



RESEARCH ARTICLE

10.1029/2023JD039427

Key Points:

- Individual short-term cloud feedback components in climate models are assessed by comparing against satellite observations
- Model biases are mainly driven by tropical marine low-cloud, high-cloud altitude, and extratropical high-cloud optical depth feedbacks
- Skill in simulating short-term cloud feedbacks is not correlated with skill in simulating long-term cloud feedbacks

Supporting Information:

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

Correspondence to:

L.-W. Chao,
chao5@llnl.gov

Citation:

Chao, L.-W., Zelinka, M. D., & Dessler, A. E. (2024). Evaluating cloud feedback components in observations and their representation in climate models. *Journal of Geophysical Research: Atmospheres*, 129, e2023JD039427. <https://doi.org/10.1029/2023JD039427>

Received 8 JUN 2023

Accepted 4 JAN 2024

Evaluating Cloud Feedback Components in Observations and Their Representation in Climate Models

Li-Wei Chao^{1,2} , Mark D. Zelinka² , and Andrew E. Dessler¹ 

¹Department of Atmospheric Sciences, Texas A&M University, College Station, TX, USA, ²Lawrence Livermore National Laboratory, Livermore, CA, USA

Abstract This study quantifies the contribution of individual cloud feedbacks to the total short-term cloud feedback in satellite observations over the period 2002–2014 and evaluates how they are represented in climate models. The observed positive total cloud feedback is primarily due to positive high-cloud altitude, extratropical high- and low-cloud optical depth, and land cloud amount feedbacks partially offset by negative tropical marine low-cloud feedback. Seventeen models from the Atmosphere Model Intercomparison Project of the sixth Coupled Model Intercomparison Project are analyzed. The models generally reproduce the observed moderate positive short-term cloud feedback. However, compared to satellite estimates, the models are systematically high-biased in tropical marine low-cloud and land cloud amount feedbacks and systematically low-biased in high-cloud altitude and extratropical high- and low-cloud optical depth feedbacks. Errors in modeled short-term cloud feedback components identified in this analysis highlight the need for improvements in model simulations of the response of high clouds and tropical marine low clouds. Our results suggest that skill in simulating interannual cloud feedback components may not indicate skill in simulating long-term cloud feedback components.

Plain Language Summary Cloud feedback—the radiative response of clouds to changes in temperature—is determined by the contributions from various cloud types and exhibits large uncertainty. Here, we use satellite observations to evaluate how the individual cloud feedback components in response to interannual variability are represented in the latest generation of climate models. The total cloud feedback is positive in the observations, mainly driven by changes in the altitude of high clouds, changes in the cloud cover over land, and changes in the reflectivity of high and low clouds over extratropical regions. The climate models were driven by observed sea-surface temperature, sea-ice conditions, and radiative forcing, so they can be directly compared to the observations. We found the models simulate a total cloud feedback that agrees with the observed positive total cloud feedback in general. However, the models consistently overestimate the decrease of cloud amount over land and in oceanic regions of tropical descent and consistently underestimate the increase of high-cloud altitude and the decrease of extratropical high- and low-cloud optical depth. Models that better simulate cloud feedback in response to short-term fluctuations do not perform better in simulating long-term cloud feedback under global warming.

1. Introduction

As the Earth becomes warmer due to human activities, cloud properties change and modulate the energy flow entering and leaving the Earth, resulting in a radiative feedback. Several lines of evidence indicate that the net cloud feedback is positive and amplifies warming (Sherwood et al., 2020, hereafter **S20**; IPCC, 2021). The total cloud feedback is the net effect of many individual components involving a variety of cloud types and processes that often offset one another. Therefore, evaluating climate models' performance in accurately simulating individual cloud feedback components is crucial to reveal sources of errors in modeled cloud feedback, which provides insights into the role of clouds in a warming climate and the representation of clouds in climate models.

To perform model evaluation, we require some form of ground-truth for comparison. Zelinka et al. (2022, hereafter **Z22**) evaluated cloud feedbacks in response to long-term greenhouse gas forcing in global climate models (GCMs) by comparing them against the expert judgment values in **S20**. **S20** identified a set of six well-established cloud feedbacks, hereafter referred to as assessed cloud feedbacks (see Table 1 for their definitions), and quantified their values and uncertainties using multiple independent sources including GCMs, observations, high-resolution models, and theory.

© 2024. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Table 1
The Definition of Individual Cloud Feedback Components Used in This Study

Name	Components	Surface	Region
Assessed Cloud Feedbacks			
High-Cloud Altitude	High-cloud altitude	All	90°S–90°N
Tropical Marine Low-Cloud	Low-cloud amount and optical depth	Ocean descent	30°S–30°N
Tropical Anvil Cloud	High-cloud amount and optical depth	Ocean ascent	30°S–30°N
Land Cloud Amount	High- and low-cloud amount	Land	90°S–90°N
Mid-Latitude Marine Low-Cloud Amount	Low-cloud amount	Ocean	30–60°N/S
High-Latitude Low-Cloud Optical Depth	Low-cloud optical depth	All	40–70°N/S
Unassessed Cloud Feedbacks			
30–90 High-Cloud Optical Depth	High-cloud optical depth	All	30–90°N/S
30–40/70–90 Low-Cloud Optical Depth	Low-cloud optical depth	All	30–40/70–90°N/S
Tropical Land Cloud Optical Depth	High- and low-cloud optical depth	Land	30°S–30°N
Tropical Marine Descent High-Cloud	High-cloud amount and optical depth	Ocean descent	30°S–30°N
Tropical Marine Ascent Low-Cloud	Low-cloud amount and optical depth	Ocean ascent	30°S–30°N
Others	High-cloud amount	Ocean	30–90°N/S
	Low-cloud amount	Ocean	60–90°N/S
	Low-cloud altitude	All	90°S–90°N
Residuals	Obscuration Covariance	All	90°S–90°N
	Zelinka Decomposition Residual	All	90°S–90°N
	Bony Omega Space Residual	–	–
Sum of Assessed Cloud Feedbacks	–	–	–
Sum of Unassessed Cloud Feedbacks	–	–	–
Total Cloud Feedback	–	All	90°S–90°N

This approach has some limitations. First, results from GCMs were used to inform the quantitative values of several expert-assessed long-term cloud feedbacks, which reduced their independence as a gauge of GCM performance. Second, S20 approximated the total cloud feedback as the sum of six components, and hence the model evaluation conducted in Z22 was limited to these six assessed cloud feedback values. Other cloud feedback components not discussed in S20 could be large in nature and/or strongly biased in models. For instance, Z22 showed that the inter-model spread of the feedbacks that were not assessed in S20 is nonnegligible and that their sum is a major contributor to the increase in multi-model mean cloud feedback between the latest generation of GCMs and the previous generation (Zelinka et al., 2022). Third, for some feedbacks, such as the tropical anvil cloud area feedback, S20's assessed value relied heavily on a single observational study, calling into question its utility as a benchmark for the long-term feedback.

Finally, S20's assessed feedback values were derived using expert judgment, which is informed by the literature available at the time. This means, for example, that the recent spate of evidence of weak trade cumulus feedbacks (e.g., Cesana & Del Genio, 2021; Myers et al., 2021; Vial et al., 2023; Vogel et al., 2022)—which would presumably decrease the assessed tropical marine low cloud feedback—is not accounted for since these studies were published after S20. Still, we note that other recent studies support the overall cloud feedback values assessed in S20 (Ceppi & Nowack, 2021) and opinions differ on whether the assessed values of low cloud feedbacks should indeed be updated in light of the aforementioned studies (Sherwood & Forest, 2023).

In this study, we evaluate models against a different source of evidence: satellite observations. Here we estimate short-term cloud feedbacks in response to interannual variability inferred from the observations and evaluate the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al., 2016) models' ability to simulate the individual components of the cloud feedback. This study not only investigates the assessed cloud feedbacks but also analyzes the cloud feedback components that are not explicitly identified in S20.

The comparison against observations provides an in-depth evaluation that is aligned with real-world climate conditions. While avoiding the limitations of Z22 noted above, we acknowledge several caveats with our analysis. First, the cloud feedback in response to short-term climate fluctuations may differ from the long-term feedback relevant to climate sensitivity. This is not a limitation per se, since it is important to assess whether models can simulate cloud responses on all timescales, not just in response to CO₂ induced global warming. Furthermore, in Section 3.8 we explicitly assess whether model skill in simulating short-term cloud feedbacks implies skill in simulating long-term cloud feedbacks. Second, the cloud radiative response to forcing agents (e.g., aerosol-cloud interactions)—which may vary considerably across models and between models and observations—is not explicitly accounted for. We expect this to be a minor effect, as we quantify the response of clouds to interannual fluctuations in surface temperature which are unlikely to have a coincident aerosol signature (unlike, for example, if our analysis relied on secular trends during this time). Indeed, we derive statistically indistinguishable feedbacks in atmospheric simulations with and without historical forcing variations, confirming that forcing variations make a minor contribution to the interannual cloud feedback during 2002–2014 (not shown). Third, the diagnosis of cloudiness is similar but not identical in observations and models, as the observations measure two daytime samples per day (Aqua and Terra satellite overpasses) while the model's cloud simulator output is based on cloudiness from all sunlit timesteps. We anticipate this to be a minor effect as well since studies have shown that temporal sampling has limited effect on cloudiness estimates at the global scale (e.g., Pincus et al., 2012). Finally, our analysis is subject to potential observational uncertainty arising from satellite instrument or calibration errors. However, the CERES and MODIS instrument calibration is stable as shown in previous studies (e.g., CERES Science Team, 2022; Loeb et al., 2016; Yue et al., 2017), and recent studies found that the cloud feedbacks estimated from four different satellite data sets are consistent, suggesting that the measurement error is small (Myers et al., 2021; Scott et al., 2020).

2. Data and Methods

2.1. Satellite Observations

The cloud feedback is estimated as the change in cloud-induced top-of-atmosphere (TOA) radiation anomalies (ΔR_{cloud}) per degree of global average surface temperature change. ΔR_{cloud} is calculated as the product of cloud fraction anomalies and cloud radiative kernels, both of which are a function of cloud top pressure (CTP) and cloud visible optical depth (τ). Two data sets of observed monthly cloud fraction are considered in this study: the Clouds and the Earth's Radiant Energy System (CERES) Flux By Cloud Type (FBCT) Ed4A product (Sun et al., 2022) and the Moderate Resolution Imaging Spectroradiometer (MODIS) COSP Level-3 product (Pincus et al., 2023), both available starting July 2002. Both products provide cloud properties from the MODIS instrument on Terra and Aqua spacecrafts but using different cloud retrieval algorithms, Minnis et al. (2011) for CERES-FBCT and Platnick et al. (2017) for MODIS-COSP. To be consistent with CERES-FBCT product, which only provides cloud properties with an optical depth above 0.25, we discard the first τ bin (0–0.3) in the MODIS-COSP product, such that both have seven CTP bins and six τ bins. MODIS COSP reports cloud fraction joint histograms for fully cloudy and partly cloudy scenes, and we sum them to obtain the total cloud fraction following previous studies (e.g., Wall et al., 2022). The conclusion of the paper and the cloud feedback values are consistent when including or excluding partly cloudy cloud fraction histograms. Note that both data sets utilize only daytime observations to ensure consistency of cloud-type classification (Pincus et al., 2023; Sun et al., 2022).

Cloud radiative kernels (Zhou et al., 2013) quantify the sensitivity of TOA radiation to CTP- and τ -stratified cloud fraction changes. The first τ bin (0–0.3) in the radiative kernels is also discarded for consistency, yielding six τ bins in all data sets. Multiplying the kernel by the change in cloud fraction histogram yields the TOA radiative impact of each cloud type. The total ΔR_{cloud} is then the summation of ΔR_{cloud} for the 42 cloud categories. ΔR_{cloud} consists of shortwave (SW) and longwave (LW) components and is defined as positive downward (i.e., planetary heating).

The cloud feedback is calculated separately using CERES-FBCT and MODIS-COSP products. These are then averaged to produce a single combined observational estimate of the feedback, with 90% and 66% confidence intervals calculated as the square root of the sum of the square of uncertainties from the two data sets, divided by two. The uncertainties are averaged under the assumption that the two data sets have uncorrelated errors. Uncertainty intervals are very similar if calculated using a bootstrapping method that does not rely on a priori assumptions about the distribution of uncertainties from the two data sets. Feedbacks estimated using the two

observational data sets have quantitative differences that are discussed further in Section 3.6, but the main conclusions of the paper are unchanged if using either data set in isolation.

2.2. CMIP6 Models

One of the main goals of this study is to evaluate climate models' performance in simulating cloud feedbacks measured from real-world climate variability. We analyze Atmospheric Model Intercomparison Project (AMIP) simulations in CMIP6, where the models are forced with anthropogenic and natural forcing and observed fields of sea surface temperatures and sea ice concentrations. The use of AMIP models ensures a fair comparison with the observations and allows us to focus on the representation of cloud radiative response in the atmospheric model. This is especially important given the substantial influence of sea surface temperature patterns on the sign and strength of individual feedbacks, which is discussed in Section 3.7.

The calculation of individual cloud feedbacks in models requires cloud fraction anomalies stratified by cloud top pressure and cloud optical depth, which are produced from the ISCCP simulator (Klein & Jakob, 1999; Webb et al., 2001). This limits our analysis to 17 climate models (listed in Table S1 in Supporting Information S1) that participated in the Cloud Feedback Model Intercomparison Project (CFMIP; Webb et al., 2017). We calculated modeled cloud feedback from all available members from individual climate models. Since the MODIS instrument cannot retrieve thin clouds with $\tau < 0.25$, we exclude the first τ bin (0–0.3) in modeled cloud fraction for consistency. The modeled cloud feedback components computed excluding or including the first τ bin are highly correlated ($r = 0.96$), suggesting that excluding the first τ bin does not significantly affect the study's key findings (Figure S1 in Supporting Information S1).

2.3. Decomposition of Cloud Feedback Into Contributions From Individual Cloud Types

To decompose the contributions of different cloud types to total ΔR_{cloud} , we follow Z22. ΔR_{cloud} is first separated into contributions from low clouds (CTP > 680 hPa) and high clouds (CTP ≤ 680 hPa). Within each regime, we further break down ΔR_{cloud} into the components that are solely due to changes in cloud amount, altitude, and optical depth, along with a residual term (Zelinka et al., 2012, 2013, 2016).

Changes in low clouds observed by passive sensors may be due to changes in high clouds that mask or reveal low clouds. To avoid this ambiguity, we separate low cloud feedback into three components: that due to changes in unobscured low clouds, that due to changes in obscuration by upper-level clouds, and a small covariance term (see Equation 4 in Scott et al. (2020) and also Myers et al. (2021)). The unobscured low cloud component is treated as the true low cloud feedback and is further decomposed into cloud amount, altitude, and optical depth components. The obscuration component refers to changes in the fraction of low-level clouds that occur exclusively because of variations in the fraction of higher-level clouds that can cover or uncover the lower clouds beneath them. As in Z22, this component is combined with the high-cloud amount component, treating them as a single entity.

In the tropics (30°N–30°S), stratocumulus and trade cumulus dominate the marine descent regions, while anvil clouds dominate the marine ascent regions. To distinguish the clouds in different large-scale circulation regimes, we compute tropical marine ascent/descent feedbacks separately following Bony et al. (2004). Cloud radiative kernel and cloud fraction fields over the tropical ocean are first aggregated into 10-hPa/day wide bins of monthly averaged 500-hPa vertical velocity (ω_{500}) with area-weighting. ΔR_{cloud} is then calculated in ω_{500} space rather than geographic space and is further decomposed into cloud amount, altitude, and optical depth components in ω_{500} space. Tropical ascent and descent components are computed by summing over the bins where $\omega_{500} < 0$ and $\omega_{500} \geq 0$, respectively.

Lastly, the individual components of the cloud feedback are estimated by regressing ΔR_{cloud} components at each grid point (or each ω_{500} bin) against the global average surface air temperature anomaly. Both SW and LW components are calculated, and their sum is what we report below. For observed cloud feedbacks, the surface temperature and ω_{500} are obtained from ERA5 reanalysis (Hersbach et al., 2020). Monthly average data are used for all calculations, and the anomalies are deviations from the climatological seasonal cycle. While the cloud feedback is defined as the changes of cloud radiative anomalies per degree of global temperature change, this does not imply that global average surface temperature directly controls changes in clouds in every location. Rather, individual

cloud responses at a given location are responding to their proximate controlling factors (e.g., inversion strength, circulation, and humidity), which themselves change as global temperature changes. All feedbacks reported in this study are scaled by the fractional area of the globe over which they operate, so they represent relative contributions to global mean total cloud feedback. We analyze the cloud feedback over the globe to be consistent with the definition in S20 and Z22. Confining the analysis to 60°N–60°S where FBCT sampling is better yields similar results (Table S4 in Supporting Information S1).

In summary, in both models and observations we compute LW, SW, and net cloud feedbacks at each geographic location over both land and ocean. Over the tropical oceans, we additionally distinguish feedbacks within regimes of ascent and descent. All feedbacks are broken down into amount, altitude, and optical depth components for both high and low clouds (with careful accounting of obscuration effects). While this allows for a wide variety of possible groupings in which to compare modeled and observed feedbacks, we adhere to the groupings used in Z22 since—in the case of the assessed feedbacks—they represent relatively well-studied feedbacks with known governing mechanisms.

2.4. Significance Test for Biased Cloud Feedback Components

This study identifies biased cloud feedback components from AMIP simulations by considering the uncertainty in both observations and models. We perform all significance tests using a bootstrapping method, as it avoids a priori assumptions about the shape of the data distribution. First, the distribution of each cloud feedback component is estimated through Monte Carlo sampling (e.g., Chao & Dessler, 2021; Dessler & Forster, 2018; Donohoe et al., 2014). For each sample, we calculate the feedback by randomly selecting a number of points (with replacement) from the time series equal to the effective sample size adjusted for temporal autocorrelation (Santer et al., 2000), and conducting linear regression. This process is repeated 100,000 times for both observations and individual model members. Next, a new distribution representing the difference between modeled and observed feedbacks is generated through Monte Carlo sampling. Biased cloud feedback components are identified if the differences are statistically significant (i.e., significantly depart from zero) at 90% or 66% confidence intervals.

2.5. Model Evaluation Strategy

When evaluating models' short-term cloud feedbacks against observations, one has to carefully account for uncertainty in the feedback estimates arising from meteorological noise. As shown below, this noise can be substantial for feedbacks derived from interannual fluctuations due to meteorological fluctuations that are not systematically dependent on global surface temperature, especially when using only 12.5 years of data. This means that for the following analyses, we do not compare central feedback estimates between models and observations but rather the distributions of both following the aforementioned bootstrapping method. Moreover, we consider all available realizations of the individual models' AMIP simulations, which provides another indication of how much weather noise can affect the feedback estimates. Because inter-member differences in central feedback estimates within any given model can be comparable to the biases with respect to observations, this limits our ability to conclusively determine that any particular model is biased. This is further complicated by the fact that many models provide only a single realization and could therefore be unfairly judged as biased or unbiased owing to a small sample size. We therefore focus hereafter on identifying biases that are systematic across models and that are relatively robust across the multiple realizations of any given model.

3. Evaluation of Modeled Cloud Feedbacks Against Observations

3.1. Comparisons of Total Cloud Feedback in Models Against Observations

Figure 1a shows the short-term total cloud feedback inferred from the observations and from CMIP6 AMIP simulations. We first focus on the period of July 2002 to December 2014, determined by the overlap between the satellite observations and CMIP6 AMIP experiments. All models have at least one member agreeing with the observations within the 90% confidence intervals and only one model member has a value that is biased at 90% confidence. We note that the inter-member spread arising from internal meteorological variations could be large for specific models (one standard deviation of ensemble members varies across models from 0.10 to 0.57 W/m²/K). These internal variations remain in longer analyzed periods (e.g., 20-year cloud feedback in AMIP simulation) and in other simulations (e.g., 12-year and 20-year segments of piControl simulation) (not shown).

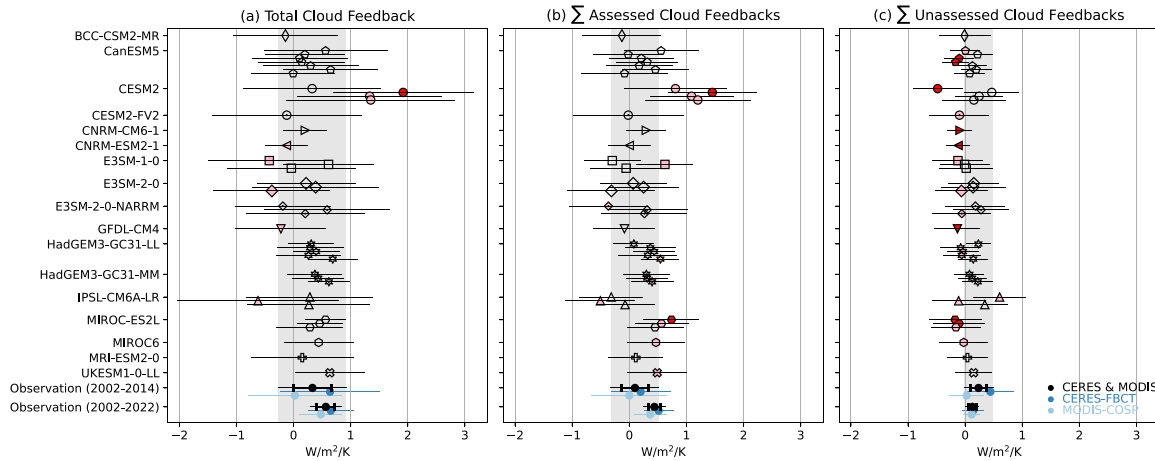


Figure 1. Short-term (a) total, (b) sum of assessed, and (c) sum of unassessed cloud feedbacks inferred from satellite observations from 2002 to 2014 and 2002 to 2022, and CMIP6 AMIP simulations between 2002 and 2014. For the observations (CERES and MODIS), the best-estimated value and 66% (90%) confidence intervals are indicated as the black circles and thick (thin) error bars. Gray shading represents 90% confidence intervals for observations from 2002 to 2014. The observations are the average of CERES-FBCT (blue) and MODIS-COSP (light blue) as described in the text. The individual model ensemble member values are shown as markers. Components that are significantly biased relative to the observations at 90% (66%) confidence are denoted with red (pink) markers.

We also compare the spatial pattern of total cloud feedback between observations and models. CMIP6 AMIP multi-model mean qualitatively reproduces some spatial features of observed total cloud feedback, such as the positive feedback over the northeastern Pacific, equatorial Pacific, and southwestern Pacific (Figure S2 in Supporting Information S1). The multi-model mean is obtained by averaging the first member of each model to ensure consistent treatment with all models, but the results are similar if including multiple members.

The total cloud feedback is further decomposed into the individual assessed feedback components that are closely tied to well-studied processes (i.e., assessed cloud feedbacks that are discussed individually in Section 3.2), and the feedback components that are not assessed in S20 (i.e., unassessed cloud feedbacks that are discussed individually in Section 3.3). The description of each cloud feedback component is provided in Table 1. Table 2 summarizes the values of cloud feedback components estimated from the observations, along with the number of models significantly differing from the observations. The values for individual models and members are listed in Tables S1 and S2 in Supporting Information S1. The breakdown is rationalized in Z22 (see their Figure S3), and the summation of all cloud feedback components equals the total cloud feedback, with no overlap in their definitions. Most of the variance in the total cloud feedback is accounted for by the sum of assessed components. All models except CESM2 and MIROC-ES2L simulate the sum of assessed cloud feedbacks that match well with the observations. However, for the sum of unassessed cloud feedbacks, the majority of models (13 of 17 models) have at least one member that exhibits significant bias at 66% confidence, and several of these are biased at 90% confidence. All of these significant biases are negative except one realization in IPSL-CM6A-LR. Note that there are multiple ways to interpret the number of models that are biased, and in this study we adopt a more stringent approach to minimize the risk of underestimating biased models.

3.2. Comparisons of Assessed Cloud Feedback Components

In this section, we evaluate models' ability to capture the individual assessed feedback components (Figure 2). The observational value of the high-cloud altitude feedback is $+0.40 \pm 0.22$ W/m²/K (90% confidence intervals; Figure 2a). The altitude of high-cloud tops increases as temperature increases, which hinders the system's ability to radiatively damp warming and results in positive feedback (Hartmann & Larson, 2002; Zelinka & Hartmann, 2010). Models are systematically low-biased, with at least one member from every model significantly underestimating the feedback magnitude at 66% confidence, and many members underestimating it at 90% confidence. Not a single member of any model overestimates the observed value of the high cloud altitude feedback. Inclusion of clouds in the thinnest optical depth category only slightly increases the multi-model mean value (i.e., from $+0.02 \pm 0.12$ W/m²/K to $+0.06 \pm 0.13$ W/m²/K (one standard deviation)), suggesting that the bias is not simply arising from ignoring the thinnest clouds in models. Note that the magnitude of high-cloud altitude

Table 2

The Individual Cloud Feedback Components and 90% Confidence Intervals ($W/m^2/K$) Inferred From Satellite Observations (Combined CERES-FBCT and MODIS-COSP) From 2002 to 2022 and 2002 to 2014, and the Number of Models (Out of 17 Models) Having at Least One Member That is Significantly Biased Low or High at 90% (66%) Confidence With Respect to the Observations (2002 to 2014)

Name	Observations (2002–2022)	Observations (2002–2014)	Number of models with significant low/high bias
Assessed Cloud Feedbacks			
High-Cloud Altitude	0.42 ± 0.09	0.40 ± 0.22	12/0 (17/0)
Tropical Marine Low-Cloud	-0.04 ± 0.14	-0.51 ± 0.28	0/10 (0/14)
Tropical Anvil Cloud	0.01 ± 0.05	-0.06 ± 0.09	5/2 (9/2)
Land Cloud Amount	0.00 ± 0.05	0.14 ± 0.13	0/8 (0/11)
Mid-Latitude Marine Low-Cloud Amount	-0.00 ± 0.05	0.04 ± 0.12	3/1 (4/5)
High-Latitude Low-Cloud Optical Depth	0.05 ± 0.04	0.09 ± 0.08	11/0 (17/0)
Unassessed Cloud Feedbacks			
30–90 High-Cloud Optical Depth	0.04 ± 0.05	0.20 ± 0.10	12/0 (17/0)
30–40/70–90 Low-Cloud Optical Depth	0.06 ± 0.02	0.09 ± 0.05	15/0 (16/0)
Tropical Land Cloud Optical Depth	0.02 ± 0.02	0.04 ± 0.06	3/0 (7/3)
Tropical Marine Descent High-Cloud	0.04 ± 0.06	0.11 ± 0.13	2/0 (4/0)
Tropical Marine Ascent Low-Cloud	0.03 ± 0.05	-0.11 ± 0.10	0/5 (0/9)
Others	-0.04 ± 0.03	-0.00 ± 0.07	0/2 (1/8)
30–90 Marine High-Cloud Amount	0.00 ± 0.03	0.04 ± 0.06	
60–90 Marine Low-Cloud Amount	-0.02 ± 0.02	-0.04 ± 0.04	
Low-Cloud Altitude	-0.03 ± 0.01	0.00 ± 0.01	
Residuals	-0.02 ± 0.05	-0.09 ± 0.13	1/2 (1/6)
Obscuration Covariance	0.01 ± 0.01	0.01 ± 0.02	
Zelinka Decomposition Residual	-0.01 ± 0.01	-0.03 ± 0.01	
Bony Omega Space Residual	-0.03 ± 0.05	-0.07 ± 0.12	
Sum of Assessed Cloud Feedbacks	0.44 ± 0.19	0.10 ± 0.42	0/2 (2/7)
Sum of Unassessed Cloud Feedbacks	0.13 ± 0.11	0.23 ± 0.25	6/0 (13/1)
Total Cloud Feedback	0.57 ± 0.27	0.33 ± 0.59	0/1 (5/1)

feedback is sensitive to choice of cloud radiative kernel, but models are significantly biased low with respect to the observations regardless of which radiative kernel is used (Figure S9a in Supporting Information S1). Sensitivity of results to choice of cloud radiative kernel is discussed further in Section 3.5. This bias is not seen in the long-term high-cloud altitude feedback in Z22, where the CMIP6 multi-model mean matches the expert-assessed value well, and 11 of 12 models have best estimates that fall within expert judgment's 90% confidence intervals. Note that the cloud radiative kernels used in Z22 (Zelinka et al., 2012) and in this study (Zhou et al., 2013) are comparable as they are constructed using the same methodology, including the same radiative transfer model. The multi-model mean spatial pattern of high-cloud altitude feedback inferred from CMIP6 AMIP models is similar to the major observed spatial features, notably over the Pacific Ocean, but models fail to reproduce positive values observed over several regions including the Indian Ocean, North Atlantic, Eurasia, and Australia (Figures S3a and S3e in Supporting Information S1).

Six models (CanESM5, E3SM-1-0, GFDL-CM4, IPSL-CM6A-LR, MIROC6, and MRI-ESM2-0) show at least one member with moderate to strong negative high-cloud altitude feedback. For these six models, the response of cloud fraction from the ISCCP simulator indicates a decrease in high-cloud altitude with increasing temperature, even when including clouds in the thinnest optical depth category. In contrast, cloud fraction profiles taken directly from the models (cl) or from the CALIPSO simulator (clcalipso) show an increase in high-cloud altitude. Future work is needed to reconcile the inconsistent high-cloud responses between various cloud fraction diagnostics in AMIP simulations, which are not seen in corresponding long-term warming simulations (i.e., abrupt-4xCO₂).

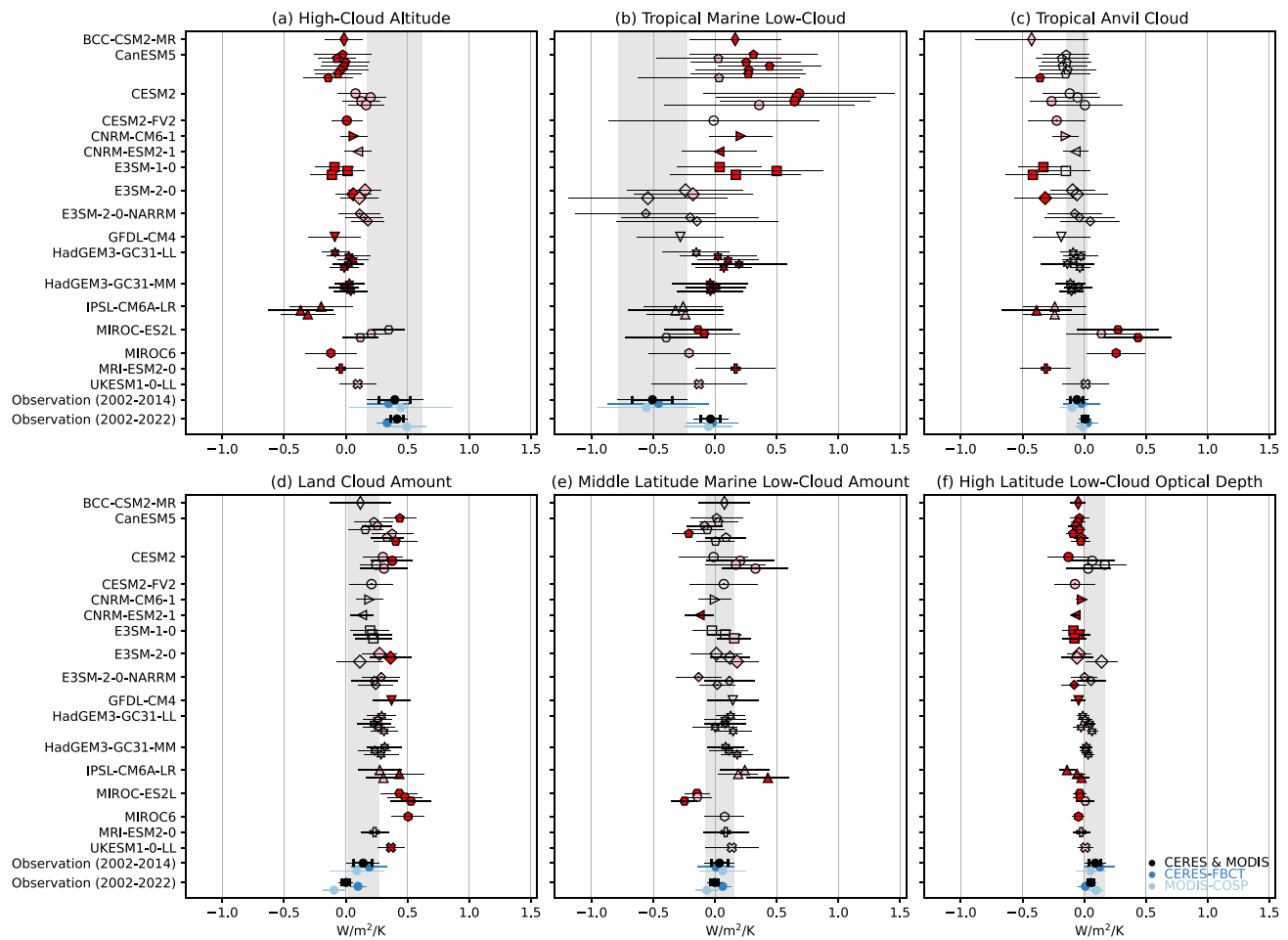


Figure 2. As in Figure 1, but for the assessed cloud feedback components. Note that the range of the x axis is smaller than in Figure 1.

The tropical marine low-cloud feedback is $-0.51 \pm 0.28 \text{ W/m}^2/\text{K}$ in this 12-year period, primarily due to a strong negative SW cloud amount component that arises from increases in low-cloud cover as temperature increases. Note that this feedback component is highly sensitive to the time period considered, which will be discussed in Section 3.7. Ten of seventeen models have at least one member matching the sign of the observed feedback, but the models are generally biased too positive for this feedback. Notably, at least one member from fourteen of the models has a tropical marine low-cloud feedback that is significantly high-biased at 66% confidence, with many being high-biased at 90% confidence. There are no members of any model that have a negative tropical marine low-cloud feedback value that is significantly larger in magnitude than what is observed. Marked disagreement is apparent among models and between models and the observed tropical marine low cloud feedback in regimes of weak descent (Figure S4a in Supporting Information S1).

The tropical anvil cloud feedback inferred from the observations is weakly negative ($-0.06 \pm 0.09 \text{ W/m}^2/\text{K}$), as the negative high cloud optical depth feedback is largely offset by the positive high cloud amount feedback in regions of tropical oceanic ascent. As temperature increases, the high-cloud fraction decreases over ascending regions, which results in less SW reflection (not shown). At the same time, the optical thickness of high clouds increases, reflecting more SW radiation, and these two effects nearly perfectly counteract each other (not shown). In both the amount and optical depth feedbacks, the LW component counteracts the SW component with opposite sign but smaller magnitude. The majority of AMIP models exhibit a weak to moderate negative feedback, with nine models having at least one member exhibiting negative anvil cloud feedbacks that are strong enough to be significantly low-biased with respect to observations at 66% confidence. However, the bias is less systematic than for other feedback components, as two models have at least one member with positive anvil feedbacks that are

high-biased at 66% confidence. These anvil cloud feedback biases (of either sign) are mainly driven by the high cloud amount with minor contribution from optical depth component (not shown).

The observed land cloud amount feedback is $+0.14 \pm 0.13 \text{ W/m}^2/\text{K}$, where low- and high-clouds contribute around 60% and 40%, respectively. The positive feedback is consistent with previous literature based on GCMs and observations, where relative humidity decreases over land as temperature increases and leads to a reduction in cloud cover (e.g., Kamae et al., 2016; Zhang & Klein, 2013). Every member of each model has a positive feedback, in agreement with observations. However, eleven models have at least one member exhibiting land cloud amount feedbacks that are significantly high-biased at 66% confidence, and no models have any members with feedbacks that are significantly low-biased with respect to observations. The regions contributing to positive feedback values include western Asia/eastern Europe, southern Australia, southeastern U.S., and northern South America. The CMIP6 multi-model mean agrees with the sign of the observations over subtropical regions but fails to reproduce the strong positive values over Eurasia and the southeastern U.S., while also overestimating the positive feedback over Australia and Africa (Figures S3b and S3f in Supporting Information S1).

The majority of models have mid-latitude (30–60°N/S) marine low-cloud amount feedbacks that agree reasonably well with the small value estimated in observations ($+0.04 \pm 0.12 \text{ W/m}^2/\text{K}$), with all but three models having at least one member that lies within the statistical uncertainty of the observed value. While nine models have at least one member that is significantly biased with respect to observations at 66% confidence, these biases are not systematic across models, with some biased low and others biased high. The general agreement between observations and models is also apparent in the spatial pattern, where the multi-model mean captures the positive values over northeastern and southwestern Pacific, as well as the negative values over northwestern and southeastern Pacific (Figures S3c and S3g in Supporting Information S1).

In contrast to the neutral high-latitude (40–70°N/S) low-cloud optical depth feedback assessed in S20 on long-term time scales, the observations suggest a positive short-term feedback of $+0.09 \pm 0.08 \text{ W/m}^2/\text{K}$, in close agreement with a previous observational analysis (i.e., $+0.10 \pm 0.03 \text{ W/m}^2/\text{K}$ (1.645 standard deviation); Wall et al. (2022)). In observations, low clouds transition from optically thick to thin as global mean temperature increases, leading to a reduction of SW reflection and a positive optical depth feedback. Changes in low cloud optical depth arise from a complex mix of temperature-dependent processes including changes in the moist adiabatic lapse rate, inversion strength, boundary layer decoupling, phase partitioning, and changes in adiabaticity (e.g., Terai et al., 2019), which are not further explored here. The CMIP6 models are systematically low-biased for this feedback, with every model having at least one member (and in many cases, all members) exhibiting a low-biased feedback that is significant at 66% confidence. No members of any model have a positive high latitude low cloud optical depth feedback that is significantly larger than estimated from observations. This is consistent with Terai et al. (2016), who found that low cloud optical depth increases too strongly as temperature increases in most CMIP5 AMIP models.

In summary, despite substantial uncertainty arising from weather noise, several systematic biases in modeled assessed cloud feedback components are apparent: The positive high cloud altitude and high latitude low cloud optical depth feedbacks are systematically underestimated in models, with many model realizations actually having the wrong sign and no realizations having values that are significantly larger than observed. Several model realizations also exhibit negative anvil cloud feedbacks that are too large in magnitude relative to the small negative anvil cloud feedback in observations. These low biases are countered somewhat by high biases in other components. Notably, most model realizations underestimate the strength of the negative tropical marine low cloud feedback, with many simulating a positive feedback instead, and some model realizations significantly overestimate the strength of the positive land cloud amount feedback. Hence most of the individual assessed components are systematically biased in models, but these biases closely compensate, resulting in an overall lack of systematic errors in the sum of assessed feedbacks and the total cloud feedback (Figures 1a and 1b).

3.3. Comparisons of Unassessed Cloud Feedback Components

S20 quantified a set of six cloud feedback components that are well-studied and/or important, and model evaluation in Z22 was done only for those components. An advantage of our approach is that we are not restricted to S20's feedbacks and can evaluate feedback components that were unassessed in S20. As mentioned above, the observations show a positive unassessed cloud feedback on the interannual timescale (significant at 90% confidence). However, the majority of CMIP6 AMIP simulations underestimate its magnitude (Figure 1c).

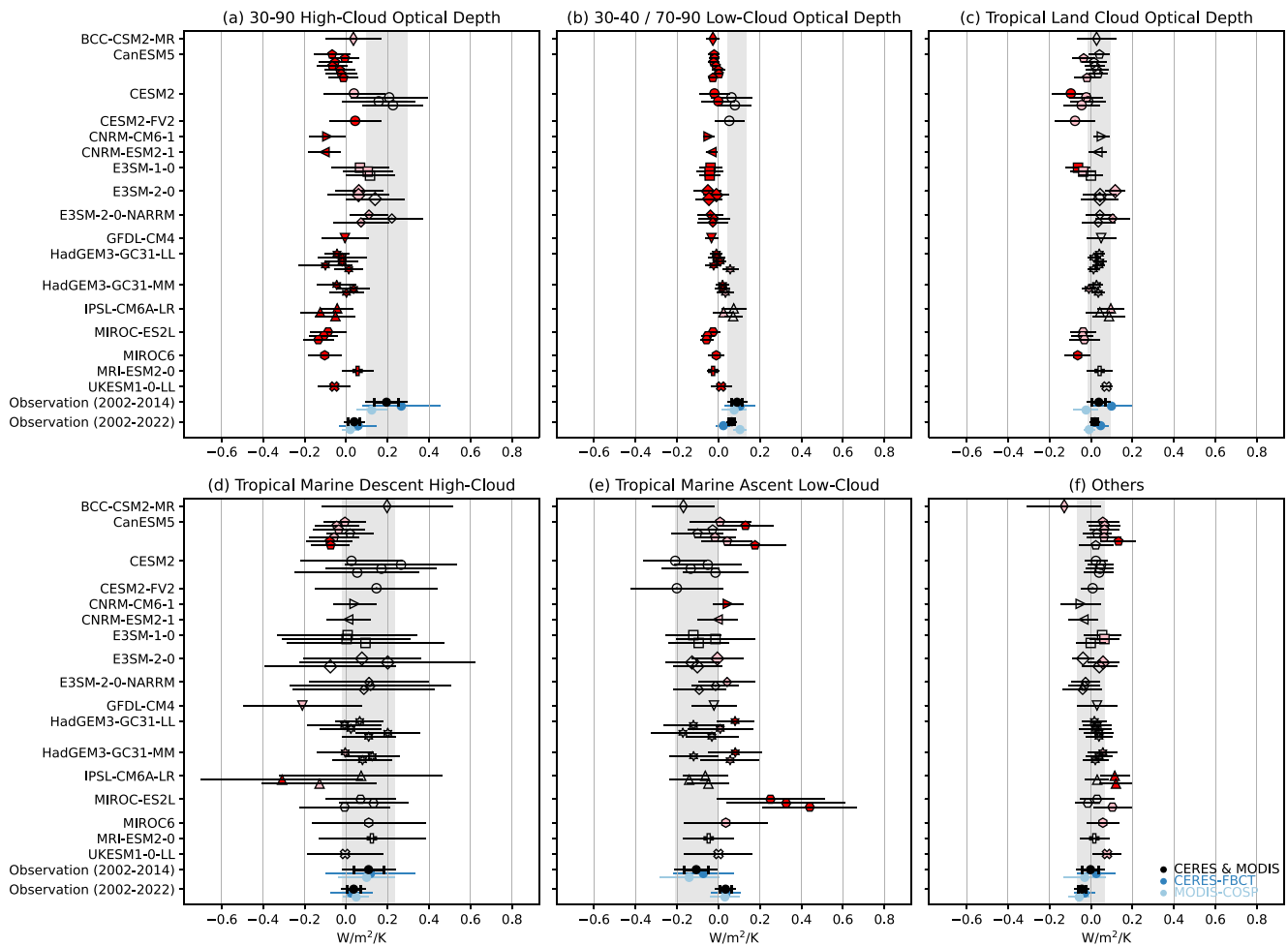


Figure 3. As in Figure 1, but for the unassessed cloud feedback components. Note that the range of the x axis is smaller than in Figure 1.

We decompose the unassessed cloud feedback into 6 components plus a residual term (Table 2 and Figure 3). Note that the unassessed components could be grouped in many ways. Here we start from the components used in Z22 and further aggregate those that share similar characteristics for simplicity. The observed unassessed cloud feedback receives the largest contribution from the extratropical (30–90°N/S) high-cloud optical depth feedback (i.e., $+0.20 \pm 0.10$ W/m²/K, Figure 3a). This component is also one of the main contributors to model-observation discrepancy. Every model has at least one realization that is biased low at 66% significance, with many of the realizations being significantly biased at 90% confidence. Not a single realization has an extratropical high cloud optical depth feedback that is significantly larger than the observed value. As global temperature increases, the high clouds shift from optically thick to optically thin, thereby reflecting less SW radiation. As shown in Figures S5a and S5d in Supporting Information S1, this comes mainly from clouds over land over the northern hemisphere (around 52%) and southern Pacific (around 30%). Underestimation over these two regions contributes to a weaker positive or even weak negative value in CMIP6 AMIP simulations.

The extratropical low-cloud optical depth component is weakly positive in the observations (i.e., $+0.09 \pm 0.05$ W/m²/K, Figure 3b). Here extratropical is defined as poleward of 30° latitude but excluding the 40–70° band that is already counted in the assessed high-latitude low-cloud optical depth feedback. The mechanism driving positive values is similar to the high-latitude low-cloud optical depth, with 30–40° latitude having decreased cloud optical depth as temperature changes and 70–90° latitude having little cloud fraction changes among optical depth bins. The models systematically underestimate the feedback, with all but one model having at least one realization that is significantly low-biased at 66% confidence and many low-biased at 90% confidence. Not a single realization has a feedback that exceeds the observed value. The underestimation occurs at nearly every location within the domain (Figures S5b and S5e in Supporting Information S1).

Dominated by high clouds, the tropical land cloud optical depth feedback is $+0.04 \pm 0.06$ W/m²/K in the observations (Figure 3c). Most CMIP6 models (13 of 17 models) have at least one member that agrees within observed 90% confidence intervals. Ten models have at least one member that significantly differs from the observations, but the biases are not systematic as seven models exhibit low-biased feedbacks while three models exhibit high-biased feedbacks. The observed positive feedback mainly originates from East Asia and the northern part of South America, but these spatial features are not well-captured in CMIP6 multi-model mean (Figures S5c and S5f in Supporting Information S1).

The observations show a positive tropical marine descent high-cloud feedback ($+0.11 \pm 0.13$ W/m²/K; Figure 3d). This is mainly due to a positive optical depth component from high clouds becoming optically thinner with warming. In contrast, the net high cloud amount component is small despite increases in high cloud cover with warming owing to opposing LW and SW effects (not shown). All models and members exhibit a feedback value that agrees with the observations except for four models that have at least one member that is low-biased at 66% confidence.

The tropical marine ascent low-cloud feedback is -0.11 ± 0.10 W/m²/K in the observations (Figure 3e), due to a negative cloud optical depth component partially offset by a positive cloud amount component. The low clouds shift from optically thin ($\tau = 0.3$ – 3.6) to moderate optical thickness ($\tau = 3.6$ – 23) with little change in the optically thick regime ($\tau > 23$) (not shown). Meanwhile, the low-cloud fraction decreases as temperature increases, reflecting less SW radiation. Around half of the models (9 of 17 models) have at least one member that significantly overestimates the feedback at 66% confidence, while no model simulations are significantly low-biased.

The term “Others” includes extratropical (30–90°N/S) marine high-cloud amount, 60–90°N/S marine low-cloud amount, and low-cloud altitude feedbacks, as these three components have smaller magnitudes in both observations and CMIP6 AMIP simulations. Together they make a minor contribution to the unassessed cloud feedback and the model values are in reasonable agreement with the observations (Table 2 and Figure 3f).

The unassessed cloud feedback includes three residual terms: the cloud radiative kernel decomposition residual (Zelinka et al., 2013, 2016), a covariance term from low-cloud obscuration calculation (Myers et al., 2021; Scott et al., 2020), and a residual term from performing tropical marine cloud feedbacks in ω_{500} space (Bony et al., 2004) (Tables 1 and 2). The first two terms have a small magnitude in both observations and models, but the last one, due to nonlinearity when aggregating geospatial information to ω_{500} space, has a relatively large magnitude and inter-model spread (not shown). We have also tested an alternative method in which feedbacks are computed in geographic space and then aggregated over ascent and descent regions based on the climatological 500-hPa vertical velocity at each month and each grid point. The residual term vanishes in the latter method, but our default method better ensures that we capture feedbacks due to cloud property changes within specific vertical motion regimes, which is desirable for connecting the feedbacks to their underlying physical processes.

In summary, the sum of unassessed cloud feedbacks is significantly positive in observations during the 2002–2014 period, and the optical depth components over four cloud regimes are the major contributors: extratropical (30–90°N/S) high-cloud, extratropical low-cloud, tropical land, and tropical high-cloud over oceanic descent regions (Figures 3a–3d). The CMIP6 ensemble tends to underestimate the magnitudes of the first three optical depth feedbacks, specifically for extratropical high- and low-cloud optical depth feedbacks. This highlights a possible bias in models' representation of high and low cloud optical depth and its response to warming.

3.4. Sources of Model-Observation Discrepancies

To evaluate the skill of individual models in reproducing observed cloud feedbacks, we compute the mean squared error (MSE) over all cloud feedback components for each model (Figure 4). The use of MSE allows us to quantify the relative contribution from each cloud type to the total MSE since the sum of the squared errors from individual cloud feedback components (divided by the number of components) equals the total MSE for a model. As shown in Figure 4, the model-observation discrepancies in most models come from three main components: the tropical marine low-cloud (light blue), high-cloud altitude (dark blue), and extratropical high-cloud optical depth feedbacks (dark red), with minor contributions from extratropical low-cloud optical depth (dark green and light green) and tropical anvil cloud feedback (dark orange). We highlight the components that are significantly biased high and low (90% confidence) with stippling and hatching. The three main components mentioned above contribute over 50% of MSE for all models except MIROC-ES2L, in which a wider diversity of components contribute.

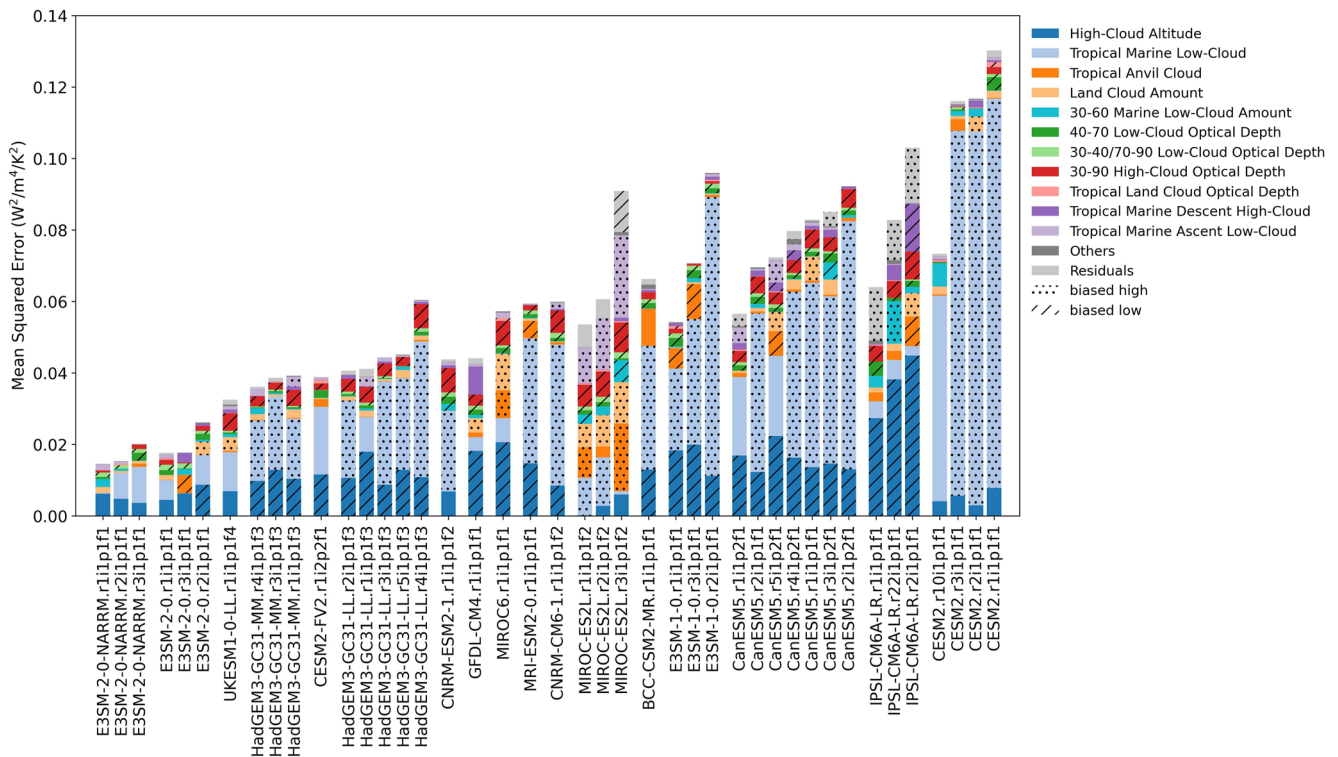


Figure 4. The contribution from each cloud feedback component to mean squared error (MSE). Components that are significantly biased high and low at 90% confidence are identified with stippling and hatching, respectively.

Our results suggest that caution is needed when comparing a single climate model simulation with a single observation in either 12-year or 20-year analysis (i.e., the current longest satellite observation record), since both modeled and observed results could be affected by internal variability. Despite quantitative differences, individual realizations of a particular climate model tend to resemble each other in terms of the error properties and are typically distinct from other models (Figure 4). For instance, CESM2's bias overwhelmingly comes from an overestimated tropical marine low cloud with little role for a high cloud altitude bias. In contrast, high cloud altitude feedback is significantly low biased in all members of IPSL-CM6A-LR while their tropical marine low cloud feedback is close to observations. Additionally, in contrast to most models, a diverse mix of components are biased other than tropical marine low cloud or high cloud altitude in MIROC-ES2L. One can also discern systematic improvements across generations of models and between model configurations. For example, E3SM-2-0 and E3SM-2-0-NARRM show marked improvement in all categories over their predecessor, E3SM-1-0. And the large biases in the tropical marine low-cloud feedback in CESM2 are much weaker and insignificant in CESM2-FV2.

3.5. Sensitivity of Model-Observation Comparison to Choice of Cloud Radiative Kernel

In this section, we test the sensitivity of our results to the choice of cloud radiative kernel. Figure S6 in Supporting Information S1 shows the comparison of observed cloud feedback using the cloud radiative kernels of Zhou et al. (2013, 2022), Zhang et al. (2021), and CERES FBCT (Sun et al., 2022). Zhou et al. (2013, 2022) kernels are derived using reanalysis thermodynamic fields supplied to the Fu-Liou and RRTM radiative transfer models, respectively. The Zhang et al. (2021) kernel is calculated from the GISS model radiation code that is used in constructing the International Satellite Cloud Climatology Project (ISCCP) H data sets. CERES FBCT is derived empirically using clear-sky and overcast fluxes from CERES observations coincident with each MODIS cloud type. Cloud feedbacks inferred from the first three products are consistent in general, while the CERES FBCT-estimated cloud feedbacks have smaller values. This is most apparent for the LW high cloud altitude and amount components, which is consistent with the fact that the FBCT LW kernel shows a weaker increase in strength with altitude than the radiative transfer-based kernels (Figure S7 in Supporting Information S1). The SW

components, in contrast, are in better agreement among the kernels (not shown). Further investigation is needed to determine whether this discrepancy implies that radiative transfer model-based kernels give a biased estimate of the radiative impact of certain cloud types (perhaps due to assumptions made in constructing them) or that the empirical nature of the CERES FBCT introduces errors (e.g., sampling at most two times per day during the sunlit MODIS overpasses, with many cloud types rarely observed from space). Repeating all analyses using the FBCT kernel instead results in quantitative differences in feedback magnitudes and MSE values, which results in a smaller positive total cloud feedback that is closer to the cloud feedback derived from cloud radiative effect (i.e., the difference between all-sky and clear-sky radiative fluxes) adjusted for non-cloud effects (Raghuraman et al., 2023). Yet, the choice of a different kernel does not change any of the conclusions. We find that the models overestimate the tropical marine low-cloud feedback and underestimate high-cloud altitude and extratropical high- and low-cloud optical depth feedbacks (Figures S8 to S11 in Supporting Information S1).

3.6. Sensitivity of Model-Observation Comparison to Choice of Cloud Fraction Product

In addition to being sensitive to choice of cloud radiative kernels, cloud feedback values estimated from observations depend on which cloud fraction product is used (Figures 1–3, Table S4 in Supporting Information S1). Compared to the moderate positive cloud feedback derived using CERES-FBCT, the MODIS-COSP-derived feedback is near-zero. This is because all but two feedback components are less positive or more negative in this data set (Table S4 in Supporting Information S1). The two exceptions are the positive high-cloud altitude and middle latitude marine low-cloud amount feedbacks, which are stronger in MODIS-COSP. None of the differences in the individual cloud feedback components or the total cloud feedback between the two observational products are statistically significant at 90% confidence (comparing cloud feedbacks from the two data sets using the Monte Carlo method described in Section 2.4, not shown). We find that the main conclusions of the study remain robust, independent of whether the models are evaluated against CERES-FBCT or MODIS-COSP observations (Table S4 in Supporting Information S1). Since both CERES-FBCT and MODIS-COSP report cloud properties measured by the MODIS instrument, the differences in feedbacks must be related in some way to differences in their cloud retrieval algorithms. Understanding the nature and causes of these differences is important for future work but is beyond the scope of this study.

3.7. Observed Cloud Feedback Inferred From Different Periods

This study also analyzes the full period of available observations, from July 2002 to December 2022. The best estimates of the total cloud feedback inferred from 12-yr and 20-yr observations are both positive (0.33 and 0.57 W/m²/K), implying that the radiative response of clouds amplifies interannual temperature fluctuations. The regression uncertainty decreases when analyzing a longer period, and the likelihood of a negative cloud feedback is reduced from 16% to less than 0.1%.

This analysis indicates that several observed cloud feedback components vary in strength and even sign across time periods. With more recent years included, the tropical marine low-cloud feedback switches from a strong negative to a weak negative value (Figure 2b), the positive extratropical high-cloud optical depth feedback becomes smaller (Figure 3a), and the tropical marine ascent low-cloud feedback switches from moderately negative to weakly positive (Figure 3e). Changes in these feedback components between the two periods are significant at 90% confidence (following the methodology described in Section 2.4 but comparing 12-yr and 20-yr observations). While it is difficult to fully rule out random meteorological noise, we can ask whether feedback differences between pairs of realizations within a single model subject to the same SST boundary conditions within the 2002–2014 period are ever statistically significant at 90% confidence. If they are, then one cannot rule out noise as the reason for statistically significant differences in feedbacks. This is the case for the tropical marine ascent low-cloud feedback. In contrast, not a single pair of realizations of any model exhibit significant differences at 90% confidence in tropical marine low-cloud and extratropical high-cloud optical depth feedbacks (not shown). From this we conclude that the observed differences in these two components exceed what can be reasonably expected to arise from noise, and are instead reflecting a true dependence on time period.

The spatial pattern of temperature response per degree of interannual global warming is different between the July 2002 to December 2014 and July 2002 to December 2022 periods (Figure S12 in Supporting Information S1), suggesting that the dependence of these cloud feedbacks on the choice of periods is related to the pattern effect

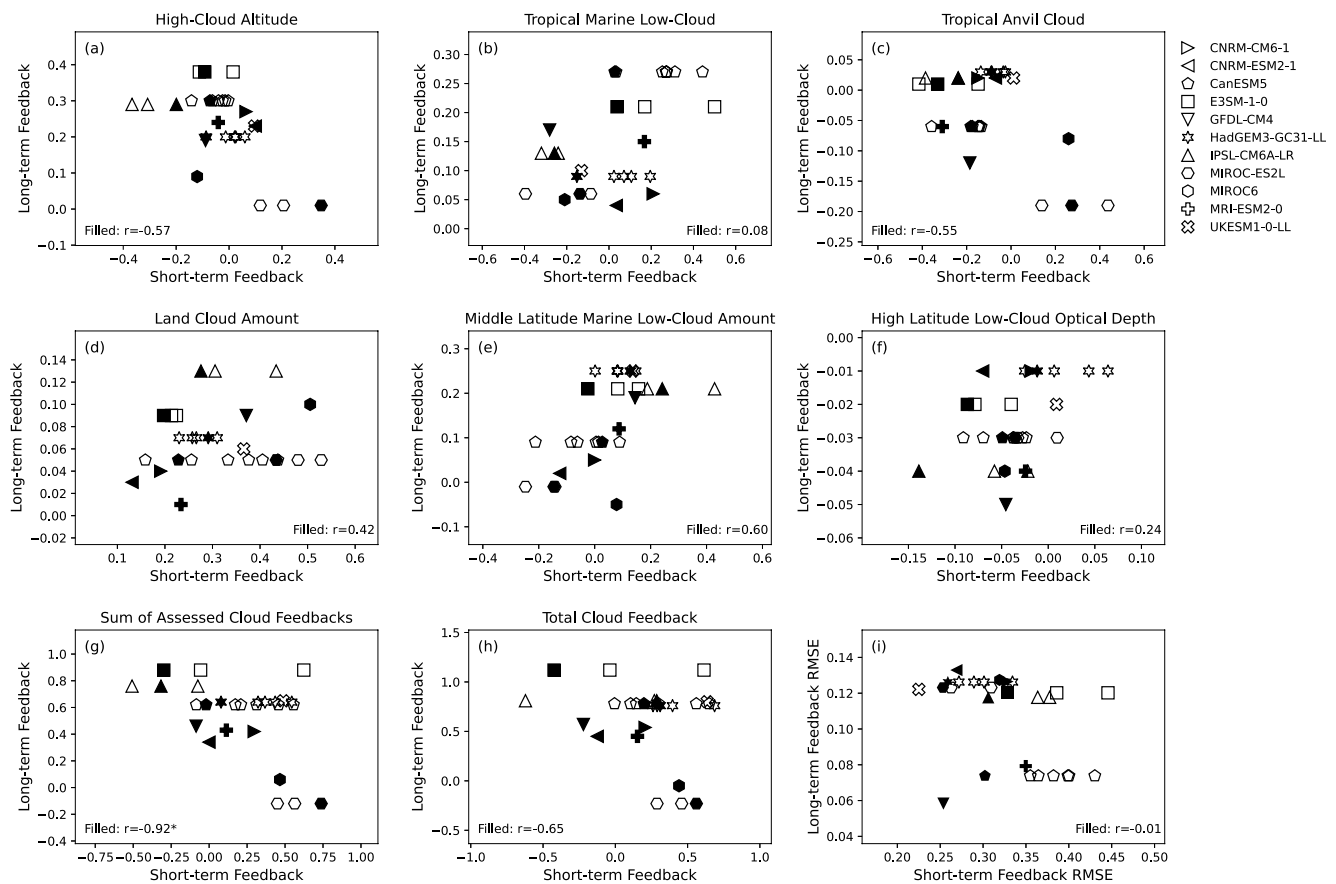


Figure 5. The scatterplot of (a–f) individual assessed, (g) sum of assessed, and (h) total cloud feedbacks (in $\text{W/m}^2/\text{K}$) inferred from the short and long timescales for each model common to this analysis and that of Z22. (i) Root mean squared error (RMSE; in $\text{W/m}^2/\text{K}$) of cloud feedbacks on the long timescale scattered against those on the short timescale. See text for the description of RMSE on two timescales. For each model, the realization used in both Z22 and this analysis is shown as the filled marker, while other realizations are shown as open markers. Correlation coefficients calculated from filled markers are shown, with significant correlations at 95% confidence intervals (p -value < 0.025) marked by asterisks.

(Stevens et al., 2016), which describes how the radiative response depends not only on global average surface warming but also on the spatial pattern of warming. Specifically, studies have shown that the pattern effect is mainly due to the low-cloud feedback, which is strongly modulated by variations in inversion strength that are themselves tied to the large-scale gradients in surface temperature (e.g., Andrews & Webb, 2018; Ceppi & Gregory, 2017; Zhou et al., 2016). Recent studies have demonstrated the impacts of the pattern effect on total climate feedback and cloud feedback during the satellite era (e.g., Chao et al., 2022; Loeb et al., 2020). In this study, we confirm the substantial impacts of the pattern effect on low cloud feedback, but also highlight other cloud feedbacks that exhibit substantial dependence on time period considered. Future work should clarify the mechanism underlying the temporal variations in the strength of the extratropical high-cloud optical depth feedback, and how it may be related to variations in surface warming pattern.

3.8. Relations Between Short-Term and Long-Term Cloud Feedbacks and Their Errors

How do cloud feedbacks estimated on the interannual timescale relate to those in response to long-term warming? Plotting the six assessed cloud feedback values computed on interannual timescales against those in response to $4\times\text{CO}_2$ (taken from Z22) for the 11 models that have both sets of simulations indicates that there is no significant relation (at 95% confidence) between the two timescales for any individual assessed component nor total cloud feedback, except for a significant negative correlation for the sum of assessed feedbacks (Figure 5a–5h). This result suggests that cloud feedback estimates derived from interannual global temperature fluctuations may be of limited value for long-term climate projections or for constraining ECS. A more fruitful approach to leverage observations may be to perform cloud controlling factor analyses in which the sensitivity of cloud properties to

environmental controlling factors is determined from observed fluctuations and then scaled by model-projected changes in these factors (e.g., Ceppi & Nowack, 2021; Cesana & Del Genio, 2021; Gordon & Klein, 2014; Klein et al., 2017; Myers et al., 2021; Myers & Norris, 2016; Qu et al., 2015; Terai et al., 2016).

Do models that perform better in simulating short-term cloud feedback also perform better in simulating the cloud feedback in response to long-term warming? We estimate the RMSE (model minus observations) across the six assessed short-term cloud feedbacks and scatter them against their 4xCO₂ counterparts (model minus expert judgment) taken from Z22. There is no significant correlation between the skill in simulating climate feedback on short-term and long-term timescales (Figure 5i). Models with better skill in matching long-term cloud feedback components have diverse performances in simulating short-term cloud feedback, and vice versa. The models that fall in the lower left quadrant represent smaller-than-average errors at both interannual and long-term timescales (e.g., GFDL-CM4), and the models falling in the upper right quadrant are the ones with larger-than-average errors at both timescales, such as E3SM-1-0 and IPSL-CM6A-LR. Two of the models that perform best with respect to S20 on the long timescale are among the worst performers here on the short timescale (CanESM5 and MRI-ESM2-0).

Note that the scatter plot should be interpreted with caution as several uncertainties can influence the results. For instance, here we only assess the skill of simulating cloud feedback based on six assessed cloud types. Including unassessed cloud feedbacks does not affect the short-term feedback RMSE much (not shown), but it is unknown whether the long-term RMSE could be altered by including more components. Given the fact that several currently unassessed cloud feedbacks are not negligible in most models' abrupt-4xCO₂ simulations (Z22), it is likely that long-term RMSE would change if one or more of these components is quantified in a future assessment. Second, the calculation of long-term feedback RMSE relies on the expert assessment in S20; should another assessment exercise be conducted in the future, the expert assessment of the cloud feedback components may change. Lastly, the interannual cloud feedback is estimated during a relatively short period (i.e., 12.5 years). Extending the AMIP experiments beyond 2014 to more closely match the longer record of CERES observations (e.g., Raghuraman et al., 2021; Raghuraman et al., 2023; Schmidt et al., 2023) will allow for a more robust comparison between modeled and observed short-term cloud feedbacks.

4. Discussion and Conclusions

This study quantifies individual short-term cloud feedback components using satellite observations and CMIP6 AMIP simulations. Building upon Zelinka et al. (2022)'s work in evaluating modeled long-term cloud feedback, this work provides another test of the CMIP models by comparing them to observed cloud feedbacks in response to interannual temperature fluctuations. We present a novel analysis of cloud feedback components observed in recent decades and provide an assessment of how well the models perform in matching individual cloud feedback components.

In observations, we find a positive total cloud feedback that is mainly due to positive high-cloud altitude component, land cloud amount component, and extratropical high- and low-cloud optical depth components, partially offset by negative tropical marine low-cloud amount component. The extratropical high- and low-cloud optical depth feedbacks, which were not assessed in S20, are both positive and not negligible on the interannual timescale. Our results also show evidence of the pattern effect on tropical marine low-cloud feedback and extratropical high-cloud optical depth feedbacks, as their values differ significantly between the 2002–2014 and the 2002–2022 periods that are characterized by different patterns of warming.

CMIP6 AMIP models simulate a total cloud feedback that agrees with the observations in general. However, the models systematically overestimate tropical marine low-cloud and land cloud amount feedback and underestimate high-cloud altitude and extratropical high- and low-cloud optical depth feedbacks. The model-observation discrepancies are mainly from the tropical marine low-cloud, high-cloud altitude, and extratropical high-cloud optical depth feedbacks.

The biases in AMIP models present a number of important issues that warrant future investigation. Many previous studies focused on the uncertainty in tropical marine low-cloud feedback over subsidence regions. This study finds that the substantial uncertainty in this feedback persists for the short-term cloud feedback in AMIP simulations, primarily due to an inability to accurately simulate the response of the marine low-cloud fraction to temperature changes. Additionally, our results highlight the need to investigate high-cloud feedback and the

mechanisms driving high clouds, in agreement with the key findings in Z22. Despite a well-developed theoretical background, AMIP models tend to underestimate the increase of high-cloud altitude as global mean temperature increases. Moreover, the unassessed cloud feedback is biased low in most AMIP models mainly due to a systematically weaker extratropical high-cloud optical depth feedback. More work is needed to better understand the processes governing high-cloud optical depth in nature (including those that determine cloud phase, total water content, particle sizes, and other cloud microphysical properties) and its changes in response to interannual temperature variations. This echoes recent studies arguing for the importance of examining high-cloud optical depth feedbacks (Gasparini et al., 2023; Li et al., 2019; McKim et al., 2024; Zelinka et al., 2022). The consistent underestimation in high cloud feedbacks motivates further work to analyze how high clouds respond to their cloud controlling factors.

This study also demonstrates that the magnitude of short-term cloud feedback components is dependent on several factors that must be carefully accounted for when evaluating models against observations. First, statistical noise arising from meteorological fluctuations that are not systematically dependent on global mean temperature is substantial, limiting the robustness of any quantitative estimate derived from a single short observational data set and the rigor of any model evaluation that relies on a small number of realizations. Second, the magnitudes of short-term cloud feedbacks vary depending on the choice of cloud radiative kernel and observed cloud fraction data set. While the main conclusions of this paper are largely insensitive to these choices, caution is needed when quantifying the short-term cloud feedback from a single radiative kernel or observational cloud data set. Finally, cloud feedbacks estimated from coincident interannual cloud and temperature fluctuations are not correlated across models with their long-term counterparts in response to greenhouse warming. It is therefore inappropriate to view feedbacks quantified on observable timescales as being interchangeable with the long-term feedbacks that largely determine equilibrium climate sensitivity. Targeted analyses using cloud controlling factors may be necessary to more rigorously establish the link between observable cloud sensitivities on short timescales and the uncertain cloud feedback in response to long-term warming.

One of the ultimate goals in climate science is narrowing the range of equilibrium climate sensitivity in GCMs. While we do not find that model skill in simulating short-term feedback translates into skill in simulating long-term feedbacks, the evaluation of climate models on interannual timescales against observations provides an apples-to-apples comparison that avoids evaluation against partly-subjective expert judgment values and helps reveal compensating errors in modeled cloud feedback components, which is a crucial step in further improving the representation of clouds in climate models.

Data Availability Statement

All data used in this study are publically available at <https://ceres.larc.nasa.gov/data/#fluxbycltyp-level-3> (CERES FBCT), https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MCD06COSP_M3_MODIS (MODIS COSP), and <https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5> (ERA5). CMIP6 models used in this study are listed in Table S1 in Supporting Information S1 and are available at <https://esgf-node.llnl.gov/search/cmip6/>. Cloud radiative kernels are available at Zelinka (2021) (<https://doi.org/10.5281/zenodo.5514137>). The code is available at <https://doi.org/10.5281/zenodo.10481816>.

References

- Andrews, T., & Webb, M. J. (2018). The dependence of global cloud and lapse rate feedbacks on the spatial structure of tropical Pacific warming. *Journal of Climate*, 31(2), 641–654. <https://doi.org/10.1175/JCLI-D-17-0087.1>
- Bony, S., Dufresne, J. L., Le Treut, H., Morcrette, J. J., & Senior, C. (2004). On dynamic and thermodynamic components of cloud changes. *Climate Dynamics*, 22(2), 71–86. <https://doi.org/10.1007/s00382-003-0369-6>
- Ceppi, P., & Gregory, J. M. (2017). Relationship of tropospheric stability to climate sensitivity and Earth's observed radiation budget. *Proceedings of the National Academy of Sciences*, 114(50), 13126–13131. <https://doi.org/10.1073/pnas.1714308114>
- Ceppi, P., & Nowack, P. (2021). Observational evidence that cloud feedback amplifies global warming. *Proceedings of the National Academy of Sciences of the United States of America*, 118(30), e2026290118. <https://doi.org/10.1073/pnas.2026290118>
- CERES Science Team. (2022). CERES_FluxByCldTyp-Day/Month_Ed4A data quality summary. Retrieved from <https://ceres.larc.nasa.gov/data/>
- Cesana, G. V., & Del Genio, A. D. (2021). Observational constraint on cloud feedbacks suggests moderate climate sensitivity. *Nature Climate Change*, 11(3), 213–218. <https://doi.org/10.1038/s41558-020-00970-y>
- Chao, L.-W., & Dessler, A. E. (2021). An assessment of climate feedbacks in observations and climate models using different energy balance frameworks. *Journal of Climate*, 34(24), 9763–9773. <https://doi.org/10.1175/jcli-d-21-0226.1>

Acknowledgments

This work was supported by National Science Foundation Grant AGS-2243602 and NASA FINESST Grant 80NSSC20K1606, both to Texas A&M University. The work of MDZ and LWC was additionally supported by the U.S. Department of Energy (DOE) Regional and Global Model Analysis program area and was performed under the auspices of the DOE by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. We thank Shiv Priyam Raghuraman and two anonymous reviewers, whose constructive comments greatly improved this paper.

- Chao, L.-W., Muller, J. C., & Dessler, A. E. (2022). Impacts of the unforced pattern effect on the cloud feedback in CERES observations and climate models. *Geophysical Research Letters*, 49(2), e2021GL096299. <https://doi.org/10.1029/2021GL096299>
- Dessler, A. E., & Forster, P. (2018). An estimate of equilibrium climate sensitivity from interannual variability. *Journal of Geophysical Research: Atmospheres*, 123(16), 8634–8645. <https://doi.org/10.1029/2018JD028481>
- Donohoe, A., Armour, K. C., Pendergrass, A. G., & Battisti, D. S. (2014). Shortwave and longwave radiative contributions to global warming under increasing CO₂. *Proceedings of the National Academy of Sciences of the United States of America*, 111(47), 16700–16705. <https://doi.org/10.1073/pnas.1412190111>
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model inter-comparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- Gasparini, B., Sullivan, S. C., Sokol, A. B., Kärcher, B., Jensen, E., & Hartmann, D. L. (2023). Opinion: Tropical cirrus – From micro-scale processes to climate-scale impacts. *Atmospheric Chemistry and Physics*, 23(24), 15413–15444. <https://doi.org/10.5194/acp-23-15413-2023>
- Gordon, N. D., & Klein, S. A. (2014). Low-cloud optical depth feedback in climate models. *Journal of Geophysical Research: Atmospheres*, 119(10), 6052–6065. <https://doi.org/10.1002/2013JD021052>
- Hartmann, D. L., & Larson, K. (2002). An important constraint on tropical cloud-climate feedback. *Geophysical Research Letters*, 29(20), 12-11–12-14. <https://doi.org/10.1029/2002GL015835>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- IPCC. (2021). Climate change 2021: The physical science basis. In *Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge University Press.
- Kamae, Y., Ogura, T., Watanabe, M., Xie, S.-P., & Ueda, H. (2016). Robust cloud feedback over tropical land in a warming climate. *Journal of Geophysical Research: Atmospheres*, 121(6), 2593–2609. <https://doi.org/10.1002/2015JD024525>
- Klein, S. A., Hall, A., Norris, J. R., & Pincus, R. (2017). Low-cloud feedbacks from cloud-controlling factors: A review. *Shallow Clouds, Water Vapor, Circulation, and Climate Sensitivity*, 38(6), 135–157. <https://doi.org/10.1007/s10712-017-9433-3>
- Klein, S. A., & Jakob, C. (1999). Validation and sensitivities of frontal clouds simulated by the ECMWF model. *Monthly Weather Review*, 127(10), 2514–2531. [https://doi.org/10.1175/1520-0493\(1999\)127<3C2514:VASOFC%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<3C2514:VASOFC%3E2.0.CO;2)
- Li, R. L., Storelvmo, T., Fedorov, A. V., & Choi, Y.-S. (2019). A positive iris feedback: Insights from climate simulations with temperature-sensitive cloud–rain conversion. *Journal of Climate*, 32(16), 5305–5324. <https://doi.org/10.1175/JCLI-D-18-0845.1>
- Loeb, N. G., Su, W., & Kato, S. (2016). Understanding climate feedbacks and sensitivity using observations of Earth's energy budget. *Current Climate Change Reports*, 2(4), 170–178. <https://doi.org/10.1007/s40641-016-0047-5>
- Loeb, N. G., Wang, H., Allan, R. P., Andrews, T., Armour, K., Cole, J. N. S., et al. (2020). New generation of climate models track recent unprecedented changes in Earth's radiation budget observed by CERES. *Geophysical Research Letters*, 47(5), e2019GL086705. <https://doi.org/10.1029/2019GL086705>
- McKim, B., Bony, S., & Dufresne, J.-L. (2024). Physical and observational constraints on the anvil cloud area feedback. *Nature Geoscience*. <https://doi.org/10.22541/au.167769953.39966398/v2>
- Minnis, P., Sun-Mack, S., Young, D. F., Heck, P. W., Garber, D. P., Chen, Y., et al. (2011). CERES edition-2 cloud property retrievals using TRMM VIRS and terra and aqua MODIS data—Part I: Algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11), 4374–4400. <https://doi.org/10.1109/TGRS.2011.2144601>
- Myers, T. A., & Norris, J. R. (2016). Reducing the uncertainty in subtropical cloud feedback. *Geophysical Research Letters*, 43(5), 2144–2148. <https://doi.org/10.1002/2015GL067416>
- Myers, T. A., Scott, R. C., Zelinka, M. D., Klein, S. A., Norris, J. R., & Caldwell, P. M. (2021). Observational constraints on low cloud feedback reduce uncertainty of climate sensitivity. *Nature Climate Change*, 11(6), 501–507. <https://doi.org/10.1038/s41558-021-01039-0>
- Pincus, R., Hubanks, P. A., Platnick, S., Meyer, K., Holz, R. E., Botambekov, D., & Wall, C. J. (2023). Updated observations of clouds by MODIS for global model assessment. *Earth System Science Data*, 15(6), 2483–2497. <https://doi.org/10.5194/essd-15-2483-2023>
- Pincus, R., Platnick, S., Ackerman, S. A., Hemler, R. S., & Patrick Hofmann, R. J. (2012). Reconciling simulated and observed views of clouds: MODIS, ISCCP, and the limits of instrument simulators. *Journal of Climate*, 25(13), 4699–4720. <https://doi.org/10.1175/JCLI-D-11-00267.1>
- Platnick, S., Meyer, K. G., King, M. D., Wind, G., Amarasinghe, N., Marchant, B., et al. (2017). The MODIS cloud optical and microphysical products: Collection 6 updates and examples from terra and aqua. *IEEE Transactions on Geoscience and Remote Sensing*, 55(1), 502–525. <https://doi.org/10.1109/TGRS.2016.2610522>
- Qu, X., Hall, A., Klein, S. A., & Caldwell, P. M. (2015). The strength of the tropical inversion and its response to climate change in 18 CMIP5 models. *Climate Dynamics*, 45(1–2), 375–396. <https://doi.org/10.1007/s00382-014-2441-9>
- Raghuraman, S. P., Paynter, D., Menzel, R., & Ramaswamy, V. (2023). Forcing, cloud feedbacks, cloud masking, and internal variability in the cloud radiative effect satellite record. *Journal of Climate*, 36(12), 4151–4167. <https://doi.org/10.1175/JCLI-D-22-0555.1>
- Raghuraman, S. P., Paynter, D., & Ramaswamy, V. (2021). Anthropogenic forcing and response yield observed positive trend in Earth's energy imbalance. *Nature Communications*, 12(1), 4577. <https://doi.org/10.1038/s41467-021-24544-4>
- Santer, B. D., Wigley, T. M. L., Boyle, J. S., Gaffen, D. J., Hnilo, J. J., Nychka, D., et al. (2000). Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *Journal of Geophysical Research*, 105(D6), 7337–7356. <https://doi.org/10.1029/1999JD901105>
- Schmidt, G. A., Andrews, T., Bauer, S. E., Durack, P. J., Loeb, N. G., Ramaswamy, V., et al. (2023). CERESMIP: A climate modeling protocol to investigate recent trends in the Earth's energy imbalance. *Frontiers in Climate*, 5. <https://doi.org/10.3389/fclim.2023.1202161>
- Scott, R. C., Myers, T. A., Norris, J. R., Zelinka, M. D., Klein, S. A., Sun, M., & Doelling, D. R. (2020). Observed sensitivity of low-cloud radiative effects to meteorological perturbations over the global oceans. *Journal of Climate*, 33(18), 7717–7734. <https://doi.org/10.1175/jcli-d-19-1028.1>
- Sherwood, S. C., & Forest, C. (2023). Opinion: Can uncertainty in climate sensitivity be narrowed further? *EGU sphere*. <https://doi.org/10.5194/egusphere-2023-1647>
- Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, 58(4), e2019RG000678. <https://doi.org/10.1029/2019rg000678>
- Stevens, B., Sherwood, S. C., Bony, S., & Webb, M. J. (2016). Prospects for narrowing bounds on Earth's equilibrium climate sensitivity. *Earth's Future*, 4(11), 512–522. <https://doi.org/10.1002/2016EF000376>
- Sun, M., Doelling, D. R., Loeb, N. G., Scott, R. C., Wilkins, J., Nguyen, L. T., & Mlynarczyk, P. (2022). Clouds and the Earth's radiant energy system (CERES) FluxByCldTyp edition 4 data product. *Journal of Atmospheric and Oceanic Technology*, 39(3), 303–318. <https://doi.org/10.1175/jtech-d-21-0029.1>

- Terai, C. R., Klein, S. A., & Zelinka, M. D. (2016). Constraining the low-cloud optical depth feedback at middle and high latitudes using satellite observations. *Journal of Geophysical Research: Atmospheres*, 121(16), 9696–9716. <https://doi.org/10.1002/2016JD025233>
- Terai, C. R., Zhang, Y., Klein, S. A., Zelinka, M. D., Chiu, J. C., & Min, Q. (2019). Mechanisms behind the extratropical stratiform low-cloud optical depth response to temperature in ARM site observations. *Journal of Geophysical Research: Atmospheres*, 124(4), 2127–2147. <https://doi.org/10.1029/2018JD029359>
- Vial, J., Albright, A. L., Vogel, R., Musat, I., & Bony, S. (2023). Cloud transition across the daily cycle illuminates model responses of trade cumuli to warming. *Proceedings of the National Academy of Sciences of the United States of America*, 120(8), e2209805120. <https://doi.org/10.1073/pnas.2209805120>
- Vogel, R., Albright, A. L., Vial, J., George, G., Stevens, B., & Bony, S. (2022). Strong cloud–circulation coupling explains weak trade cumulus feedback. *Nature*, 612(7941), 696–700. <https://doi.org/10.1038/s41586-022-05364-y>
- Wall, C. J., Storelvmo, T., Norris, J. R., & Tan, I. (2022). Observational constraints on southern ocean cloud-phase feedback. *Journal of Climate*, 35(15), 5087–5102. <https://doi.org/10.1175/JCLI-D-21-0812.1>
- Webb, M., Senior, C., Bony, S., & Morcrette, J. J. (2001). Combining ERBE and ISCCP data to assess clouds in the Hadley Centre, ECMWF and LMD atmospheric climate models. *Climate Dynamics*, 17(12), 905–922. <https://doi.org/10.1007/s003820100157>
- Webb, M. J., Andrews, T., Bodas-Salcedo, A., Bony, S., Bretherton, C. S., Chadwick, R., et al. (2017). The cloud feedback model intercomparison project (CFMIP) contribution to CMIP6. *Geoscientific Model Development*, 10(1), 359–384. <https://doi.org/10.5194/gmd-10-359-2017>
- Yue, Q., Kahn, B. H., Fetzner, E. J., Wong, S., Frey, R., & Meyer, K. G. (2017). On the response of MODIS cloud coverage to global mean surface air temperature. *Journal of Geophysical Research: Atmospheres*, 122(2), 966–979. <https://doi.org/10.1002/2016JD025174>
- Zelinka, M. (2021). mzelinka/cloud-radiative-kernels [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.5514137>
- Zelinka, M. D., & Hartmann, D. L. (2010). Why is longwave cloud feedback positive? *Journal of Geophysical Research*, 115(D16), D16117. <https://doi.org/10.1029/2010JD013817>
- Zelinka, M. D., Klein, S. A., & Hartmann, D. L. (2012). Computing and partitioning cloud feedbacks using cloud property histograms. Part I: Cloud radiative kernels. *Journal of Climate*, 25(11), 3715–3735. <https://doi.org/10.1175/jcli-d-11-00248.1>
- Zelinka, M. D., Klein, S. A., Qin, Y., & Myers, T. A. (2022). Evaluating climate models' cloud feedbacks against expert judgment. *Journal of Geophysical Research: Atmospheres*, 127(2), e2021JD035198. <https://doi.org/10.1029/2021JD035198>
- Zelinka, M. D., Klein, S. A., Taylor, K. E., Andrews, T., Webb, M. J., Gregory, J. M., & Forster, P. M. (2013). Contributions of different cloud types to feedbacks and rapid adjustments in CMIP5. *Journal of Climate*, 26(14), 5007–5027. <https://doi.org/10.1175/jcli-d-12-00555.1>
- Zelinka, M. D., Zhou, C., & Klein, S. A. (2016). Insights from a refined decomposition of cloud feedbacks. *Geophysical Research Letters*, 43(17), 9259–9269. <https://doi.org/10.1002/2016GL069917>
- Zhang, Y., Jin, Z., & Sikand, M. (2021). The top-of-atmosphere, surface and atmospheric cloud radiative kernels based on ISCCP-H datasets: Method and evaluation. *Journal of Geophysical Research: Atmospheres*, 126(24), e2021JD035053. <https://doi.org/10.1029/2021JD035053>
- Zhang, Y., & Klein, S. A. (2013). Factors controlling the vertical extent of fair-weather shallow cumulus clouds over land: Investigation of diurnal-cycle observations collected at the ARM southern great plains site. *Journal of the Atmospheric Sciences*, 70(4), 1297–1315. <https://doi.org/10.1175/jas-d-12-0131.1>
- Zhou, C., Liu, Y., & Wang, Q. (2022). Calculating the climatology and anomalies of surface cloud radiative effect using cloud property histograms and cloud radiative kernels. *Advances in Atmospheric Sciences*, 39(12), 2124–2136. <https://doi.org/10.1007/s00376-021-1166-z>
- Zhou, C., Zelinka, M. D., Dessler, A. E., & Yang, P. (2013). An analysis of the short-term cloud feedback using MODIS data. *Journal of Climate*, 26(13), 4803–4815. <https://doi.org/10.1175/jcli-d-12-00547.1>
- Zhou, C., Zelinka, M. D., & Klein, S. A. (2016). Impact of decadal cloud variations on the Earth's energy budget. *Nature Geoscience*, 9(12), 871–874. <https://doi.org/10.1038/NGEO2828>