine the extent to which 20-yr metrics (intra-annual, interannual, or feedback) can be used to estimate corresponding 100-yr feedbacks, we perform regressions between the short-term and long-term modeled quantities. The significance of the regression coefficient is determined by the p value for a two-tailed Student's t test using a t statistic and degrees of freedom of the regression line and tested against the null hypothesis that the regression coefficient is zero (i.e., the metrics are unrelated). Results are judged to be significant at >95% significance of a nonzero slope. The significant regression slopes are subsequently used to estimate long-term feedbacks from reanalysis observations.

We regress each model's ensemble-average 20-yr metric against the corresponding ensemble-average 100-yr feedback values (indicated by the 13 asterisks shown in Figs. 1, 3–7, with solid black regression lines). Our goal is to identify any relationships between metrics that hold across most models. No single model can be identified as the “best” model; furthermore, individual models tend to have limited ranges of feedback and other metric values, increasing uncertainty in regression calculations. Thus, we have more confidence in common behavior in intermodel versus intramodel ensembles, under the assumption that similarities across models are more likely to correspond to actual climate behaviors.

Regressions through all 20- and 100-yr points (not shown) yield similar slopes. These slopes have greater significances and smaller standard errors; certain models with more ensemble members counter the effects of outliers, reducing the apparent uncertainty. However, this regression methodology effectively makes the (unsupported) assumption that models with more ensemble members are closer to reality. We also calculate the regressions using the first ensemble members of each model. Again, this results in smaller standard errors and higher significances for most cases. However, this method has the disadvantage of excluding information that would improve the behavior of some models.

Also, since there is variance in both the 20- and 100-yr data, we ideally could weigh each average model value by the variance in both dimensions when computing the linear regression coefficient for the 13 model points. Attaching weights to each model based on the inverse of the variance of the 20-yr data (we could not incorporate the variance of the 100-yr data since not all models have multiple ensemble members), we find little difference in the regressions compared with the unweighted regressions. Thus, we focus on the regressions with

TABLE 4. Regression slopes (unitless) with standard errors for 20- (short term) to 100-yr (long term) same-metric comparisons for all variables and regions using the 13 model means. All significances are >99.99% unless otherwise noted. Regular and normalized [section 3d(2)] regressions are listed for the intra-annual surface albedo metrics.

Metric	Water vapor	Atmospheric temperature	Surface albedo	
M1 feedback				
Global	0.38 ± 0.12 (98.9%)	0.41 ± 0.15 (98.0%)	1.21 ± 0.34 (99.6%)	
NH	0.45 ± 0.18 (97.4%)	0.43 ± 0.23 (91.0%)	0.97 ± 0.26 (99.7%)	
SH	0.28 ± 0.11 (97.6%)	0.34 ± 0.13 (97.6%)	1.41 ± 0.25 (99.9%)	
Interannual				
Global	1.01 ± 0.02	1.00 ± 0.02	1.07 ± 0.09	
NH	1.01 ± 0.02	1.01 ± 0.02	1.04 ± 0.05	
SH	1.01 ± 0.01	1.00 ± 0.02	1.13 ± 0.15	
Intra-annual				Normalized
Global	1.06 ± 0.04	0.99 ± 0.03	0.97 ± 0.05	0.88 ± 0.04
NH	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
SH	1.01 ± 0.01	1.00 ± 0.00	0.99 ± 0.04	0.85 ± 0.04

the unweighted ensemble-average values. In this way, all models are treated equally rather than giving more weight to models with more ensemble members. We note below when conclusions are sensitive to the specific methodology.

Our analysis uses multiple ensemble members, whereas most other studies select one ensemble member per model (e.g., Colman and Hanson 2013), effectively underestimating the amount of internal variability and thus the uncertainty. Among other factors, Masson and Knutti (2013) demonstrate that care should be taken in the interpretation of short-term to long-term relationships across models when the ensembles are small.

Some relationships between short-term feedback metrics and long-term feedbacks are sensitive to the method by which feedbacks are calculated; others depend on which models are included. Given the lack of an objective technique to justify exclusion of models, we include all the models available to us. A thorough analysis of the dependence of these relationships on which models are included is beyond the scope of this paper, but regressions excluding certain outlier models are noted below.

f. Estimating long-term feedbacks using reanalysis data

We estimate long-term feedbacks from observations, assuming the modeled relationships between 20-yr metrics and 100-yr feedbacks also hold for the actual climate system. We use relationships with regressions significantly different from zero at the 95% level, along with reanalysis observations, to estimate long-term feedbacks, accounting for approximate error in the reanalysis data and the standard error in the regression slope and y intercept. We use the ERA-Interim product, derived by assimilating observational data into a forecast model to produce the global state of the atmosphere from 1979 to

the present (Dee et al. 2011), as the “truth.” We analyze 20 yr (1989–2008) of monthly averages of specific humidity and atmospheric temperature from 1000 to 10 mb and surface fields of forecast albedo and 2-m above-surface air temperature in the same way we analyze the model data. As an estimate of the reanalysis uncertainty due to the limited observational period, we use the spread in 20-yr model values. Our simple methodology is similar to but underestimates the total uncertainty of the feedback estimation framework in Masson and Knutti (2013).

3. Results and discussion

Our goal is to test how well three short-term feedback metrics (20-yr feedback, standard deviation of TOA flux anomalies, and amplitude of seasonal cycle of feedback variables) represent long-term feedbacks. We discuss relationships first between 20- and 100-yr values for the same metric and then between 20-yr metrics and 100-yr feedbacks. Long-term feedback estimates, when appropriate, are presented within each section.

a. Short-term metrics

The most desirable short-term metric would be relatively insensitive to the particular 20-yr slice used, because then one 20-yr period of observations (e.g., the existing short record) would be sufficient for exploring relationships with long-term feedbacks. Table 4 summarizes the 20–100-yr comparisons for the feedback, interannual metric, and intra-annual metric. The spread (percent ratio between the standard deviation and mean of 20-yr values) of the 20-yr feedbacks (M1) is large (14%–47%), indicating that a single 20-yr observation cannot be used to estimate the 100-yr feedbacks with much confidence. However, the correspondence between the 20- and 100-yr feedbacks using M1 is encouraging and discussed further in section 3b.

The spread in the 20-yr standard deviation (approximately 10%) suggests that one 20-yr period of observations of this metric may be better, though still not ideal, for representing the 100-yr feedback. Figure 3 shows examples of interannual, as well as intra-annual, comparisons for the NH and all three variables. While the regression lines in Figs. 3b,d,f indicate near one-to-one correspondence, the 20-yr standard deviation underestimates the 100-yr standard deviation, as indicated by the position of the points above the one-to-one line. This is likely because one 20-yr period does not capture all the extremes and low frequency cycles of the longer period. However, the regressions are all significant at $>99.9\%$, and a slope of 1 is within the standard errors. The reanalysis 20-yr standard deviation is marked by the vertical black line, which is within the range of modeled standard deviations for all variables and regions, except NH surface albedo.

The spread in the 20-yr seasonal cycle amplitude for each model is smaller still (about 1%). Also, the seasonal cycles differ more between models than within a single model, as seen in Figs. 3a,c,e. Almost all combinations of regions and variables show near one-to-one correspondence (regression slopes of 0.97–1.06), significant at $>99.9\%$. The normalized surface albedo seasonal cycles for the SH and global regions have regression slopes <1 : 0.85 and 0.88, respectively (Table 4). These smaller regression slopes are due to the influence of one or two models; most models fall on the one-to-one line. The vertical black lines mark the amplitudes of the 20-yr reanalysis data, which lie within the model spread for water vapor and atmospheric temperature seasonal amplitudes, but are less than almost all models' surface albedo amplitudes.

Because of the small spread in 20-yr seasonal amplitude values (i.e., 20-yr seasonal amplitude is a relatively stationary measure), we might have more confidence in any relationship found between the seasonal amplitude and 100-yr feedbacks than using other metrics that are more sensitive to the time slice. However, as shown below, many of the relationships found between 20-yr seasonal amplitudes and 100-yr feedbacks depend on the inclusion of a few models that behave differently from the rest. Since models have distinct seasonal amplitudes, these can be easily validated against observations, but the best performing model in terms of seasonal amplitude does not necessarily predict the correct long-term feedback (Knutti 2010).

b. 20-yr feedbacks versus 100-yr feedbacks

Regression slopes between 20- and 100-yr feedbacks calculated with M1 (the combined radiative kernel technique-regression approach) along with standard

errors and significances are summarized in Table 4 and shown in Fig. 4. In general there is positive correspondence, and M1 produces relationships with positive slopes significantly different from zero at the 95% level for all variables and regions except NH atmospheric temperature. Note, however, that, because of the large intramodel spread of the 20-yr feedbacks (14%–47%), we cannot accurately calculate feedbacks, despite the highly significant relationships between 20- and 100-yr feedbacks.

1) WATER VAPOR AND ATMOSPHERIC TEMPERATURE

The regression slopes are positive and <1 for all water vapor and atmospheric temperature feedbacks calculated using M1, suggesting that stronger 20-yr feedbacks correspond to stronger 100-yr feedbacks. In contrast, Colman and Hanson (2013) find negative, though not significant, global relationships for their water vapor and lapse rate decadal versus transient feedbacks. One reason for this apparently conflicting result could be the fact that the decadal feedback (similar to M1) in Colman and Hanson (2013) is computed in a different way than the transient feedback (similar to M2), whereas our short-term and long-term feedbacks are calculated with the same method (M1). Another difference is that several models have positive decadal lapse rate feedbacks in Colman and Hanson (2013), whereas our atmospheric temperature short-term feedbacks are all negative.

For a majority of models, 20-yr feedbacks are smaller than 100-yr feedbacks, as indicated by the position of the points on the left (right) side of the blue equal-strength line for water vapor (atmospheric temperature) in Figs. 4a–f. The two CGCMs have much larger differences between 100- and 20-yr feedbacks compared with other models, as indicated by the greatest horizontal distances from the one-to-one lines, and omitting them from the regression calculation results in larger slopes. Slopes remain below 1 because models with high 20-yr water vapor or atmospheric temperature feedbacks tend to have similar or stronger 20-yr feedbacks compared with 100-yr feedbacks, while models with smaller 20-yr feedbacks are more likely to have weaker 20-yr feedbacks than 100-yr feedbacks.

2) SURFACE ALBEDO

The surface albedo slopes are positive, and most models fall on or near the one-to-one line, indicating near-equal strength in 20- and 100-yr feedbacks calculated with M1. Obvious exceptions include both CGCMs and GISS-AOM. Global and NH regression slopes including all models are within the standard error of a slope of 1, but the SH regression slope is >1 by more than the

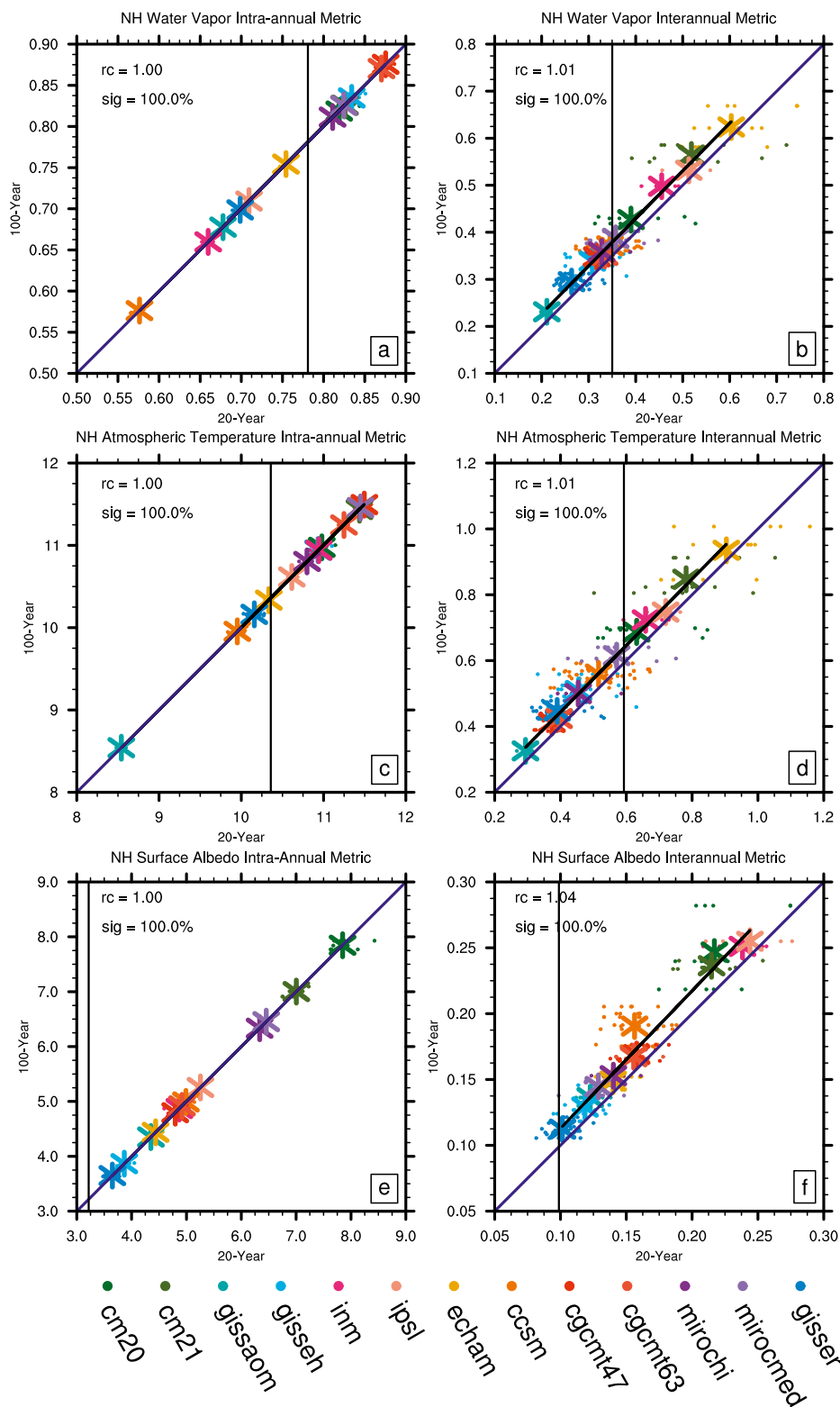


FIG. 3. Comparison of 20- and 100-yr (left) seasonal cycle amplitude (unitless) and (right) standard deviation (W m^{-2}) for NH: (top) water vapor, (middle) atmospheric temperature, and (bottom) surface albedo. See Fig. 1 for conventions. The vertical lines indicate the reanalysis values.

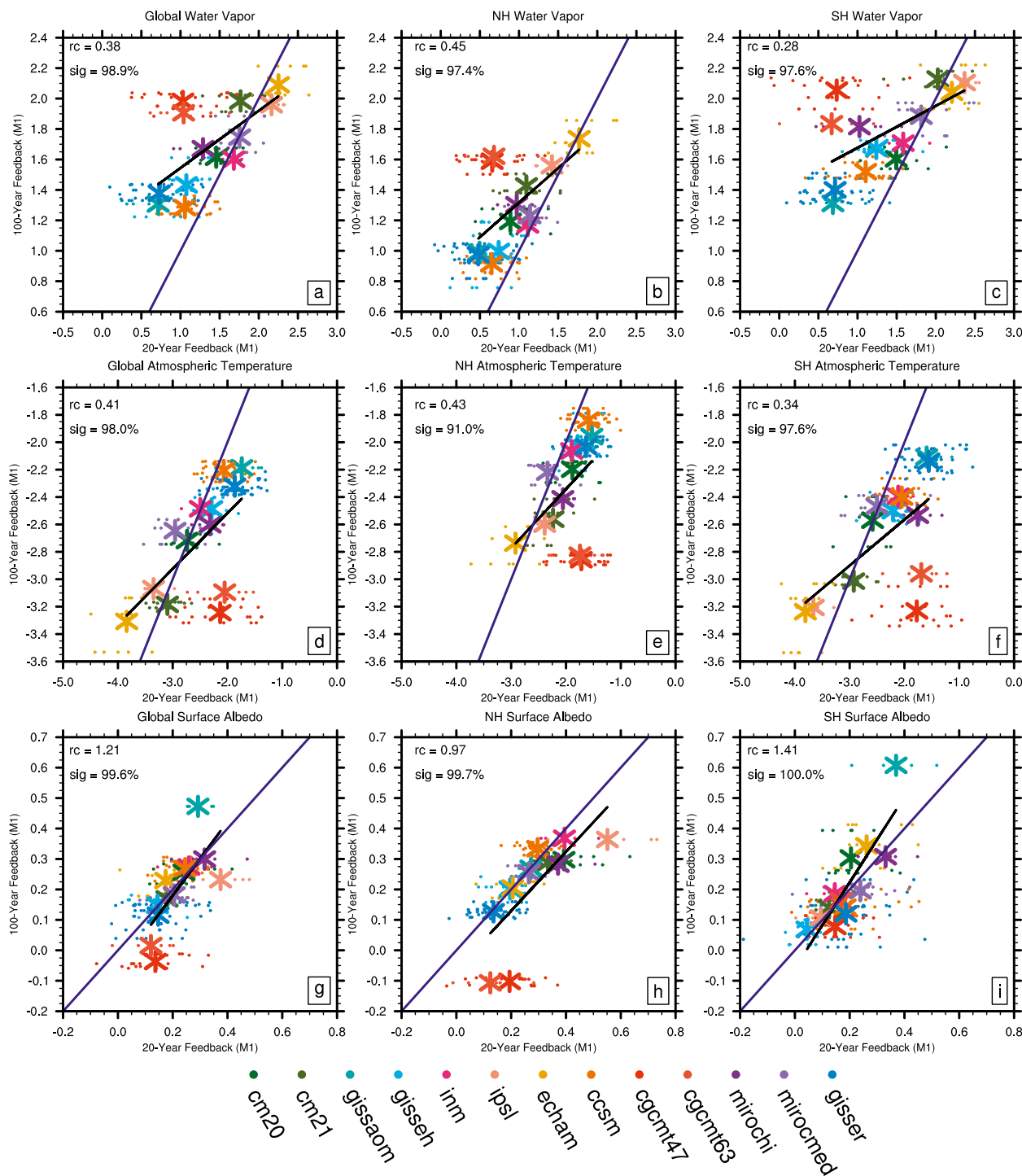


FIG. 4. Comparison of 20- and 100-yr feedbacks ($W m^{-2} K^{-1}$) calculated using M1 for (top)–(bottom) water vapor, atmospheric temperature, and surface albedo for (left)–(right) the globe, NH, and SH. See Fig. 1 for conventions.

standard error. For the globe and NH, the CGCMs have near-zero or negative 100-yr feedbacks and positive but relatively small 20-yr feedbacks. Excluding these models decreases the global and NH regression

slopes slightly while retaining significance. For the SH, omitting the CGCMs alters the slope little. However, exclusion of GISS-AOM, which has a much smaller 20-yr feedback than 100-yr feedback in the SH and is

TABLE 5. Regression slopes ($\text{W m}^{-2} \text{K}^{-1}$ per W m^{-2}) with standard errors (and significances) for 20-yr standard deviations of monthly TOA flux anomalies and 100-yr feedbacks for all three regions and variables and both feedback methods. Regressions with $>95\%$ significance are in boldface.

Region	Water vapor	Atmospheric temperature	Surface albedo
Global			
M1	1.48 ± 0.49 (98.8%)	-1.11 ± 0.49 (95.5%)	0.57 ± 1.74 (25.2%)
M2	0.43 ± 0.60 (51.5%)	0.26 ± 0.65 (29.8%)	1.04 ± 2.12 (36.7%)
NH			
M1	1.55 ± 0.59 (97.7%)	-0.54 ± 0.56 (64.3%)	1.32 ± 0.90 (82.8%)
M2	0.99 ± 0.76 (78.4%)	0.05 ± 0.73 (5.6%)	0.51 ± 1.05 (36.3%)
SH			
M1	1.34 ± 0.39 (99.5%)	-1.19 ± 0.47 (97.2%)	1.66 ± 1.00 (87.5%)
M2	0.35 ± 0.58 (44.0%)	0.43 ± 0.70 (45.0%)	3.09 ± 1.27 (96.6%)

an obvious outlier in other SH albedo regressions as well, slightly decreases the regression slope, bringing it closer to 1.

Surface albedo regression slopes are closer to 1 than the water vapor and atmospheric temperature regression slopes. Most models have similar albedo feedback magnitude between 20- and 100-yr feedbacks or are evenly split between those having larger 100- or 20-yr feedbacks. However, one must be careful in the consideration of the models that do not fit the general pattern of near-equal strength between 20- and 100-yr feedbacks.

c. 20-yr interannual metric versus 100-yr feedbacks

The interannual metric shows less intramodel spread and thus may be a more promising proxy for 100-yr feedbacks. Regression slopes between 20-yr standard deviations of TOA flux anomalies and 100-yr feedbacks, along with their standard errors and significances, are summarized in Table 5. For water vapor and atmospheric temperature, M1 regressions are significantly different from zero at the 95% level, except for NH atmospheric temperature. For surface albedo, the only relationship significantly different from zero is for the SH M2 feedback. Using 10- and 20-yr averages for M2 produces slightly different values but does not alter the results. For the significant relationships, we incorporate the spread to estimate a range of feedbacks based on the reanalysis data.

1) WATER VAPOR AND ATMOSPHERIC TEMPERATURE

Regressions between 20-yr standard deviations and 100-yr feedbacks for water vapor are positive in all regions for both methods but only significant at the 95% level for M1. The NH and SH regressions with M1 feedbacks are shown in Figs. 5a,b and the global plot (not shown) is qualitatively similar. The positive regression slopes indicate that models with larger interannual variability, as measured by the standard deviation

of monthly LW water vapor TOA flux anomalies, tend to have larger 100-yr water vapor feedbacks. Again, the CGCMs visually stand apart from the rest of the models in the NH (Fig. 5a), and excluding those models from the regression yields a steeper slope of 1.91 (99.9%) $\text{W m}^{-2} \text{K}^{-1}$ per W m^{-2} . Note that the magnitude of 100-yr water vapor feedbacks tends to be larger in the SH (and globe) than in the NH.

Regressions of 20-yr standard deviations and 100-yr M1 feedbacks for atmospheric temperature are negative for all regions, implying that models with larger interannual variability tend to have larger negative feedbacks, but the slopes are only significantly different from zero for the globe and SH. The regression slopes for M2 are positive but not significantly different from zero because of large scatter. The NH and SH regressions with M1 feedbacks are shown in Figs. 5c,d, and the global plot (not shown) is qualitatively similar. The magnitudes of SH (and global) 100-yr M1 feedbacks tend to be slightly larger than NH values. The CGCMs behave differently than the rest of the models in both the NH and SH. Excluding these models increases the magnitude of the regression slopes and the significances to -1.70 (99.9%), -1.25 (99.4%), and -1.61 (99.9%) $\text{W m}^{-2} \text{K}^{-1}$ per W m^{-2} for the globe, NH, and SH, respectively, corresponding to slope increases ranging from 0.4 to $0.7 \text{ W m}^{-2} \text{K}^{-1}$ per W m^{-2} .

For both water vapor and atmospheric temperature, significant slopes were obtained only using M1 (radiative kernel-regression method). M2 (difference in beginning and ending values) results in much more scatter of the points around the regression line. This association also holds regardless of the specific regression methodology (model-mean points, all 20-yr slices, or first ensemble member of each model) and is a reflection of the differences between M1 and M2 100-yr feedbacks (Fig. 1), suggesting that the interannual metric is more representative of the year-to-year feedback variability behavior measured by M1 as opposed to the

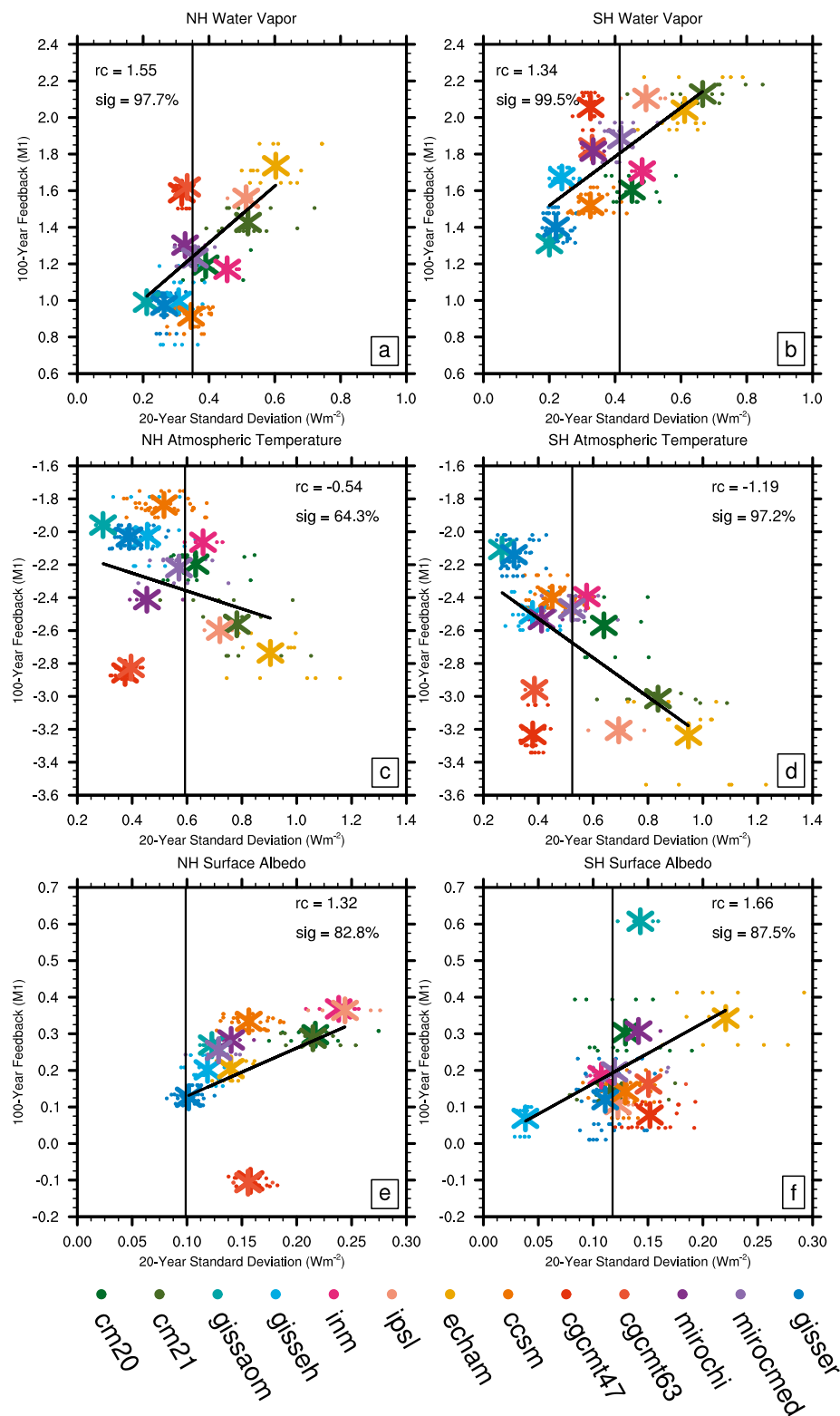


FIG. 5. Comparison of 20-yr standard deviations of monthly TOA flux anomalies (Wm^{-2}) and 100-yr feedbacks ($\text{Wm}^{-2} \text{K}^{-1}$) calculated using M1 for (top) water vapor, (middle) atmospheric temperature, and (bottom) surface albedo in the (left) NH and (right) SH. The vertical lines indicate the reanalysis values. See Fig. 1 for conventions.

TABLE 6. Regression slopes with standard errors (and significances) for 20-yr seasonal cycle amplitudes and 100-yr feedbacks for all three regions and variables and both methods. Units are $\text{W m}^{-2} \text{K}^{-1}$ per natural log of maximum specific humidity divided by minimum specific humidity, $\text{W m}^{-2} \text{K}^{-1}$ per K, and $\text{W m}^{-2} \text{K}^{-1}$ per percent albedo change for water vapor, atmospheric temperature, and surface albedo, respectively. Units for the normalized surface albedo regressions are $\text{W m}^{-2} \text{K}^{-1}$ per percent albedo change. Regressions with $>90\%$ significance are italicized, and those with $>95\%$ significance are in boldface.

Region	Water vapor	Atmospheric temperature	Surface albedo	
Global			Normalized	
M1	2.66 ± 1.89 (81.5%)	-0.19 ± 0.31 (44.4%)	0.00 ± 0.04 (1.3%)	0.08 ± 0.11 (53.8%)
M2	3.48 ± 1.57 (95.2%)	0.02 ± 0.35 (4.9%)	-0.05 ± 0.05 (67.7%)	-0.01 ± 0.14 (6.9%)
NH			Normalized	
M1	1.55 ± 0.78 (92.6%)	-0.22 ± 0.11 (93.0%)	0.04 ± 0.04 (71.0%)	<i>0.48 ± 0.27 (90.3%)</i>
M2	2.09 ± 0.76 (98.2%)	-0.27 ± 0.14 (91.5%)	0.01 ± 0.04 (13.2%)	0.33 ± 0.31 (68.9%)
SH			Normalized	
M1	1.34 ± 1.12 (74.2%)	-0.13 ± 0.13 (67.2%)	-0.01 ± 0.03 (31.6%)	0.03 ± 0.01 (89.9%)
M2	2.68 ± 0.96 (98.2%)	-0.30 ± 0.13 (95.7%)	-0.04 ± 0.03 (77.3%)	0.02 ± 0.02 (53.6%)

century-scale behavior quantified by M2. That is, regressing TOA flux and \overline{T}_{as} anomalies captures more of the year-to-year feedback variability behavior (e.g., ENSO) than M2 does. Since the twentieth-century anomalies do not correspond to a large climate trend, the interannual variation stand out more in the M1 calculation. Note that this result may differ for simulations with a larger climate trend component.

Since there are significant regressions and the reanalysis values fall within the range of modeled values for all regions for water vapor, we calculate the long-term feedback using the reanalysis values and regression relationships (e.g., solid lines in Figs. 5a,b). We estimate values of 1.67 (1.48–1.91), 1.24 (0.99–1.53), and 1.8 (1.6–2.04) $\text{W m}^{-2} \text{K}^{-1}$ for the globe, NH, and SH, respectively, where the ranges correspond to minimum and maximum feedbacks calculated using the best-fit slopes ± 1 standard error, uncertainty in the y intercept, and the spread in the 20-yr model values as an estimate of the observational error. For atmospheric temperature, feedback estimates for the globe, NH, and SH are -2.75 (-2.52 to -2.91), -2.36 (-2.0 to -2.62), and -2.67 (-2.45 to -2.84) $\text{W m}^{-2} \text{K}^{-1}$, respectively.

2) SURFACE ALBEDO

Regressions between 20-yr standard deviations and 100-yr feedbacks for surface albedo are positive; models with large interannual variability tend to have large 100-yr feedbacks. However, only the SH slope using M2 is significantly different from zero. The M1 regression slopes are smaller than M2 regression slopes globally and in the SH, but the reverse is true for the NH. We show NH and SH regressions with M1 feedbacks in Figs. 5e,f for consistency with atmospheric temperature and water vapor.

For the NH, the M1 regressions slopes are over twice the magnitude of the M2 slopes. In both cases, the

CGCMs have negative surface albedo feedbacks as opposed to the other models (Fig. 1h). Excluding CGCMs results in increased significance and a lower slope for M1 (1.12; 99.7%); the M2 slope also decreases.

Using the 100-yr M2 feedback (not shown) instead of M1 (Fig. 5f) for the SH regression, the slope nearly doubles, and the significance increases to $>95\%$. This is due almost entirely to the fact that the magnitude of ECHAM's 100-yr M2 feedback is twice as large as the M1 feedback (see Fig. 1i); most other models closely follow one-to-one correspondence between M1 and M2 feedbacks. Thus, inclusion of ECHAM results in a steeper and more significant slope for M2 regressions with 20-yr SH surface albedo standard deviation, highlighting the high sensitivity of the albedo feedback estimates on model choice.

d. 20-yr intra-annual metric versus 100-yr feedbacks

Colman and Hanson (2013) do not find significant relationships between transient climate change and decadal or interannual and feedbacks, but they find weak positive correlations between seasonal and transient feedbacks for global LW water vapor and NH lapse rate, significant at the 90% level. They also find that seasonal hemispheric feedbacks are closer to those on longer time scales than global feedbacks are. We find significant relationships between the intra-annual metric and long-term water vapor and temperature feedbacks, though we use a fundamentally different metric. Unlike Hall and Qu (2006), we do not find a significant relationship for surface albedo. Table 6 summarizes regression slopes and significances for 20-yr seasonal amplitude and 100-yr feedback comparisons.

1) WATER VAPOR AND ATMOSPHERIC TEMPERATURE

For water vapor, regression slopes of 20-yr seasonal cycle amplitudes versus 100-yr M2 feedbacks are positive

and significantly different from zero at the 95% level in all regions (Fig. 6); large short-term seasonal cycle amplitudes correspond with large long-term feedbacks. M1 regressions are also positive but slightly smaller in magnitude and less significant. Interestingly, this is in contrast to the interannual metric, where we find more significant and larger magnitude slopes using M1 as opposed to M2 (discussed more below). The global regression slope is larger, but it is less significant than the hemispheric regressions because the global seasonal amplitude is essentially a residual between the NH and SH seasonal cycles, which are out of phase. The global plot (not shown) is qualitatively similar to the NH plot (Fig. 6a) since the NH dominates the global water vapor seasonal cycle (note the larger values in Fig. 6a versus Fig. 6b).

For atmospheric temperature, NH and SH relationships between 20-yr seasonal amplitude and 100-yr feedback are negative, indicating again that models with larger 20-yr seasonal cycle amplitudes tend to have larger (more negative) 100-yr feedbacks. Again, there is no significant global relationship. In the NH, regression slopes are similar between feedback calculation methods and significantly different from zero at the 90% level. However, using 10-yr averages instead of 20-yr averages decreases the significance of the NH M2 regression, suggesting that there may be some decadal variability biasing the NH M2 feedbacks. In the SH, the M2 regression is significantly different from zero at the 95% level, but the slope is twice as large as the M1 regression slope, which is not significant. Using 10-yr averages in M2 calculations reduces the significance in the SH by only about 1%. This difference in M2 versus M1 slopes is partly due to the fact that the M2 SH atmospheric temperature feedback in ECHAM is markedly smaller than the M1 feedback, as noted earlier (Fig. 1e). In Figs. 6c,d, we present regressions with M2 feedback calculation for consistency with the water vapor plots.

Based on the relationship between 20-yr seasonal amplitude and 100-yr M2 feedback, we estimate the NH and SH water vapor feedbacks to be 1.70 (1.09–2.31) and 1.95 (1.51–2.40) $\text{W m}^{-2} \text{K}^{-1}$, respectively. Estimated atmospheric temperature feedbacks are -2.58 (-1.03 to -4.11) $\text{W m}^{-2} \text{K}^{-1}$ for the NH and -2.65 (-2.12 to -3.17) $\text{W m}^{-2} \text{K}^{-1}$ for the SH. Because the intramodel spread is small for the intra-annual metric, the uncertainty related to using a single 20-yr observation is responsible for only a small part of this range.

When we use model-mean points for the intra-annual metric (seasonal cycle), M2 results in more significant relationships than M1. However, using all 20-yr slices or just the first ensemble member results in significant ($>99\%$) regressions for both M1 and M2 for all water vapor and NH and SH atmospheric temperature

feedbacks. Since models have different numbers of ensemble members, this suggests that a few models may be responsible for the discrepancy between M1 and M2 feedback regressions. Thus, we do not have confidence that feedback calculation methodology has a consistent influence on the significances of the regressions, though it may matter for some regions.

Figure 1 indicates that M1 feedbacks are generally weaker and have smaller ranges than M2 feedbacks, which tends to decrease the M1 regression slopes compared with M2 (regardless of the regression methodology). For seasonal amplitude, the difference between the M1 and M2 regression slopes can be largely explained by the slope of the M1 to M2 feedback regression (Fig. 1), but this cannot explain the larger M1 slopes for the interannual metric. Many of the differences between methods, however, can be related to anomalous feedback behavior of one or two models. For example, ECHAM has a smaller water vapor feedback in SH for M2 compared with M1. Since ECHAM's interannual metric is on the large side (Fig. 5b), the M2 interannual regression has a smaller slope than M1. The intra-annual metric for ECHAM is on the smaller side (Fig. 6b), resulting in a higher slope for M2.

2) SURFACE ALBEDO

We do not find any significant (95%) relationships between 20-yr surface albedo seasonal amplitude and 100-yr feedbacks for either feedback calculation method (Table 6). Regressions with M2 for the NH and SH are shown in Fig. 7 for consistency with Fig. 6. The global relationship (not shown) is similar to that of the SH. Note that the slopes and significances are highly sensitive to model selection and regression methodology.

Regression slopes using the normalized seasonal amplitude metric are significantly different from zero at the 90% level for both the NH and SH via M1. Regression slopes for M2 are not significant but within standard errors of M1 slopes. The positive relationship in the NH is weaker than but consistent with the results of Hall and Qu (2006). Our results may be weaker because we use all NH points, including both snow cover and sea ice contributions. Hall and Qu (2006) focus on snow cover using only NH land points. In fact, Colman (2013) finds a significant positive relationship between seasonal and climate change for NH snow albedo feedbacks but not for the NH sea ice albedo feedback. Snow feedbacks are more dominant in the NH, where there is more land area, while the SH is more dominated by sea ice changes, which respond to more than local temperature (Rohbock 1980). This may explain why normalizing the SH seasonal amplitude does not improve the relationship with M2 feedbacks.

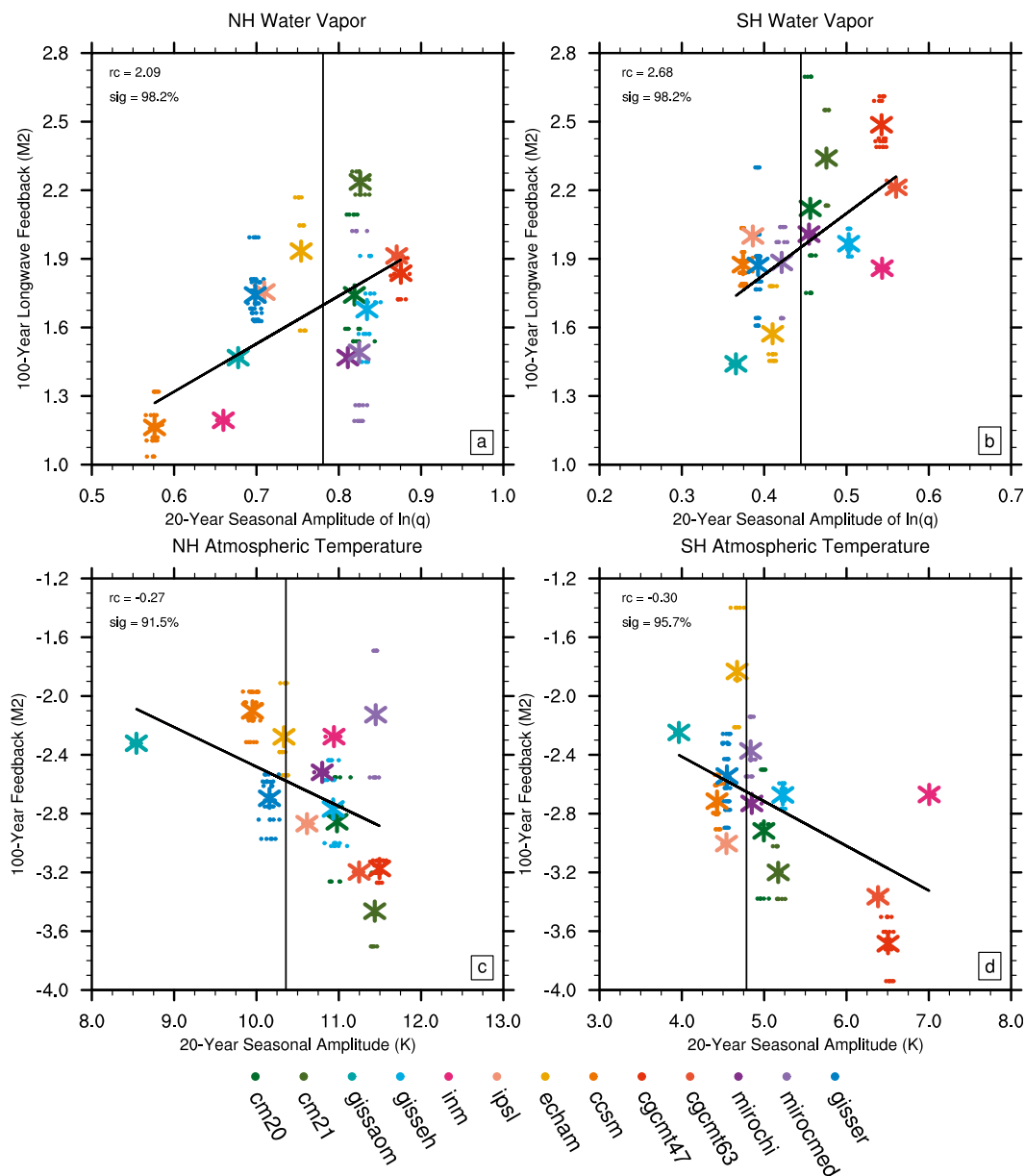


FIG. 6. Comparison of 20-yr seasonal cycle amplitudes and 100-yr feedbacks ($\text{W m}^{-2} \text{K}^{-1}$) calculated using M2 for (top) water vapor and (bottom) atmospheric temperature in the (left) NH and (right) SH. The vertical lines indicate the reanalysis values. See Fig. 1 for conventions.

4. Conclusions

The uncertainty in the climate sensitivity of current GCMs is due in part to the spread of individual climate feedbacks (Bony et al. 2006). One approach to constrain long-term climate sensitivity is to use short-term satellite data. For an observational metric to be useful, both a strong physical relationship and a strong correlation between the short-term metric and long-term feedback are needed. We test three short-term (20 yr) feedback

metrics (20-yr feedbacks, standard deviation of TOA flux anomalies, and amplitude of seasonal cycle of feedback variables) to see how well they represent long-term (100 yr) feedbacks using twentieth-century simulations from 13 GCMs. Several realizations of the short-term metrics are compared with long-term LW water vapor, LW atmospheric temperature, and shortwave surface albedo feedbacks calculated via two methods.

The first method (M1) regresses the deseasonalized monthly TOA flux anomalies against surface air temperature

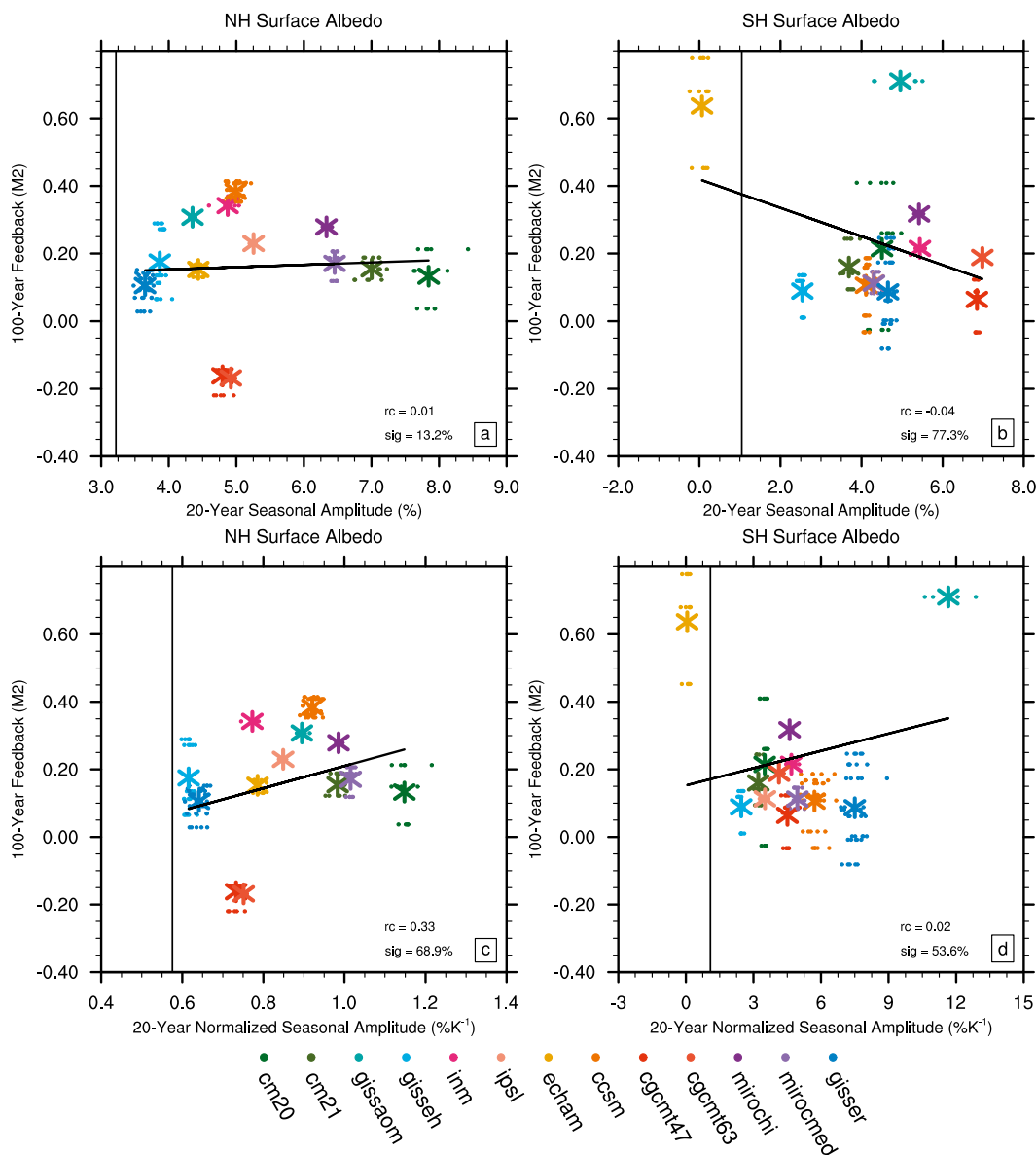


FIG. 7. (top) Comparison of 20-yr seasonal cycle amplitudes and 100-yr feedbacks ($\text{W m}^{-2} \text{K}^{-1}$) calculated using M2 for surface albedo in the (left) NH and (right) SH. (bottom) As in (top), but with seasonal amplitudes normalized by the respective seasonal amplitudes in surface air temperature. The vertical lines indicate the reanalysis values. See Fig. 1 for conventions.

anomalies. For the second method (M2), we divide the difference in the 20-yr TOA flux anomalies between the end and beginning of the run by the corresponding difference in surface air temperature anomalies. While mean feedbacks are similar between the two techniques, water vapor and atmospheric temperature M2 feedbacks are generally larger than M1 feedbacks, and the spread in intramodel means is larger as well. M2 captures only long-term feedback processes, while M1 includes the effects of some shorter-scale variability. The fact that the two methods produce different results

suggests that feedbacks behave differently on different time scales. Surface albedo feedbacks tend to be similar for the two methods, suggesting that nonlinearities in albedo feedbacks with respect to time scale are small. However, some models have very different M1 versus M2 albedo feedbacks. Thus, attention should be paid to the specific methodology used to calculate feedbacks. Note though that we expect the two to converge somewhat as the climate change signal gets larger and dominates shorter-term variability in the twenty-first century.

The short-term intra-annual metric may be useful because it is relatively stationary over the twentieth century as exemplified by the small ($<1\%$) spread (Fig. 3). Thus, it can be closely estimated from a single 20-yr observation, and there will be less observational uncertainty for any relationship found between the seasonal cycle amplitude and 100-yr feedback than for other relationships. The spreads in the 20-yr interannual metric and feedbacks are $>10\%$, and the metrics overlaps considerably among models. For the later calculation of feedback estimates, we use these spreads as error estimates for the uncertainty in reanalysis observations due to the short time period.

The relationships between the 20-yr interannual metric and 100-yr feedbacks are significant for water vapor and global and SH atmospheric temperature but only for feedbacks calculated with M1. On the other hand, the only significant relationship for albedo is in the SH using M2, suggesting processes operating on different time scales.

Relationships between 20-yr seasonal cycle amplitudes and 100-yr feedbacks are significantly different from zero ($>95\%$) for water vapor for all three regions but only for feedback calculation M2. Intra-annual atmospheric temperature relationships with M2 are significant at the 90% level but only for NH and SH. More regression slopes are significant when all 20-yr points or just the first ensemble members of each model are used rather than the model means, but we have less confidence in these methodologies, since they either emphasize particular models or include fewer data. We choose to be conservative and use the regression methodology indicating the lowest significances. We do not find any significant intra-annual relationships for surface albedo for any region, though our NH results using M1 are generally consistent with the results of Hall and Qu (2006).

Many of these relationships depend on the inclusion of a few models that behave differently from the rest. Even though several relationships are improved by excluding, for example, the CGCMs, we cannot exclude any model from the analysis on the grounds that they behave differently from the rest of the models. Without further information, they remain legitimate representations of the climate system. Additionally, the outlier models vary for the particular metrics and regions. Development of robust, objective criteria to select proper models to include is outside the scope of this paper but important (B. M. Sanderson and R. Knutti 2012, personal communication).

Based on the seasonal amplitude regressions, estimates of the NH and SH water vapor feedbacks are 1.70 (1.09 – 2.31) and 1.95 (1.51 – 2.40) $\text{W m}^{-2} \text{K}^{-1}$, respectively, compared with interannual regression estimates of 1.24

(0.99 – 1.35) and 1.8 (1.6 – 2.04) $\text{W m}^{-2} \text{K}^{-1}$. Note that different feedback calculation methods are used for the seasonal (M2) and interannual (M1) estimates. For atmospheric temperature, estimated M2 feedback values from seasonal amplitudes are -2.58 (-1.03 to -4.11) $\text{W m}^{-2} \text{K}^{-1}$ for the NH and -2.65 (-2.12 to -3.17) $\text{W m}^{-2} \text{K}^{-1}$ for the SH; M1 estimates derived from interannual regressions are -2.36 (-2.0 to -2.62) and -2.67 (-2.45 to -2.84) $\text{W m}^{-2} \text{K}^{-1}$. These values are within the range of Table 3 values, so the method produces reasonable feedback estimates. Unfortunately, the range of model behavior results in a large uncertainty. In the case of the interannual metric, natural variability within each model contributes to the uncertainty, reducing our ability to estimate the interannual behavior from one 20-yr time slice. For both the interannual and intra-annual metrics, uncertainty in the intermodel relationships between 20-yr metrics and 100-yr feedbacks contributes to the feedback uncertainty. Thus, this method cannot provide a more constrained feedback estimate without an objective methodology for weighing the importance or correctness of different models.

Finally, we highlight that many of the 20-yr metric versus 100-yr feedback regressions differed based on the 100-yr feedback calculation method. Interannual metric regressions are better with M1 feedbacks because the standard deviation of TOA flux anomalies (interannual metric) characterizes the year-to-year variability, and the M1 feedback calculation is based on the year-to-year covariability of TOA flux anomalies and surface air temperature anomalies. On the other hand, the more significant regression slopes for the intra-annual metrics could be with either M1 or M2, depending on the specific regression methodology and region.

Though some of the differences between methods 1 and 2 can be explained by the difference in feedback magnitudes for a model, this work highlights that care should be taken when drawing conclusions about long-term climate change feedbacks from short-term observed climate feedbacks, since the feedback methods are emphasizing somewhat different time scales and processes. In particular, the feedback method resulting in the more significant regression slopes may change with time, as century-scale anthropogenic warming becomes more important in the twenty-first century.

Acknowledgments. This work was supported by National Science Foundation Grant ATM-0904092. We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI), and the WCRP Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multimodel dataset. Support of this dataset is

provided by the Office of Science, U.S. Department of Energy. We thank Drs. Christoph Thomas, Eric Skillingstad, and Alexandra Jonko for helpful discussion and comments and Dr. Andrew Dessler for sharing data with us. Finally, we thank Dr. Kyle Armour and an anonymous reviewer for their thoughtful and thorough comments.

REFERENCES

- Andrews, T., and P. M. Forster, 2008: CO₂ forcing induces semi-direct effects with consequences for climate feedback interpretations. *Geophys. Res. Lett.*, **35**, L04802, doi:10.1029/2007GL032273.
- Armour, K. C., C. M. Bitz, and G. H. Roe, 2013: Time-varying climate sensitivity from regional feedbacks. *J. Climate*, **26**, 4518–4534.
- Barkstrom, B. R., 1984: The Earth Radiation Budget Experiment (ERBE). *Bull. Amer. Meteor. Soc.*, **65**, 1170–1185.
- Boer, G. J., and B. Yu, 2003: Climate sensitivity and response. *Climate Dyn.*, **20**, 415–429.
- Bony, S., and Coauthors, 2006: How well do we understand and evaluate climate change feedback processes? *J. Climate*, **19**, 3445–3482.
- Cess, R. D., 1974: Radiative transfer due to atmospheric water vapor: Global considerations of the earth's energy balance. *J. Quant. Spectrosc. Radiat. Transfer*, **14**, 861–871, doi:10.1016/0022-4073(74)90014-4.
- Chung, E.-S., B. J. Soden, and B.-J. Sohn, 2010: Revisiting the determination of climate sensitivity from relationships between surface temperature and radiative fluxes. *Geophys. Res. Lett.*, **37**, L10703, doi:10.1029/2010GL043051.
- Colman, R. A., 2013: Surface albedo feedbacks from climate variability and change. *J. Geophys. Res. Atmos.*, **118**, 2827–2834, doi:10.1002/jgrd.50230.
- , and S. B. Power, 2010: Atmospheric radiative feedbacks associated with transient climate change and climate variability. *Climate Dyn.*, **34**, 919–933, doi:10.1007/s00382-009-0541-8.
- , and B. J. McAvaney, 2011: On tropospheric adjustment to forcing and climate feedbacks. *Climate Dyn.*, **36**, 1649–1658, doi:10.1007/s00382-011-1067-4.
- , and L. I. Hanson, 2013: On atmospheric radiative feedbacks associated with climate variability and change. *Climate Dyn.*, **40**, 475–492, doi:10.1007/s00382-012-1391-3.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.
- Dessler, A. E., 2010: A determination of the cloud feedback from climate variations over the past decade. *Science*, **330**, 1523–1527, doi:10.1126/science.1192546.
- , 2013: Observations of climate feedbacks over 2000–10 and comparisons to climate models. *J. Climate*, **26**, 333–342.
- , and S. Wong, 2009: Estimates of the water vapor climate feedback during El Niño–Southern Oscillation. *J. Climate*, **22**, 6404–6412.
- , Z. Zhang, and P. Yang, 2008: Water-vapor climate feedback inferred from climate fluctuations, 2003–2008. *Geophys. Res. Lett.*, **35**, L20704, doi:10.1029/2008GL035333.
- Flanner, M. G., K. M. Shell, M. Barlage, D. K. Perovich, and M. A. Tschudi, 2011: Radiative forcing and albedo feedback from the Northern Hemisphere cryosphere between 1979 and 2008. *Nat. Geosci.*, **4**, 151–155, doi:10.1038/ngeo1062.
- Forster, P. M., and M. Collins, 2004: Quantifying the water vapor feedback associated with post-Pinatubo global cooling. *Climate Dyn.*, **23**, 207–214, doi:10.1007/s00382-004-0431-z.
- , and J. M. Gregory, 2006: The climate sensitivity and its components diagnosed from earth radiation budget data. *J. Climate*, **19**, 39–52.
- Gregory, J., and M. Webb, 2008: Tropospheric adjustment induces a cloud component in CO₂ forcing. *J. Climate*, **21**, 58–71.
- , and Coauthors, 2004: A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.*, **31**, L03205, doi:10.1029/2003GL018747.
- Hall, A., and X. Qu, 2006: Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophys. Res. Lett.*, **33**, L03502, doi:10.1029/2005GL025127.
- Held, I. M., and K. M. Shell, 2012: Using relative humidity as a state variable in climate feedback analysis. *J. Climate*, **25**, 2578–2582.
- Jonko, A. K., K. M. Shell, B. M. Sanderson, and G. Danabasoglu, 2012: Climate feedbacks in CCSM3 under changing CO₂ forcing. Part I: Adapting the linear radiative kernel technique to feedback calculations for a broad range of forcings. *J. Climate*, **25**, 5260–5272.
- Knutti, R., 2010: The end of model democracy? *Climatic Change*, **102**, 395–404, doi:10.1007/s10584-010-9800-2.
- , G. A. Meehl, M. R. Allen, and D. A. Stainforth, 2006: Constraining climate sensitivity from the seasonal cycle in surface temperature. *J. Climate*, **19**, 4224–4233.
- Lin, B., Q. Min, W. Sun, Y. Hu, and T. Fan, 2011: Can climate sensitivity be estimated from short-term relationships of top-of-atmosphere net radiation and surface temperature? *J. Quant. Spectrosc. Radiat. Transfer*, **112**, 177–181, doi:10.1016/j.jqsrt.2010.03.012.
- Masson, D., and R. Knutti, 2013: Predictor screening, calibration, and observational constraints in climate model ensembles: An illustration using climate sensitivity. *J. Climate*, **26**, 887–898.
- Randall, D. A., and Coauthors, 2007: Climate models and their evaluation. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds., Cambridge University Press, 589–662.
- Robock, A., 1980: The seasonal cycle of snow cover, sea ice and surface albedo. *Mon. Wea. Rev.*, **108**, 267–285.
- Sanderson, B. M., and K. M. Shell, 2012: Model-specific radiative kernels for calculating cloud and noncloud climate feedbacks. *J. Climate*, **25**, 7606–7624.
- Shell, K., J. Kiehl, and C. Shields, 2008: Using the radiative kernel technique to calculate climate feedbacks in NCAR's Community Atmospheric Model. *J. Climate*, **21**, 2269–2282.
- Soden, B. J., 1997: Variations in the tropical greenhouse effect during El Niño. *J. Climate*, **10**, 1050–1055.
- , and I. M. Held, 2006: An assessment of climate feedbacks in coupled ocean–atmosphere models. *J. Climate*, **19**, 3354–3360.
- , —, R. A. Colman, K. M. Shell, J. T. Kiehl, and C. A. Shields, 2008: Quantifying climate feedbacks using radiative kernels. *J. Climate*, **21**, 3504–3520.
- Thompson, S. L., and S. G. Warren, 1982: Parameterization of outgoing infrared radiation derived from detailed radiative calculations. *J. Atmos. Sci.*, **39**, 2667–2680.
- Trenberth, K. E., J. T. Fasullo, C. O'Dell, and T. Wong, 2010: Relationships between tropical sea surface temperatures and top-of-atmosphere radiation. *Geophys. Res. Lett.*, **37**, L03702, doi:10.1029/2009GL042314.
- Winton, M., K. Takahashi, and I. M. Held, 2010: Importance of ocean heat uptake efficacy to transient climate change. *J. Climate*, **23**, 2333–2344.