LGM paleoclimate constraints inform cloud parameterizations and equilibrium climate sensitivity in CESM2

Jiang Zhu[1], Bette L. Otto-Bliesner[1], Esther C. Brady[1], Christopher J. Poulsen[2], Jonah K. Shaw[3], Jennifer E. Kay[3]

[1]*Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, Colorado, USA*

[2]*Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, Michigan, USA*

[3]*Department of Atmospheric and Oceanic Sciences and Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, USA*

Key Points:

- Excessive LGM global cooling in CESM2 is attributed to the simulation of cloud microphysical processes including the ice nucleation
- A new configuration (PaleoCalibr) with a cloud-ice-number limiter removed and microphysical substepping produces realistic LGM temperature
- PaleoCalibr simulates realistic modern climate and clouds with reduced equilibrium climate sensitivity and cloud feedback

Plain Language Summary: The Community Earth System Model version 2 (CESM2) shows a much higher equilibrium climate sensitivity (ECS > 5°C) than its predecessor models (≤ 4°C), which, if true, implies a greater future warming than previously thought and a more severe challenge for climate adaptation and mitigation. Therefore, it is critical to determine whether the high ECS is realistic and what causes its increase. In a previous study, we suggested that the high ECS is likely unrealistic because CESM2 simulates excessive cooling for an ice age climate—the Last Glacial Maximum (LGM; ~21,000 years ago). In this study, we investigate which aspects of CESM2 are responsible for the extreme LGM cooling and the high ECS. We find that the simulated LGM climate is very sensitive to treatments of cloud microphysical processes, and that removing an inappropriate limiter on cloud ice number and using a smaller time-step size in the microphysics largely eliminate the excessive LGM cooling. With these microphysical modifications, CESM2 simulates a much lower ECS (~4°C) and matches present-day observations well. Our study suggests that an ECS >5°C is likely unrealistic and highlights the importance of using past climates to inform and validate the model development including the treatment of clouds.

Abstract: The Community Earth System Model version 2 (CESM2) simulates a high equilibrium climate sensitivity (ECS > 5°C) and a Last Glacial Maximum (LGM) that is substantially colder than proxy temperatures. In this study, we use the LGM global temperature from geological proxies as a benchmark to examine the role of cloud parameterizations in simulating the LGM cooling in CESM2. Through substituting different versions of cloud schemes in the atmosphere model, we attribute the excessive LGM cooling to the new schemes of cloud microphysics and ice nucleation. Further exploration suggests that removing an inappropriate limiter on cloud ice number (NoNimax) and decreasing the time-step size (substepping) in cloud microphysics largely eliminate the excessive LGM cooling. NoNimax produces a more physically consistent treatment of mixed-phase clouds, which leads to more cloud ice content and a weaker shortwave cloud feedback over mid-to-high latitudes and the Southern Hemisphere subtropics. Microphysical substepping further weakens the shortwave cloud feedback. Based on NoNimax and microphysical substepping, we have developed a paleoclimate-calibrated CESM2 (PaleoCalibr), which simulates well the observed 20th century warming and spatial characteristics of key cloud and climate variables. PaleoCalibr has a lower ECS (~4°C) and a 20% weaker aerosol-cloud interaction than CESM2. PaleoCalibr represents a physically and numerically better treatment of cloud microphysics and, we believe, is a more appropriate tool than CESM2 in climate change studies, especially when a large climate forcing is involved. Our study highlights the unique value of paleoclimate constraints in informing the cloud parameterizations and ultimately the future climate projection.

1. Introduction

The Community Earth System Model version 2 (CESM2) is the newest and most comprehensive model of the CESM family and is a participant in the Coupled Model Intercomparison Projects phase 6 (CMIP6) (Bacmeister et al., 2020; Danabasoglu et al., 2020; Meehl, Arblaster, et al., 2020). One of the most conspicuous differences between CESM2 and its predecessor models is its high equilibrium climate sensitivity (ECS; Bacmeister et al., 2020; Bitz et al., 2011; Gettelman et al., 2019; Gettelman, Kay, & Shell, 2012; Kiehl, Shields, Hack, & Collins, 2006). In the early versions of CESM (the Climate System Model version 1, the Community Climate System Model versions 2–4, and CESM1), ECS ranges from 2.0 to 4.0 K, increasing with the model version and spanning the likely (66%) range from multiple synthesis reports (Figure 1; Charney et al., 1979; IPCC, 2013; Sherwood et al., 2020). In CESM2, ECS has risen to values higher than 5.0°C and well beyond the likely range in different synthesis reports including the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC, 2021). These increases of ECS with CESM model versions have been attributed to increases of model resolution and improvements of physical parameterizations (Bacmeister et al., 2020; Bitz et al., 2011; Gettelman et al., 2019; Gettelman et al., 2012; Kiehl et al., 2006). Specifically, the higher ECS in CESM2, configured with the Community Atmosphere Model version 6 (CAM6), than that in CESM1 with CAM5 (hereafter CESM1) is attributed to changes in the atmospheric parameterizations of stratiform cloud microphysics, unified turbulence, ice nucleation, and convection, as well as the adjustment of aerosol-cloud interactions to match the 20th century temperature record (Gettelman et al., 2019). A high ECS similar to that of CESM2 has been reported in other CMIP6 models and similarly attributed to the simulation of cloud processes (Zelinka et al., 2020).

Whether the high ECS of CESM2 and many other CMIP6 models is realistic remains uncertain and is difficult to address using evidence from present-day observations. On the one hand, CESM2 reproduces well the magnitude of the 20th century global warming in instrumental records and outperforms CESM1 in many observation-based climate metrics (Danabasoglu et al., 2020). In particular, CESM2 simulates a more realistic cloud phase distribution with more supercooled liquid water over the Southern Ocean, largely correcting a major model deficiency in CESM1 (Bjordal, Storelvmo, Alterskjær, & Carlsen, 2020; Gettelman et al., 2020; Kay et al., 2016; Kay et al., 2012). The higher amount of supercooled liquid water is attributed to the updated ice nucleation and cloud microphysical schemes in the atmosphere model and should lead to less cloud phase transitioning (ice to liquid) in response to warming and a weaker (less negative) cloud-phase feedback than in CESM1 (Gettelman et al., 2020). Thus, the stronger cloud feedback and higher ECS in CESM2 is an expected outcome of model improvements (Bjordal et al., 2020; Frey & Kay, 2018; Tan, Storelvmo, & Zelinka, 2016). On the other hand, process understanding from satellite observations suggests that the high-ECS models including CESM2 overestimate the cloud feedback over tropical shallow cumulus regions (Cesana & Del Genio, 2021; Myers et al., 2021). In addition, recent work shows models (including CESM2) underestimate a negative cloud feedback from cloud lifetime changes (Mülmenstädt et al., 2021). There is no doubt that the representation of cloud feedbacks in climate models remains as a large source of uncertainty in climate model projections. Thus, CESM2's successful simulation of the 20th century warming could result from coexisting and compensating model biases due to excessive sensitivities to both aerosol and greenhouse gas (GHG) increases. In this case, the resultant cooling and warming during the historic period offset each other (Kiehl, 2007; Meehl, Senior, et al., 2020; C. Wang, Soden, Yang, & Vecchi, 2021).

Paleoclimate constraints represent a unique and independent way to assess the climate sensitivity of models and consist of performing paleoclimate simulations that incorporate reconstructed climate forcings and assessing them against proxy reconstructions of paleotemperature (e.g. Manabe & Broccoli, 1985). Simulations of the Last Glacial Maximum (LGM; an extreme ice-age climate of ~21,000 years ago) have been performed using many versions of the CESM models and exhibit a close relationship between global cooling and ECS (Figure 1; correlation coefficient = –0.97) (Brady, Otto-Bliesner, Kay, & Rosenbloom, 2013; Otto-Bliesner et al., 2006; Shin et al., 2003; Zhu et al., 2017; Zhu et al., 2021; Zhu & Poulsen, 2021). CESM2, for instance, has the highest ECS (5.6°C) and also simulates the lowest LGM global temperature among the CESM models, a temperature that is at least 5°C lower than a recent observationally based estimate and the CESM1 LGM global temperature (Tierney et al., 2020; Zhu et al., 2021). CESM2 also overestimates global and regional temperature responses for past warm climates including the Early Eocene (an extreme greenhouse climate of ~50 million years ago) and the Pliocene (the most recent warm climate of ~3.2 million years ago with atmospheric CO_2 comparable to today's) (Feng, Otto-Bliesner, Brady, & Rosenbloom, 2020; Zhu, Poulsen, & Otto-Bliesner, 2020). Taken together, these paleoclimate simulations suggest that CESM2 is too sensitive to large external forcings and that its high ECS and strong cloud feedback are likely unrealistic. The excessive cooling in the CESM2 LGM simulation has been attributed to the strong shortwave cloud feedback in the Southern Hemisphere subtropics and mid-to-high latitudes (Zhu et al., 2021). However, it remains unclear which aspects of the cloud feedback processes (such as processes related to stratiform cloud microphysics, unified turbulence, ice nucleation, and convection) in CESM2 are causing the unrealistic climate sensitivity.

In this study, we use LGM constraints to examine details of the cloud feedback processes in CESM2 and to develop a paleoclimate-calibrated version of CESM2 that has a realistic sensitivity to LGM forcings. We adopt the fully coupled LGM configuration in Zhu et al. (2021) and utilize the fact that CESM2 with CAM5 simulates a much more realistic LGM global surface temperature than with CAM6. We evaluate the impact of individual CAM6 cloud schemes on simulated LGM global cooling through simulations in which CAM6 schemes are replaced, one at a time, with older CAM5 schemes. Additionally, we explore physical and numerical aspects of key cloud parameterizations. Finally, we compare the paleoclimate-calibrated version of CESM2 to present-day observations including the scale-aware and definition-aware diagnostics available in satellite simulators. Our study demonstrates that paleoclimate information in concert with modern observations should be used to evaluate cloud parameterizations, which critically determine climate sensitivity.

2. Models and experiments

CESM2 consists of state-of-the-art models of the atmosphere, ocean, land, sea ice, and river and has the capability to simulate ice-sheet dynamics (Danabasoglu et al., 2020). Among the substantial science and infrastructure improvements from CESM1 to CESM2, updates to the cloud-related parameterizations in CAM6 are the primary reason for the high sensitivity to external forcings (Gettelman et al., 2019; Zhu et al., 2021). Specifically, the separate schemes of the moist turbulence in planetary boundary layer, shallow convection, and cloud macrophysical quantities have been replaced with a unified treatment, the Cloud Layers Unified by Binormals (CLUBB; Bogenschutz et al., 2013; Larson & Golaz, 2005). CLUBB is a higher-order turbulence closure scheme that uses a double-Gaussian probability density function to provide a self-consistent closure treatment of higher-order turbulence moments of vertical velocity,

148 temperature, and moisture, as well as boundary layer cloud properties of both stratocumulus and
149 cumulus. CAM6 uses an updated cloud microphysics scheme (MG2) that predicts rather than
150 diagnoses the mass and number concentration of rain and snow (Gettelman & Morrison, 2015).
151 In addition, MG2 links the mixed-phase ice nucleation to aerosols, rather than just temperature,
152 through the implementation of a classical-theory-based heterogeneous ice nucleation scheme
153 (HetFrz; Hoose, Kristjánsson, Chen, & Hazra, 2010; Y. Wang, Liu, Hoose, & Wang, 2014).
154 Additional updates and modifications have been implemented to schemes of aerosols, deep
155 convection, orographic gravity wave, and boundary layer form drag (Danabasoglu et al., 2020).

156 We employ the same LGM initial and boundary conditions as in Zhu et al. (2021). GHGs
157 are 190 ppm, 375 ppb, and 200 ppb for CO_2, CH_4, and N_2O, respectively. Ice sheets are from the
158 ICE-6G reconstruction at 21 ka (thousand years before present) with changes in land surface
159 properties, surface topography, and land-sea mask (Peltier, Argus, & Drummond, 2015). Earth
160 orbital parameters are fixed at the 21-ka values. Preindustrial aerosol emissions and vegetation
161 cover are used in all the LGM simulations. Similar to Zhu et al. (2021), coupled preindustrial
162 (PI) and LGM simulations are run with prescribed satellite vegetation phenology (unless noted),
163 which allows us to focus on the radiative climate feedback without the need to be concerned
164 about the vegetation phenology feedback. Different from Zhu et al. (2021), a lower horizontal
165 resolution of the atmosphere and land is used to save computing resources (1.9×2.5° instead of
166 0.9×1.25°; referred to as FV2 and FV1, respectively). CESM2 FV2 differs from FV1 in the
167 tuning parameters, which are required to achieve an overall top-of-atmosphere (TOA) energy
168 balance for the preindustrial simulation.

169 We perform paired PI and LGM simulations using different configurations of the
170 atmosphere model within the fully coupled CESM2 framework (Table 1). The first two

configurations use CAM6 and CAM5 as the atmosphere component model, respectively (referred to *CAM6* and *CAM5*; hereafter italic font is used for a specific CESM2 configuration). To explore the reason for the greater LGM cooling in *CAM6* than in *CAM5*, additional sensitivity configurations are tested with one cloud scheme in CAM6 either replaced with the older CAM5 version or altered from the default setting (*cf.* Gettelman et al., 2019). In *HetFrzOff*, we use the *CAM6* configuration, except that the new heterogeneous ice nucleation scheme (HetFrz) is replaced with the older scheme in CAM5. In *ClubbOff*, we replace the unified moist turbulence scheme (CLUBB) in CAM6 with the corresponding CAM5 schemes. In *Mg2Off*, we replace the new cloud microphysics scheme (MG2) with the older version (MG1). Considering its overall importance, we developed additional configurations (*NoNimax* and *Mg2Sub8*) to further examine details of the cloud microphysics (see Section 3.2 and 3.3 for the rationale for these sensitivity configurations). In *NoNimax*, a limiter on the cloud ice number concentration is removed in MG2. In *Mg2Sub8*, a microphysical substep of 8 is used (the default value being 1), which decreases the MG2 time-step size from 300s to 37.5s. An additional configuration (*NnSub8*) that combines *NoNimax* and *Mg2Sub8* is also tested (substep numbers of 4 and 16 are also performed but only briefly discussed in this paper). We emphasize that no parameter tuning is performed in any of the configurations, so the difference between *CAM6* and a sensitivity configuration is due to the cloud scheme or modification in question. These fully coupled simulations with various configurations are performed for 100 model years after initializing from the same PI or LGM state. Although many of the simulations have not reached equilibrium in surface climate after 100 model years (Table 1), they are sufficiently integrated to demonstrate the sensitivity of the simulated LGM cooling and cloud feedback to individual cloud schemes and modifications (see results below). Averages of the last 30 years of each simulation are used for analysis.

194 To directly show the impacts of each configuration on ECS and to link the cloud feedbacks
195 in paleoclimate and present-day climate simulations, paired PI and 2×CO_2 simulations with each
196 CESM2 configuration are performed using a slab ocean model (SOM). The same mixed layer
197 depth and heat transport convergence ("q-flux" hereafter; derived from a coupled PI simulation)
198 are prescribed in each SOM simulation. No parameter tuning is performed for the SOM
199 simulations except for *ClubbOff*, in which the relative humidity threshold for low clouds (rhminl)
200 is increased from 0.95 to 0.99 to prevent the model from producing a PI global mean surface
201 temperature (GMST) higher than 20°C. Each SOM simulation is carried out for 80 years and has
202 reached equilibrium (TOA net radiation < |0.1| W m^{-2}) with the last 30 years used for calculation
203 of ECS (denoted as ECS_{SOM}) and the shortwave cloud feedback.

204 We use the approximate partial radiative perturbation method (APRP) to quantify the
205 shortwave cloud feedback (Taylor et al., 2007). APRP uses monthly model output of radiation
206 fields to build a simplified radiation model and quantify the shortwave feedbacks. The shortwave
207 cloud feedback parameters in the paired PI and LGM in a fully coupled configuration and the
208 paired PI and 2×CO_2 in a SOM configuration are denoted as $\lambda_{sw\_cld\_LGM}$ and $\lambda_{sw\_cld\_2\times}$,
209 respectively. The longwave feedback in the simulations is not quantified because it is not a major
210 driver for the differences in ECS and the LGM temperature response between CESM2
211 configurations (Gettelman et al., 2019; Zhu et al., 2021).

212 We use the LGM proxy sea-surface temperature (SST)-derived global cooling of 5.6°C
213 (4.6–6.8°C; 95% confidence interval) (Tierney et al., 2020) as a benchmark to evaluate the
214 performance of each CESM2 configuration.

After the individual cloud schemes and changes are evaluated, a paleoclimate-calibrated CESM2 configuration (PaleoCalibr) in FV2 is developed. A suite of DECK (Diagnostic, Evaluation and Characterization of Klima) simulations and a CMIP6 historical simulation are performed (Eyring et al., 2016), which follows the experimental setup of the simulations using the standard CESM2. Results from the PaleoCalibr preindustrial, historical Atmospheric Model Intercomparison Project (AMIP), historical, and abrupt $4\times CO_2$ simulations are discussed. The historical AMIP simulation is run with the satellite simulator to facilitate a direct comparison with satellite observations (Swales, Pincus, & Bodas-Salcedo, 2018). We also have additional atmosphere-only simulations with prescribed SST and sea ice from observation at 2000 CE that are used to quantify the aerosol-cloud interaction and to test sensitivity to some parameters (see details below).

3. Sensitivity of LGM global temperature to cloud microphysical processes

3.1 Role of individual cloud schemes

CESM2(CAM6) with a ~2° atmosphere significantly overestimates the LGM global cooling, which is consistent with the results with a ~1° atmosphere in Zhu et al. (2021). The LGM ΔGMST in *CAM6* reaches –9.0°C with a large TOA imbalance of approximately –1.0 W m^{-2} after 100 model years, suggesting that if the simulation were extended further, additional cooling would be expected (red in Figure 2; Table 1). In contrast, ΔGMST in *CAM5* is –6.3°C (brown) and falls within the proxy suggested range of LGM global cooling (gray patch). Similar to *CAM5*, *Mg2Off* (green) and *HetFrzOff* (orange) have LGM ΔGMSTs of –6.3 and –5.9°C after 100 years, respectively, which also fall within the proxy range. In contrast, *ClubbOff* has a ΔGMST of –8.9°C that is comparable to the *CAM6* value (blue versus red).

The different LGM ΔGMSTs in these configurations are linked to the strength of shortwave cloud feedback. $\lambda_{sw\_cld\_LGM}$ in *CAM6* is 0.81 W m^{-2} K^{-1}, more than double the *CAM5* value of 0.29 W m^{-2} K^{-1} (Figure 3 red versus brown; Table 1). $\lambda_{sw\_cld\_LGM}$ in *CAM6* is larger than in *CAM5* over all latitudes, especially over the Southern Hemisphere (SH) subtropics and the Southern Ocean (SO). $\lambda_{sw\_cld\_LGM}$ is 0.49, 0.37, and 0.64 W m^{-2} K^{-1} in *Mg2Off*, *HetFrzOff*, and *ClubbOff*, respectively. In the subtropics, *Mg2Off* produces a $\lambda_{sw\_cld\_LGM}$ comparable to *CAM5* (green vs brown), indicating that the stronger subtropical $\lambda_{sw\_cld\_LGM}$ in *CAM6* than in *CAM5* is largely due to the new cloud microphysics scheme (MG2). Over the SO, *HetFrzOff* produces a $\lambda_{sw\_cld\_LGM}$ comparable to *CAM5* (orange vs brown), indicating that the new ice nucleation scheme (HetFrz) explains the stronger SO $\lambda_{sw\_cld\_LGM}$ in *CAM6*. In addition, *HetFrzOff* produces a weaker shortwave cloud feedback than *CAM6* over the SH subtropics, although not as weak as in *CAM5*, indicating that the ice nucleation of mixed-phase clouds is important for both SH high latitudes and subtropics. *HetFrzOff* simulates greater amounts of ice nucleation particles and cloud ice than *CAM6* (Figure 4b), which, we speculate, produces more negative cloud phase and lifetime feedbacks due to the greater cloud phase transition from ice to liquid in response to warming (Mülmenstädt et al., 2021; Tan et al., 2016).

The tests of individual cloud schemes (*Mg2Off*, *HetFrzOff*, and *ClubbOff*) suggest that the cloud microphysical processes, including those related to mixed-phase and liquid clouds, are important in driving the strong cloud feedback in *CAM6* and are likely responsible for the unrealistically high CESM2 ECS. Nevertheless, the new cloud schemes were developed according to theory and process-level understanding and were found to be critical to the improved simulation of the SO cloud phase distribution (Gettelman et al., 2020). Given that the

259 new schemes in CAM6 are "better physics", we next examine details of the cloud microphysical
260 processes while using these more advanced cloud schemes.

261 **3.2 Role of a cloud-ice-number limiter**

262 A cloud-ice-number limiter (named "nimax" in MG2) sets the maximum allowed cloud ice
263 number and was first designed in MG1 to avoid excessive nucleation with the old ice nucleation
264 scheme and with the long microphysical time step. "nimax" was not modified in MG2 when the
265 new mixed-phase ice nucleation scheme was introduced. As a result, the heterogeneous ice
266 nucleation processes can increase cloud ice mass but not the number concentration, leading to
267 artificially inflated ice particle size and increased sedimentation (Shaw, McGraw, Bruno,
268 Storelvmo, & Hofer, 2021). With "nimax", the cloud ice number also has less freedom to adjust
269 to internal or forced variations. We examine the role of "nimax" through a suite of simulations
270 (*NoNimax*), in which "nimax" is removed to improve the physical consistency in mixed-phase
271 clouds. Other considerations of *NoNimax* include that "nimax" may be no longer needed due to
272 the much shorter microphysical time step (300s in MG2 versus 1800s in MG1) and that the new
273 ice nucleation scheme is more process-based and coded differently from the old scheme (Hoose
274 et al., 2010; Y. Wang et al., 2014).

275 *NoNimax* has minor impact on the preindustrial GMST but warms the LGM by close to
276 3°C in 100 years, leading to a much-improved LGM ΔGMST of –6.8°C (yellow versus red in
277 Figure 2). As expected, *NoNimax* changes the response to LGM forcing by affecting the cloud
278 feedback. The global mean $\lambda_{sw\_cld\_LGM}$ is 0.64 W m^{-2} K^{-1} in *NoNimax*, 20% smaller than the
279 *CAM6* value (Table 1). Zonal mean $\lambda_{sw\_cld\_LGM}$ in *NoNimax* is weaker over the SH subtropics and
280 the mid-to-high latitudes in both hemispheres (yellow versus red in Figure 3). To some degree,

NoNimax impacts cloud properties and feedbacks over the SH subtropics in a similar way as *HetFrzOff*, likely indicating a similar mechanism through increasing cloud ice content (Figure 4b) and strengthening the (negative) cloud phase and lifetime feedback. We note that the projected LGM ΔGMST for *NoNimax* is ~–8°C, estimated by extrapolation using TOA radiation and GMST (yellow in Figure 2d), and indicates that *NoNimax* would likely overestimate the LGM ΔGMST if the simulation were extended beyond 100 years.

NoNimax, with the limiter removed, produces an unrealistic simulation of cloud ice number concentration (Figure 5). The zonal-mean in-cloud ice number concentration in *CAM6* is largely less than 50 L^{-1} below ~400 hPa with maximum centers in the middle troposphere at the mid-latitudes and in the lower troposphere at polar regions. In the stratosphere, the zonal-mean in-cloud ice number reaches values greater than 900 L^{-1}. The high values over the stratosphere likely reflect a model bias, while values over the middle and lower troposphere are roughly of the same order as observations (e.g., DeMott et al., 2010; Patnaude, Diao, Liu, & Chu, 2021). Without the limiter, *NoNimax* simulates a zonal mean in-cloud ice number greater than *CAM6* almost everywhere. Over the Northern Hemisphere (NH) mid-latitudes, the zonal-mean in-cloud ice number reaches values >900 L^{-1} at ~400 hPa and >300 L^{-1} below; these values are roughly an order larger than observations (e.g., DeMott et al., 2010).

3.3 Role of substepping in microphysics

The large overestimation of in-cloud ice number in *NoNimax* motivates us to explore whether substepping in the microphysics helps to improve the simulation. Microphysical substepping decreases the time-step size through increasing the substep number of microphysical calculations per calculation of the other model physical parameterizations. For simplicity, we

perform a suite of atmosphere-only simulations forced by the observed climatological SST and sea ice from 2000 CE. These simulations are run with *NoNimax* and with an increased microphysical substep of 2, 4, 8, 16, and 32, respectively (Figure 5). At the NH mid-latitudes, the zonal mean in-cloud ice number at 400 hPa decreases from ~900 L^{-1} to ~300 L^{-1} with 2 MG2 substeps (*NnSub2*; microphysical time step of 150 s) and to ~100 L^{-1} with 4 substeps (*NnSub4*; microphysical time step of 75 s). At ~700 hPa, the cloud ice number decreases from ~300 to <50 L^{-1} with substeps greater than 8. Overall, the simulated cloud ice number is converging after 8–16 MG2 substeps (a microphysical time step ≤ 37.5s).

Employing microphysical substepping together with *NoNimax* further decreases the shortwave cloud feedback and the simulated LGM global cooling, in addition to improving the simulation of cloud ice number. Three pairs of coupled PI and LGM simulations are performed with *NoNimax* and microphysical substep of 4, 8, and 16 (referred to as *NnSub4*, *NnSub8*, and *NnSub16*, respectively). The LGM ΔGMST are –6.4, –6.5, and –6.5°C after 100 model years in *NnSub4*, *NnSub8*, and *NnSub16*, respectively. The global mean $\lambda_{sw\_cld\_LGM}$ are 0.53, 0.49, and 0.49 W m^{-2} K^{-1}, respectively. The decrease of $\lambda_{sw\_cld\_LGM}$ with the microphysical substepping reaches saturation at 8 substeps: successive substep increases from 1 to 4, 8, and 16 decrease $\lambda_{sw\_cld\_LGM}$ by 0.11, 0.04, and 0.00 W m^{-2} K^{-1}, respectively. Consistent with the global mean, the zonal mean $\lambda_{sw\_cld\_LGM}$ also exhibits convergence with an MG2 substep of 8 or higher (Figure 6). Although the decrease of the LGM global cooling is small from *NoNimax* (with a substep of 1) to *NnSub8* (ΔGMST of –6.8 versus –6.4°C after 100 model years), the projected LGM ΔGMST is much larger (–8.0 versus –7.0°C; yellow versus black in Figure 2d), which is consistent with the large impact on the shortwave cloud feedback (Figure 3) and TOA radiation (Table 1; see also the ECS_{SOM}).

To understand the processes that weaken the cloud feedback with an increase in microphysical substeps, we perform an additional pair of coupled PI and LGM simulations with *CAM6* and 8 microphysical substeps (referred to *Mg2Sub8*). *Mg2Sub8* has the active "nimax" limiter and simulates a low tropospheric cloud ice number like *CAM6* (not shown). Compared to *CAM6*, the global mean $\lambda_{sw\_cld\_LGM}$ in *Mg2Sub8* decreases by 0.09 W m^{-2} K^{-1} from 0.81 to 0.72 W m^{-2} K^{-1} (Table 1). In comparison, $\lambda_{sw\_cld\_LGM}$ decreases by 0.15 W m^{-2} K^{-1} between *NoNimax* and *NnSub8* with "nimax" removed. Both configurations with (*CAM6* and *Mg2Sub8*) and without "nimax" (*NoNimax* and *NnSub8*) consistently show decreases of $\lambda_{sw\_cld\_LGM}$ in the subtropics and SH mid-latitudes (Figure 6). These results suggest that a large part of the weakening of cloud feedback with microphysical substepping is through pathways other than changing cloud ice number, such as the warm rain process (Gettelman et al., 2021).

3.4 Connected cloud feedback between LGM and 2×CO_2 simulations

A strong correlation between $\lambda_{sw\_cld\_LGM}$ and $\lambda_{sw\_cld\_2\times}$ is found across the major configurations that are explored in this study (Figure 3). In the global mean, the correlation coefficient between $\lambda_{sw\_cld\_LGM}$ and $\lambda_{sw\_cld\_2\times}$ is 0.95 (Figure 3d). A similar strong correlation (–0.93) is also found between LGM ΔGMST and ECS_{SOM} among these CESM2 configurations (Table 1). Averaged across all the configurations, $\lambda_{sw\_cld\_2\times}$ is larger than $\lambda_{sw\_cld\_LGM}$ by 0.11 W m^{-2} K^{-1}, which is consistent with previous findings that the shortwave cloud feedback increases with GMST (Zhu & Poulsen, 2020; Zhu, Poulsen, & Tierney, 2019). In the zonal mean, $\lambda_{sw\_cld\_2\times}$ is larger than $\lambda_{sw\_cld\_LGM}$ over the middle-to-high latitudes (Figure 3c), likely linked to the more positive cloud-phase feedback in response to warming than to cooling (Zhu & Poulsen, 2020). $\lambda_{sw\_cld\_LGM}$ is larger than $\lambda_{sw\_cld\_2\times}$ in the tropics, which could be linked to the stronger glacial trade winds and the impact on low clouds through increasing latent heat flux (Zhu et al., 2021).

This high correspondence between global and regional shortwave cloud feedback in paleoclimate and present-day simulations (as well as between the LGM ΔGMST and ECS_{SOM}) supports the notion that paleoclimate information can be used to constrain the cloud feedback and ECS (Zhu et al., 2021). Moreover, $\lambda_{sw\_cld\_LGM}$ is obtained with paired, short PI and LGM simulations of 100 model years, which may still have large GMST trends and TOA energy imbalances (Figure 2; Table 1). The high correlation between $\lambda_{sw\_cld\_LGM}$ and $\lambda_{sw\_cld\_2\times}$ (obtained in equilibrated SOM simulations) suggests that our major findings on the shortwave cloud feedback depend little on the equilibration state of the coupled simulations. This is further supported by the high correlation (0.95) between $\lambda_{sw\_cld\_LGM}$ and $\lambda_{sw\_cld\_2\times}$ if shorter coupled PI and LGM simulations of 50 years are used.

4. A paleoclimate-calibrated configuration of CESM2

As in Zhu et al. (2021); (Zhu et al., 2020), we find that CESM2 with CAM6 is not well suited for simulating climates under large radiative forcing that exceeds the historical range. To mitigate this shortcoming, we develop a paleoclimate-calibrated CESM2 configuration (PaleoCalibr) based on *NnSub8*, although we note that other configurations of CESM2, including *CAM5*, *HetFrzOff*, and *Mg2Off* also produce acceptably realistic LGM global cooling (Figure 2). An advantage of *NnSub8* over these other configurations is that it uses the advanced cloud schemes in CAM6, in particular the ice nucleation and microphysical schemes that are based more on theory or process-level understanding. Another advantage of *NnSub8* is that it represents a minimal departure in model code from *CAM6* and probably its future versions. Moreover, *HetFrzOff* and *CAM5* simulates a negative shortwave cloud feedback over the Southern Ocean at ~50–60°S (Figure 3b), which is inconsistent with the positive values estimated using the satellite observations (Myers et al., 2021). Based on *NnSub8*, PaleoCalibr incorporates additional minor

372 tuning. The CLUBB gamma parameter is lower from 0.280 to 0.275 to decrease the TOA
373 radiation imbalance in the preindustrial simulation. The dust emission scaling factor
374 (dust_emis_fact) is lower from 0.70 to 0.55 to ensure a more realistic global mean dust aerosol
375 optical depth. Additionally, a new and simple limiter on cloud ice number ($Ni < 1000\ L^{-1}$) is
376 added at the end of the microphysical calculations to ensure a realistic simulation of cloud ice
377 number over the stratosphere. This additional cloud-ice-number limiter has no impact on the
378 cloud feedback and LGM cooling, which has been confirmed in additional test simulations (not
379 shown).

380 We perform 500-year simulations for both the preindustrial and LGM using PaleoCalibr
381 and the same experimental setups as the standard CESM2 runs. PaleoCalibr PI has a similar
382 GMST as CESM2 (13.9 versus 14.1°C) and a small TOA energy imbalance ($0.03\ W\ m^{-2}$) at the
383 end of the simulation. PaleoCalibr LGM has a ΔGMST of –6.7°C and a TOA radiation
384 imbalance of $-0.08\ W\ m^{-2}$ (Figure 7). The projected LGM ΔGMST, using a linear regression
385 between LGM GMST and TOA radiation, is approximately –7.3°C in PaleoCalibr, which is
386 marginally too cold when compared with the proxy estimation (Tierney et al., 2020). We contend
387 that the PaleoCalibr LGM is acceptably realistic and suitable for glacial climate research,
388 considering the uncertainty in the ice sheet forcing and the absence of LGM dust forcing in our
389 simulations (Abe-Ouchi et al., 2015; Ohgaito et al., 2018). We note that the land
390 biogeochemistry (BGC) model is inactive in the PaleoCalibr LGM simulation but is active in
391 PaleoCalibr PI. The PI simulation (and the associated historical and abrupt $4\times CO_2$ simulations)
392 with land BGC is consistent with the available standard CESM2 DECK simulations. We decided
393 not to include land BGC in the PaleoCalibr LGM because it produces an extra LGM cooling
394 of >1°C after 100 simulation years (not shown) due to vegetation phenology feedbacks. This

vegetation phenology-induced LGM cooling is consistent with results in CESM1.2 (Zhu & Poulsen, 2021) but we do not know how realistic it is, given that the land vegetation processes are highly parameterized for the present climate and may not work well under a much colder environment with a much lower CO_2 (Lawrence et al., 2019).

4.1 A realistic simulation of the present-day climate

We first evaluate the cloud simulation of PaleoCalibr in an AMIP historical simulation with an active Cloud Feedback Model Intercomparison Project Observational Simulator Package (COSP). We use the Taylor diagram (Taylor, 2001) for a compact visualization of the model performance (Figure 8). The Taylor diagram evaluates a suite of model variables against observations using multiple metrics including the area-weighted pattern correlation and normalized root-mean-squared differences (RMSDs), as well as the relative bias. Cloud observations that are used in the model evaluation include the climatology of TOA cloud radiative forcing from Clouds and the Earth's Radiant Energy System Energy Balanced and Filled Edition-4.1 (CERES-EBAF; Loeb et al., 2018) and the cloud fraction products from the International Satellite Cloud Climatology Project (ISCCP) (Pincus, Platnick, Ackerman, Hemler, & Patrick Hofmann, 2012), the Multiangle Imaging Spectro-Radiometer (MISR) (Marchand, Ackerman, Smyth, & Rossow, 2010), and the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) (Chepfer et al., 2010), as well as the liquid and ice cloud fraction from CALIPSO. Averages between 2000 and 2014 CE are used in the model-data comparison, except that CALIPSO cloud fraction between 2008 and 2020 CE is used.

PaleoCalibr improves the simulations of cloud fraction and its liquid-ice partition over the standard CESM2 but has degradations in cloud radiative forcing. In the total cloud fraction,

417 PaleoCalibr shows smaller centered pattern errors than CESM2, i.e., normalized RMSDs that are
418 closer to one when compared with ISCCP (1.33 versus 1.37; labeled "1"), MISR (1.43 versus
419 1.50; labeled "2"), and CALIPSO (1.40 versus 1.47; labeled "3"). PaleoCalibr cloud fraction also
420 has a greater pattern correlation with CALIPSO (0.90 versus 0.85) than CESM2. The phase
421 partition of cloud fraction in PaleoCalibr shows large improvements over the standard CESM2
422 with greater pattern correlation with the CALIPSO liquid clouds (0.86 versus 0.76; label "4")
423 and smaller centered pattern error in both liquid (1.31 versus 1.38 in the normalized RMSDs)
424 and ice (1.47 versus 1.61 in the normalized RMSDs; Label "5") clouds. In SWCF, PaleoCalibr
425 exhibits a slightly smaller centered pattern error (1.08 vs 1.11; label "6") than CESM2 but has
426 degradations in the pattern correlation with observations (0.85 to 0.88). PaleoCalibr LWCF
427 degrades slightly from CESM2 with larger centered pattern error (0.91 versus 0.99 in the
428 normalized RMSD from observation; label "7"). For all the metrics that are examined in
429 PaleoCalibr and CESM2, the relative biases from observations (marker size in the Taylor
430 diagram) fall within the same category, indicating that the improvements / degradations in
431 PaleoCalibr come from a redistribution of cloud properties across the globe rather than a uniform
432 shift.

433 From a spatial view, PaleoCalibr improves the cloud simulation in the Arctic but shows
434 mixed results over other places (Figure 9). The standard CESM2 overestimates the Arctic cloud
435 fraction in CALIPSO by as much as 20%, which results primarily from a larger modeled liquid
436 cloud fraction. PaleoCalibr largely removes the model bias in CESM2 by simulating a smaller
437 liquid cloud fraction that agrees much better with observation. In the subtropics and mid-
438 latitudes, PaleoCalibr simulates a greater cloud fraction that agrees better with satellite
439 observations (Figure 9a,b), but at the expenses of degradations in the SWCF (Figure 9d),

reflecting a stubborn "too-few-and-too-bright" model bias (Nam, Bony, Dufresne, & Chepfer, 2012). In the deep tropics, PaleoCalibr simulates a smaller cloud fraction than CESM2, which agrees worse with observations. Over the SO, PaleoCalibr cloud fraction and its phase partition are similar to the standard CESM2, suggesting that the improvement from CESM1 to CESM2 in the SO clouds is largely preserved in PaleoCalibr. On average, PaleoCalibr has a more positive SWCF over middle-to-high latitudes and more negative SWCF over the lower latitudes than CESM2.

We next evaluate the coupled simulation of PaleoCalibr in a CMIP historical simulation. PaleoCalibr reproduces the magnitude of the global warming (~1.1°C; Figure 10) from 1850 to 2014 CE in the Hadley Centre-Climate Research Unit Temperature Anomalies (HADCRU4) and the Goddard Institute for Space Studies Surface Temperature Analysis (GISTEMP) (Jones et al., 2012; Lenssen et al., 2019). The large internal variability relative to the forced response prevents us from a more quantitative evaluation of the temporal characteristics of the historical simulation (green in Figure 10; Kay et al., 2015), but a visual examination suggests that the performance of PaleoCalibr is as good as the ensemble of three CESM2 historical simulations (red, orange, and pink in Figure 10).

The spatial characteristics of the PaleoCalibr historical simulation match observations and reanalysis reasonably well with skills largely similar to the standard CESM2, which is summarized in a Taylor diagram (Figure 11). The model performance is evaluated against observations of SST from the Extended Reconstructed Sea Surface Temperature version 5 (ERSST; Huang et al., 2017), precipitation from the Global Precipitation Climatology Project version 2.3 (GPCP; Adler et al., 2018), TOA cloud radiative forcing from CERES-EBAF (Loeb et al., 2018), and the surface air temperature, sea-level pressure, and zonal wind at 300 hPa from

463 the ERA5 (Hersbach et al., 2020). All the metrics are calculated for the mean fields averaged
464 between 1979 and 2014, except for the CERES-EBAF cloud radiative forcing that is averaged
465 between 2000 and 2014. The mean bias of all the fields examined is in the same category
466 between PaleoCalibr and CESM2 (marker size in the Taylor diagram), except for precipitation
467 that has a larger relative bias in PaleoCalibr (11.2% versus 9.6%). Statistics for surface air
468 temperature (labeled "2") and longwave cloud forcing (labeled "5") are very similar between the
469 PaleoCalibr and CESM2. The shortwave cloud forcing (SWCF) (labeled "4") shows a larger
470 difference, which is consistent with results from the AMIP simulations. As a result of the more
471 positive SWCF over middle-to-high latitudes and more negative SWCF over the lower latitudes,
472 SST in PaleoCalibr is warmer over mid-to-high latitudes and colder over lower latitudes with a
473 larger centered pattern error than in CESM2 (label "3" in Figure 11; Figure 12). Small
474 degradation in the normalized RMSD is also found in zonal wind at 300 hPa (label "6") and
475 precipitation (label "7") with the former having insufficient spatial variance and the latter having
476 too much spatial variance.

477 Based on the above analysis, we conclude that PaleoCalibr performs as good as the
478 standard CESM2 in the simulation of key cloud and climate observations. We note that some
479 aspects of PaleoCalibr simulations could be improved through additional parameter tuning that
480 may have little net impact on climate sensitivity, but an extensive re-tuning of the model is
481 beyond the scope of this study.

482 **4.2 A lower ECS and weaker cloud-aerosol interactions**

483 ECS is quantified to be 3.9°C in PaleoCalibr by regressing GMST and TOA radiation in an
484 abrupt $4\times CO_2$ simulation of 150 years, which is much lower than the 5.3°C in CESM2 (Figure

13). If the first 20 years are used in the regression, the difference in the estimated ECS between PaleoCalibr and CESM2 is much smaller (ΔECS = 0.2°C; 3.2 versus 3.4°C), suggesting that the effect from PaleoCalibr changes manifests mostly at timescales longer than 1–2 decades. With additional SOM simulations with an abrupt $2\times CO_2$, ECS_{SOM} are estimated to be 4.0 and 6.1°C in PaleoCalibr and CESM2, respectively. The lower ECS in PaleoCalibr is consistent with the much smaller magnitude of LGM global cooling (Figure 7). The shortwave cloud feedback averaged over the last 20 years of the $4\times CO_2$ simulations are 0.52 and 0.74 W m^{-2} K^{-1} in PaleoCalibr and CESM2, respectively. The shortwave cloud feedback shows lower values over the mid-latitudes and the SH subtropics (Figure 14a). Compared to CESM2, the reduced shortwave cloud feedback in PaleoCalibr is more consistent with the observation-based estimates (Ceppi & Nowack, 2021; Cesana & Del Genio, 2021; Myers et al., 2021).

PaleoCalibr simulates an aerosol-cloud interaction (ACI) that is 20% weaker than CESM2. ACI is quantified as the change in the net cloud radiative forcing between a pair of atmosphere-only simulations with aerosol emissions at 2000 CE and 1850 CE, forced with the same observational SST and sea ice (IPCC, 2013). ACI are –1.3 and –1.7 W m^{-2} in PaleoCalibr and CESM2, respectively. ACI weakening is mostly found at mid-to-high latitudes, where we also observe decreases in the shortwave feedback (Figure 14). This highlights the fact that aerosol forcing and cloud feedback are not independent variables (Gettelman et al., 2019; Kiehl, 2007). A weaker GHG-induced warming and a weaker aerosol-induced cooling may explain the comparable historical warming between PaleoCalibr and CESM2 (Meehl, Senior, et al., 2020; C. Wang et al., 2021).

5. Conclusions and discussion

In this study, we have investigated the impact of key cloud parameterizations of CESM2 on the simulated LGM global temperature through coupled simulations with individual CAM6 schemes substituted one-at-a-time by older CAM5 schemes. Our investigation takes advantage of the fact that CESM2(CAM6) (referred to *CAM6*) simulates an excessive LGM cooling but the CESM2(CAM5) (*CAM5*) LGM simulation falls within the proxy suggested range (4.6–6.8°C) (Tierney et al., 2020). The different performances of the LGM simulations imply that changes in the cloud parameterizations between CAM5 and CAM6 are likely responsible for the excessive LGM cooling and therefore the unrealistic climate sensitivity of CESM2. Our simulations show that the substitution of CAM6 ice nucleation or cloud microphysics scheme with the CAM5 version (*HetFrzOff* or *Mg2Off*) produces a much more realistic LGM than the default *CAM6*. In contrast, substituting the moist turbulence scheme to the CAM5 version (*ClubbOff*) has a small impact. Specifically, the LGM ΔGMST after 100 model years are –9.0, –6.3, –6.3, –5.9, and –8.9°C in the LGM simulations using *CAM6*, *CAM5*, *Mg2Off*, *HetFrzOff*, and *ClubbOff*, respectively. The different magnitude of LGM cooling in these simulations is primarily caused by variations of the shortwave cloud feedback, which are 0.81, 0.29, 0.49, 0.37, and 0.64 W m^{-2} K^{-1}, respectively. These sensitivity tests suggest that the increased climate sensitivity in CESM2 is largely determined by cloud microphysical processes, which has guided our further examination.

Further exploration suggests that the removal of a limiter on cloud ice number ("nimax") in the microphysical scheme (*NoNimax*) weakens the shortwave cloud feedback from 0.81 to 0.64 W m^{-2} K^{-1} and reduces the simulated LGM ΔGMST to –6.9°C after 100 model years. The original implementation of "nimax" in the CESM2 microphysical scheme is physically inconsistent (Shaw et al., 2021). Without the limiter, the cloud ice number also has more

freedom to adjust to internal or forced variations. As a result, *NoNimax* simulates a larger cloud ice content, which presumably causes a more negative cloud phase and lifetime feedback (Mülmenstädt et al., 2021; Tan et al., 2016). Although *NoNimax* has a more physically based treatment of the ice nucleation processes and a largely improved LGM simulation, its projected LGM ΔGMST is too low (~–8°C) and the cloud ice number in the preindustrial simulation is about one order greater than observations.

Building upon *NoNimax*, decreasing the microphysical time-step size (microphysical substepping) lowers the cloud ice number to the observational range and further improves the LGM simulation by decreasing the cloud feedback. Through a suite of simulations with the "nimax" removed, we find that an increase of the microphysical substep from 1 to 8 (a decrease of the microphysical time-step size from 300s to 37.5s; *NnSub8*) produces a realistic cloud ice number and a convergent solution in the shortwave cloud feedback. Specifically, *NnSub8* reduces the LGM ΔGMST from –6.9°C in *NoNimax* to –6.4°C and decreases the shortwave cloud feedback from 0.64 to 0.49 W m^{-2} K^{-1} after 100 model years. The decrease of the shortwave cloud feedback (0.15 W m^{-2} K^{-1}) is largely through processes other than decreasing cloud ice number, which is supported by additional simulations with the microphysical substepping and the "nimax" limiter (*Mg2Sub8*) that show a decrease of the shortwave cloud feedback by ~0.1 W m^{-2} K^{-1} (compared to *CAM6*). Findings here are generally consistent with results in Santos, Caldwell, and Bretherton (2020), who find that multiple microphysical processes are poorly resolved with a default microphysical time step in atmosphere-only simulations of the present day. Our results further show that these poorly resolved processes have an impact on the cloud feedback and climate sensitivity and that a microphysical substep of

552 8 and higher seems to produce a convergent shortwave cloud feedback. Further study is needed
553 to examine which microphysical processes are responsible for the time-step dependence.

554 A paleoclimate-calibrated CESM2 configuration (PaleoCalibr) is developed, which consists
555 of removing the "nimax" limiter (*NoNimax*) and substepping (*Mg2Sub8*) in the cloud
556 microphysics, as well as a minimal model tuning. A historical simulation using PaleoCalibr
557 reproduces the observed 20th century warming. PaleoCalibr also simulates the spatial
558 characteristics of key cloud and climate variables very well with improvements in the cloud
559 fraction and its phase partition. In the Arctic, PaleoCalibr corrects a large model bias in CESM2
560 by removing the excessive liquid clouds. In the subtropics, PaleoCalibr simulates a slightly
561 greater cloud cover that agrees better with satellite observations but introduces a compensating
562 bias by decreasing the shortwave cloud forcing. Due to the changes in shortwave cloud forcing,
563 PaleoCalibr exhibits small degradations from CESM2 in the spatial pattern of SST and the
564 associated wind and precipitation fields in coupled simulations. Performance of PaleoCalibr
565 could be further improved through parameter tuning, which we did not perform for simplicity.

566 PaleoCalibr has a lower ECS (~4°C) than the standard CESM2 (~5–6°C) and realistic
567 LGM global cooling (~7 °C). PaleoCalibr simulates a 20% smaller aerosol cloud interaction.
568 Compared to the standard CESM2, removing the cloud-ice-number limiter represents a more
569 physically consistent treatment of the cloud ice nucleation process and microphysical
570 substepping provides better numerical performance of cloud microphysical processes. We
571 believe PaleoCalibr is a more useful tool than CESM2 in climate change studies, especially when
572 a large climate forcing is involved.

573 We note that all the test simulations and the paleoclimate-calibrated configuration in this
574 study use the CESM2 with a ~2° atmosphere. We expect that the overall impact from removing
575 the "nimax" and microphysical substepping is largely independent of model resolution, but some
576 details including the tuning parameters need to be examined if the ~1° atmosphere model is used.
577 For example, the exact microphysical substep number that produces a converging cloud feedback
578 could be different due to the different model resolution and parameters. We note further that we
579 have intentionally not performed parameter tuning for each CESM2 sensitivity configuration,
580 which leads to a warmer preindustrial GMST in coupled and SOM simulations in some
581 configurations (e.g., GMST is 17.8°C in the PI SOM simulation with *Mg2Off*). As a consequence
582 and caveat, part of the differences in ECS and cloud feedback between CESM2 configurations
583 could be caused by their state dependence, instead of changes in cloud treatment. However, we
584 believe that the impact of state dependence on the shortwave cloud feedback is small (<< 0.1 W
585 m^{-2} K^{-1}) in simulations presented here, considering that the shortwave cloud feedback in CESM2
586 increases from 0.97 to 1.07 W m^{-2} K^{-1} when the background GMST increases from 15.2 to
587 20.7°C (Zhu & Poulsen, 2020).

588 Our study highlights the unique value of paleoclimate constraints in informing the cloud
589 parameterizations and ultimately future climate projections. Among the CESM2 configurations
590 that are explored in this study, a close correlation is found in cloud feedbacks and temperature
591 responses between CO_2 increasing and paleoclimate simulations (Table 1 and Figures 1 and 3d),
592 which indicates that a common set of physical processes are active in past and future climates
593 and serves as the physical basis for a paleo-constraint on clouds and climate sensitivity.
594 Although the paleoclimate forcing and global temperature response do not provide process-level
595 constraints on cloud feedback processes, they serve as a critical "out-of-sample" test for the

596 cloud parameterizations that are usually developed to match the present-day observations. We
597 encourage the use of paleoclimate constraints as an important tool in future model development
598 and validation, as our knowledge of past climates continues to improve and climate models
599 become more complex. We have ongoing work to evaluate the performance of the paleoclimate
600 calibrated CESM2 in simulating past warm climates, such as the Early Eocene (Zhu et al., 2020).

601

602 **Acknowledgements:** The authors thank A. Gettelman, J. Bacmeister, R. Neale, and X. Zhao for
603 helpful discussions. This work was supported by National Science Foundation grants 2002397 to
604 C. Poulsen. CESM model code is available through the National Center for Atmospheric
605 Research GitHub repository (https://github.com/ESCOMP/CESM/releases/tag/cesm2.1.1). The
606 CESM project is supported primarily by the National Science Foundation (NSF). This material is
607 based upon work supported by the National Center for Atmospheric Research (NCAR), which is
608 a major facility sponsored by the NSF under Cooperative Agreement No. 1852977. Computing
609 and data storage resources, including the Cheyenne supercomputer (doi:10.5065/D6RX99HX),
610 were provided by the Computational and Information Systems Laboratory (CISL) at NCAR. All
611 simulation data will be published via the NCAR Climate Data Gateway
612 (https://www.earthsystemgrid.org/).

613

References:

Abe-Ouchi, A., Saito, F., Kageyama, M., Braconnot, P., Harrison, S. P., Lambeck, K., . . . Takahashi, K. (2015). Ice-sheet configuration in the CMIP5/PMIP3 Last Glacial Maximum experiments. *Geoscientific Model Development, 8*(11), 3621-3637. doi:10.5194/gmd-8-3621-2015

Adler, R. F., Sapiano, M. R. P., Huffman, G. J., Wang, J.-J., Gu, G., Bolvin, D., . . . Shin, D.-B. (2018). The Global Precipitation Climatology Project (GPCP) Monthly Analysis (New Version 2.3) and a Review of 2017 Global Precipitation. *Atmosphere, 9*(4), 138. doi:10.3390/atmos9040138

Bacmeister, J. T., Hannay, C., Medeiros, B., Gettelman, A., Neale, R., Fredriksen, H. B., . . . Otto-Bliesner, B. (2020). CO2 Increase Experiments Using the CESM: Relationship to Climate Sensitivity and Comparison of CESM1 to CESM2. *Journal of Advances in Modeling Earth Systems, 12*(11), e2020MS002120. doi:10.1029/2020MS002120

Bitz, C. M., Shell, K. M., Gent, P. R., Bailey, D. A., Danabasoglu, G., Armour, K. C., . . . Kiehl, J. T. (2011). Climate Sensitivity of the Community Climate System Model, Version 4. *Journal of Climate, 25*(9), 3053-3070. doi:10.1175/JCLI-D-11-00290.1

Bjordal, J., Storelvmo, T., Alterskjær, K., & Carlsen, T. (2020). Equilibrium climate sensitivity above 5 °C plausible due to state-dependent cloud feedback. *Nature Geoscience, 13*(11), 718-721. doi:10.1038/s41561-020-00649-1

Bogenschutz, P. A., Gettelman, A., Morrison, H., Larson, V. E., Craig, C., & Schanen, D. P. (2013). Higher-Order Turbulence Closure and Its Impact on Climate Simulations in the Community Atmosphere Model. *Journal of Climate, 26*(23), 9655-9676. doi:10.1175/JCLI-D-13-00075.1

Brady, E. C., Otto-Bliesner, B. L., Kay, J. E., & Rosenbloom, N. (2013). Sensitivity to Glacial Forcing in the CCSM4. *Journal of Climate, 26*(6), 1901–1925. doi:10.1175/JCLI-D-11-00416.1

Ceppi, P., & Nowack, P. (2021). Observational evidence that cloud feedback amplifies global warming. *Proceedings of the National Academy of Sciences, 118*(30), e2026290118. doi:10.1073/pnas.2026290118

Cesana, G. V., & Del Genio, A. D. (2021). Observational constraint on cloud feedbacks suggests moderate climate sensitivity. *Nature Climate Change, 11*(3), 213-218. doi:10.1038/s41558-020-00970-y

Charney, J. G., Arakawa, A., Baker, D. J., Bolin, B., Dickinson, R. E., Goody, R. M., . . . Wunsch, C. I. (1979). *Carbon dioxide and climate: a scientific assessment*: National Academy of Sciences, Washington, DC.

Chepfer, H., Bony, S., Winker, D., Cesana, G., Dufresne, J. L., Minnis, P., . . . Zeng, S. (2010). The GCM-Oriented CALIPSO Cloud Product (CALIPSO-GOCCP). *Journal of Geophysical Research: Atmospheres, 115*(D4). doi:10.1029/2009JD012251

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., . . . Strand, W. G. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems, 12*(2), e2019MS001916. doi:10.1029/2019ms001916

DeMott, P. J., Prenni, A. J., Liu, X., Kreidenweis, S. M., Petters, M. D., Twohy, C. H., . . . Rogers, D. C. (2010). Predicting global atmospheric ice nuclei distributions and their impacts on climate. *Proceedings of the National Academy of Sciences, 107*(25), 11217-11222. doi:10.1073/pnas.0910818107

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev., 9*(5), 1937-1958. doi:10.5194/gmd-9-1937-2016

Feng, R., Otto-Bliesner, B. L., Brady, E. C., & Rosenbloom, N. (2020). Increased Climate Response and Earth System Sensitivity From CCSM4 to CESM2 in Mid-Pliocene Simulations. *Journal of Advances in Modeling Earth Systems, 12*(8), e2019MS002033. doi:10.1029/2019ms002033

Frey, W. R., & Kay, J. E. (2018). The influence of extratropical cloud phase and amount feedbacks on climate sensitivity. *Climate Dynamics, 50*(7), 3097-3116. doi:10.1007/s00382-017-3796-5

Gettelman, A., Bardeen, C. G., McCluskey, C. S., Järvinen, E., Stith, J., Bretherton, C., . . . Wu, W. (2020). Simulating Observations of Southern Ocean Clouds and Implications for Climate. *Journal of Geophysical Research: Atmospheres, 125*(21), e2020JD032619. doi:10.1029/2020JD032619

Gettelman, A., Gagne, D. J., Chen, C.-C., Christensen, M. W., Lebo, Z. J., Morrison, H., & Gantos, G. (2021). Machine Learning the Warm Rain Process. *Journal of Advances in Modeling Earth Systems, 13*(2), e2020MS002268. doi:10.1029/2020MS002268

Gettelman, A., Hannay, C., Bacmeister, J. T., Neale, R. B., Pendergrass, A. G., Danabasoglu, G., . . . Mills, M. J. (2019). High Climate Sensitivity in the Community Earth System Model Version 2 (CESM2). *Geophysical Research Letters, 46*, 8329–8337. doi:10.1029/2019GL083978

Gettelman, A., Kay, J. E., & Shell, K. M. (2012). The Evolution of Climate Sensitivity and Climate Feedbacks in the Community Atmosphere Model. *Journal of Climate, 25*(5), 1453-1469. doi:10.1175/JCLI-D-11-00197.1

Gettelman, A., & Morrison, H. (2015). Advanced Two-Moment Bulk Microphysics for Global Models. Part I: Off-Line Tests and Comparison with Other Schemes. *Journal of Climate, 28*(3), 1268-1287. doi:10.1175/JCLI-D-14-00102.1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., . . . Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society, 146*(730), 1999-2049. doi:10.1002/qj.3803

Hoose, C., Kristjánsson, J. E., Chen, J.-P., & Hazra, A. (2010). A Classical-Theory-Based Parameterization of Heterogeneous Ice Nucleation by Mineral Dust, Soot, and Biological Particles in a Global Climate Model. *Journal of the Atmospheric Sciences, 67*(8), 2483-2503. doi:10.1175/2010JAS3425.1

Huang, B. Y., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., . . . Zhang, H. M. (2017). Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. *Journal of Climate, 30*(20), 8179-8205. doi:10.1175/jcli-d-16-0836.1

IPCC. (2013). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (T. F. Stocker, D. Qin, G.-K. Plattner, M. M. B. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, & P. M. Midgley Eds.). Cambridge, United Kingdom and New York, NY, USA.: Cambridge University Press.

IPCC. (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate*

Change (V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou Eds.). Cambridge, United Kingdom and New York, NY, USA.: Cambridge University Press.

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., & Morice, C. P. (2012). Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research: Atmospheres, 117*(D5). doi:10.1029/2011JD017139

Kay, J. E., Bourdages, L., Miller, N. B., Morrison, A., Yettella, V., Chepfer, H., & Eaton, B. (2016). Evaluating and improving cloud phase in the Community Atmosphere Model version 5 using spaceborne lidar observations. *Journal of Geophysical Research: Atmospheres, 121*(8), 4162-4176. doi:10.1002/2015JD024699

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., . . . Vertenstein, M. (2015). THE COMMUNITY EARTH SYSTEM MODEL (CESM) LARGE ENSEMBLE PROJECT : A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bulletin of the American Meteorological Society, 96*(8), 1333-1350. doi:10.1175/BAMS-D-13-00255.1

Kay, J. E., Hillman, B. R., Klein, S. A., Zhang, Y., Medeiros, B., Pincus, R., . . . Ackerman, T. P. (2012). Exposing Global Cloud Biases in the Community Atmosphere Model (CAM) Using Satellite Observations and Their Corresponding Instrument Simulators. *Journal of Climate, 25*(15), 5190-5207. doi:10.1175/JCLI-D-11-00469.1

Kiehl, J. T. (2007). Twentieth century climate model response and climate sensitivity. *Geophysical Research Letters, 34*(22). doi:10.1029/2007GL031383

Kiehl, J. T., Shields, C. A., Hack, J. J., & Collins, W. D. (2006). The Climate Sensitivity of the Community Climate System Model Version 3 (CCSM3). *Journal of Climate, 19*(11), 2584-2596. doi:10.1175/JCLI3747.1

Larson, V. E., & Golaz, J.-C. (2005). Using Probability Density Functions to Derive Consistent Closure Relationships among Higher-Order Moments. *Monthly Weather Review, 133*(4), 1023-1042. doi:10.1175/MWR2902.1

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., . . . Zeng, X. (2019). The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty. *Journal of Advances in Modeling Earth Systems, 11*(12), 4245-4287. doi:10.1029/2018ms001583

Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., & Zyss, D. (2019). Improvements in the GISTEMP Uncertainty Model. *Journal of Geophysical Research: Atmospheres, 124*(12), 6307-6326. doi:10.1029/2018JD029522

Loeb, N. G., Doelling, D. R., Wang, H. L., Su, W. Y., Nguyen, C., Corbett, J. G., . . . Kato, S. (2018). Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Top-of-Atmosphere (TOA) Edition-4.0 Data Product. *Journal of Climate, 31*(2), 895-918. doi:10.1175/jcli-d-17-0208.1

Manabe, S., & Broccoli, A. J. (1985). A Comparison of Climate Model Sensitivity with Data from the Last Glacial Maximum. *Journal of the Atmospheric Sciences, 42*(23), 2643-2651. doi:10.1175/1520-0469(1985)042<2643:ACOCMS>2.0.CO;2

Marchand, R., Ackerman, T., Smyth, M., & Rossow, W. B. (2010). A review of cloud top height and optical depth histograms from MISR, ISCCP, and MODIS. *Journal of Geophysical Research: Atmospheres, 115*(D16). doi:10.1029/2009JD013422

752 Meehl, G. A., Arblaster, J. M., Bates, S., Richter, J. H., Tebaldi, C., Gettelman, A., . . . Strand, G.
753 (2020). Characteristics of Future Warmer Base States in CESM2. *Earth and Space*
754 *Science, 7*(9), e2020EA001296. doi:10.1029/2020ea001296
755 Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J.-F., Stouffer, R. J., . . . Schlund,
756 M. (2020). Context for interpreting equilibrium climate sensitivity and transient climate
757 response from the CMIP6 Earth system models. *Science Advances, 6*(26), eaba1981.
758 doi:10.1126/sciadv.aba1981
759 Mülmenstädt, J., Salzmann, M., Kay, J. E., Zelinka, M. D., Ma, P.-L., Nam, C., . . . Quaas, J.
760 (2021). An underestimated negative cloud feedback from cloud lifetime changes. *Nature*
761 *Climate Change, 11*(6), 508-513. doi:10.1038/s41558-021-01038-1
762 Myers, T. A., Scott, R. C., Zelinka, M. D., Klein, S. A., Norris, J. R., & Caldwell, P. M. (2021).
763 Observational constraints on low cloud feedback reduce uncertainty of climate sensitivity.
764 *Nature Climate Change*. doi:10.1038/s41558-021-01039-0
765 Nam, C., Bony, S., Dufresne, J.-L., & Chepfer, H. (2012). The 'too few, too bright' tropical low-
766 cloud problem in CMIP5 models. *Geophysical Research Letters, 39*(21).
767 doi:10.1029/2012GL053421
768 Ohgaito, R., Abe-Ouchi, A., O'Ishi, R., Takemura, T., Ito, A., Hajima, T., . . . Kawamiya, M.
769 (2018). Effect of high dust amount on surface temperature during the Last Glacial
770 Maximum: a modelling study using MIROC-ESM. *Clim. Past, 14*(11), 1565-1581.
771 doi:10.5194/cp-14-1565-2018
772 Otto-Bliesner, B. L., Brady, E. C., Clauzet, G., Tomas, R., Levis, S., & Kothavala, Z. (2006).
773 Last glacial maximum and Holocene climate in CCSM3. *Journal of Climate, 19*(11),
774 2526-2544. doi:10.1175/JCLI3748.1
775 Patnaude, R., Diao, M., Liu, X., & Chu, S. (2021). Effects of thermodynamics, dynamics and
776 aerosols on cirrus clouds based on in situ observations and NCAR CAM6. *Atmos. Chem.*
777 *Phys., 21*(3), 1835-1859. doi:10.5194/acp-21-1835-2021
778 Peltier, W. R., Argus, D. F., & Drummond, R. (2015). Space geodesy constrains ice age terminal
779 deglaciation: The global ICE-6G_C (VM5a) model. *Journal of Geophysical Research:*
780 *Solid Earth, 120*(1), 450-487. doi:10.1002/2014JB011176
781 Pincus, R., Platnick, S., Ackerman, S. A., Hemler, R. S., & Patrick Hofmann, R. J. (2012).
782 Reconciling Simulated and Observed Views of Clouds: MODIS, ISCCP, and the Limits
783 of Instrument Simulators. *Journal of Climate, 25*(13), 4699-4720. doi:10.1175/jcli-d-11-
784 00267.1
785 Santos, S. P., Caldwell, P. M., & Bretherton, C. S. (2020). Numerically Relevant Timescales in
786 the MG2 Microphysics Model. *Journal of Advances in Modeling Earth Systems, 12*(4),
787 e2019MS001972. doi:10.1029/2019MS001972
788 Shaw, J. K., McGraw, Z., Bruno, O., Storelvmo, T., & Hofer, S. (2021). Using satellite
789 observations to evaluate model representation of Arctic mixed-phase clouds. *Earth and*
790 *Space Science Open Archive ESSOAr*.
791 Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., . . .
792 Zelinka, M. D. (2020). An assessment of Earth's climate sensitivity using multiple lines
793 of evidence. *Reviews of Geophysics, n/a*(n/a), e2019RG000678.
794 doi:10.1029/2019RG000678
795 Shin, S. I., Liu, Z., Otto-Bliesner, B., Brady, E., Kutzbach, J., & Harrison, S. (2003). A
796 Simulation of the Last Glacial Maximum climate using the NCAR-CCSM. *Climate*
797 *Dynamics, 20*(2-3), 127–151. doi:10.1007/s00382-002-0260-x

Swales, D. J., Pincus, R., & Bodas-Salcedo, A. (2018). The Cloud Feedback Model Intercomparison Project Observational Simulator Package: Version 2. *Geosci. Model Dev., 11*(1), 77-81. doi:10.5194/gmd-11-77-2018

Tan, I., Storelvmo, T., & Zelinka, M. D. (2016). Observational constraints on mixed-phase clouds imply higher climate sensitivity. *Science, 352*(6282), 224. doi:10.1126/science.aad5300

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres, 106*(D7), 7183-7192. doi:10.1029/2000JD900719

Taylor, K. E., Crucifix, M., Braconnot, P., Hewitt, C. D., Doutriaux, C., Broccoli, A. J., . . . Webb, M. J. (2007). Estimating Shortwave Radiative Forcing and Response in Climate Models. *Journal of Climate, 20*(11), 2530-2543. doi:10.1175/JCLI4143.1

Tierney, J. E., Zhu, J., King, J., Malevich, S. B., Hakim, G. J., & Poulsen, C. J. (2020). Glacial cooling and climate sensitivity revisited. *Nature, 584*(7822), 569-573. doi:10.1038/s41586-020-2617-x

Wang, C., Soden, B. J., Yang, W., & Vecchi, G. A. (2021). Compensation Between Cloud Feedback and Aerosol-Cloud Interaction in CMIP6 Models. *Geophysical Research Letters, 48*(4), e2020GL091024. doi:10.1029/2020GL091024

Wang, Y., Liu, X., Hoose, C., & Wang, B. (2014). Different contact angle distributions for heterogeneous ice nucleation in the Community Atmospheric Model version 5. *Atmos. Chem. Phys., 14*(19), 10411-10430. doi:10.5194/acp-14-10411-2014

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., . . . Taylor, K. E. (2020). Causes of Higher Climate Sensitivity in CMIP6 Models. *Geophysical Research Letters, 47*(1), e2019GL085782. doi:10.1029/2019GL085782

Zhu, J., Liu, Z., Brady, E., Otto-Bliesner, B., Zhang, J., Noone, D., . . . Tabor, C. (2017). Reduced ENSO variability at the LGM revealed by an isotope-enabled Earth system model. *Geophysical Research Letters, 44*(13), 6984-6992. doi:10.1002/2017GL073406

Zhu, J., Otto-Bliesner, B. L., Brady, E. C., Poulsen, C. J., Tierney, J. E., Lofverstrom, M., & DiNezio, P. (2021). Assessment of Equilibrium Climate Sensitivity of the Community Earth System Model Version 2 Through Simulation of the Last Glacial Maximum. *Geophysical Research Letters, 48*(3), e2020GL091220. doi:10.1029/2020GL091220

Zhu, J., & Poulsen, C. J. (2020). On the Increase of Climate Sensitivity and Cloud Feedback With Warming in the Community Atmosphere Models. *Geophysical Research Letters, 47*(18), e2020GL089143. doi:10.1029/2020GL089143

Zhu, J., & Poulsen, C. J. (2021). Last Glacial Maximum (LGM) climate forcing and ocean dynamical feedback and their implications for estimating climate sensitivity. *Clim. Past, 17*(1), 253-267. doi:10.5194/cp-17-253-2021

Zhu, J., Poulsen, C. J., & Otto-Bliesner, B. L. (2020). High climate sensitivity in CMIP6 model not supported by paleoclimate. *Nature Climate Change, 10*(5), 378-379. doi:10.1038/s41558-020-0764-6

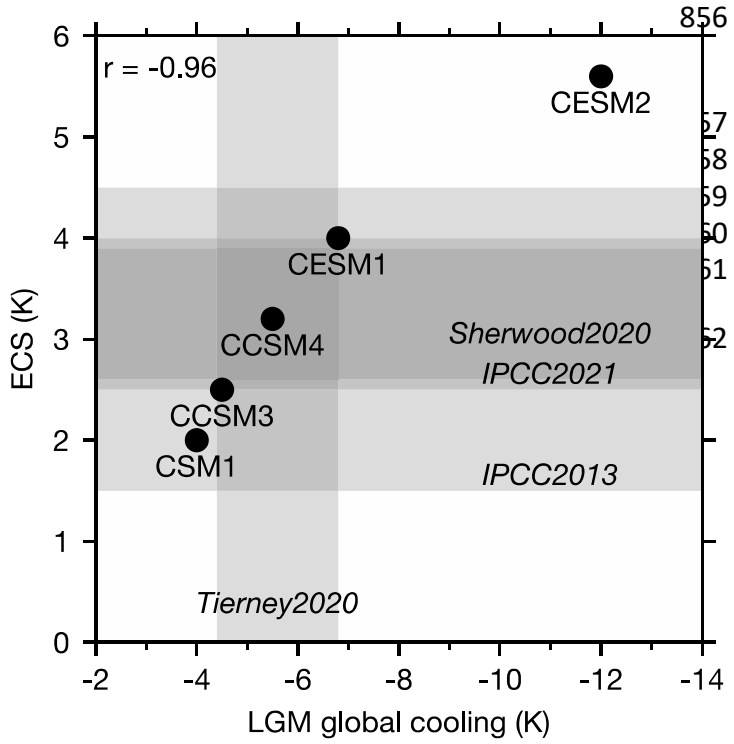
Zhu, J., Poulsen, C. J., & Tierney, J. E. (2019). Simulation of Eocene extreme warmth and high climate sensitivity through cloud feedbacks. *Science Advances, 5*(9), eaax1874. doi:10.1126/sciadv.aax1874

842

843 **Table 1.** List of top-of-atmosphere net radiation (ΔN; units: W m^{-2}) in the coupled preindustrial
844 (PI) and Last Glacial Maximum (LGM) simulations, the derived global mean surface
845 temperature change (ΔT_{LGM}; units: °C) and shortwave cloud feedback ($\lambda_{SW\_CLD\_LGM}$; units: W m^{-2} K^{-1}) in the LGM simulations, and equilibrium climate sensitivity (ECS_{SOM}; units: °C) and
847 shortwave cloud feedback ($\lambda_{SW\_CLD\_2\times}$; units: W m^{-2} K^{-1}) derived in a pair of slab ocean
848 simulations with the PI condition and an abrupt 2×CO_2 forcing. Results from the last 30 years of
849 simulations with various configuration of the atmosphere model are shown. *CAM6* uses the
850 default CESM2(CAM6); *CAM5* uses the old CAM5 cloud parameterizations; *ClubbOff* uses the
851 CAM5 shallow convection and boundary layer schemes; *HetFrzOff* uses the CAM5 ice
852 nucleation scheme; *Mg2Off* uses the CAM5 cloud microphysics; *NoNimax* removes the "nimax"
853 limiter; *Mg2Sub8* uses 8 substeps in the microphysical scheme; *NnSub8* removes the "nimax"
854 limiter and uses 8 substeps in the microphysics. See text for details of these configurations.

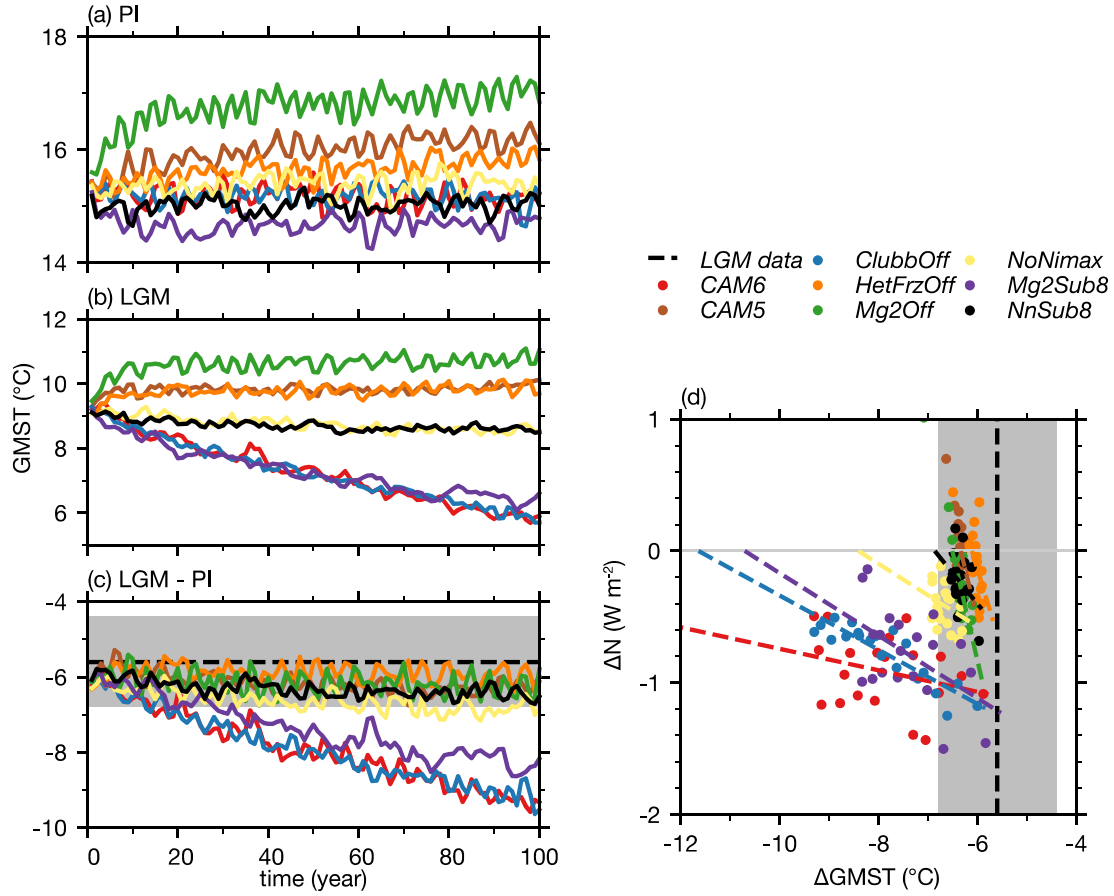
Configurations	PI ΔN	LGM ΔN	ΔT_{LGM}	$\lambda_{sw\_cld\_LGM}$	ECS_{SOM}	$\lambda_{sw\_cld\_2\times}$
CAM6	–0.18	–1.01	–9.0	0.81	6.1	0.95
CAM5	0.30	0.27	–6.3	0.29	3.7	0.32
HetFrzOff	0.42	0.15	–5.9	0.37	3.8	0.47
ClubbOff	–0.48	–1.1	–8.9	0.64	6.2	0.86
Mg2Off	0.41	0.01	–6.3	0.49	4.3	0.54
NoNimax	0.13	–0.29	–6.9	0.64	5.0	0.79
Mg2Sub8	–0.21	–0.81	–8.2	0.72	4.8	0.74
NnSub8	0.09	–0.14	–6.4	0.49	4.0	0.59

855



863 **Figure 1.** Model simulated LGM global cooling and ECS in different versions of CESM. ECS is
864 estimated using coupled simulation with a slab ocean (see text for references). Vertical patch
865 indicates the 95% confidence interval of proxy estimation of the LGM global cooling from
866 Tierney et al. (2020). Horizontal patches denote the 66% confidence intervals of ECS from the
867 IPCC Assessment Report 5 (IPCC, 2013; light gray), Assessment Report 6 (IPCC, 2021; medium
868 gray) and Sherwood et al. (2020; dark gray).

869



870

871 **Figure 2.** Time series of global mean surface temperature (GMST) in (**a**) the preindustrial (PI)
872 and (**b**) the Last Glacial Maximum (LGM) simulations using various atmosphere model
873 configurations within the coupled CESM2 framework. (**c**) Changes in GMST between paired
874 LGM and PI simulations. (**d**) Changes in top-of-atmosphere radiation (ΔN) versus GMST
875 (ΔGMST) in paired simulations. ΔN and ΔGMST (markers) are the 5-year running mean of the
876 LGM time series with the last-30-year averages of the PI simulation subtracted. A linear
877 regression between ΔN and ΔGMST is shown as dashed line for each configuration. In (**c**) and
878 (**d**), the LGM ΔGMST and the 95% uncertainty interval from Tierney et al. (2020) are shown.
879 See text and Table 1 for details of the model configurations.

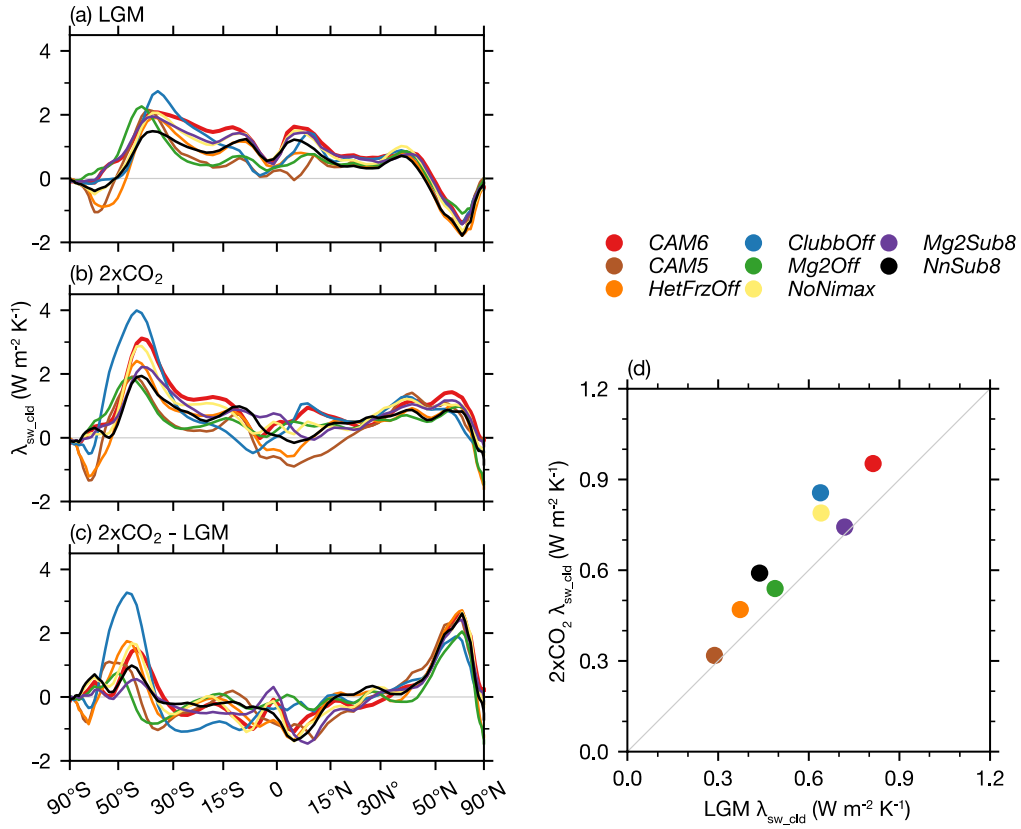


Figure 3. Zonal mean shortwave cloud feedback (λ_{sw_cld}; units: W m^{-2} K^{-1}) for various atmosphere model configurations in (**a**) the paired preindustrial and Last Glacial Maximum (LGM) simulations using fully coupled CESM2 and (**b**) the paired preindustrial and 2×CO_2 simulations using CESM2 slab ocean model. (**c**) Differences in λ_{sw_cld} between the LGM and 2×CO_2 simulations using the same atmosphere model configuration. (**d**) Scatter plot of the global mean λ_{sw_cld} in the LGM and 2×CO_2 simulations. See text and Table 1 for details of the model configurations.

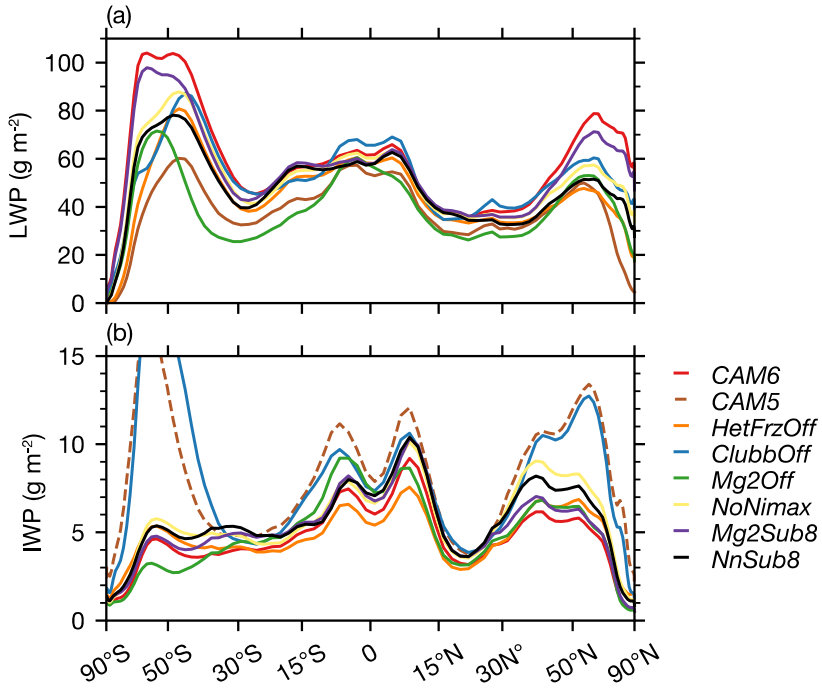


Figure 4. (**a**) Zonal mean cloud liquid water path (LWP) in the preindustrial simulations with various atmosphere model configurations in the coupled CESM2 framework. (**b**) as (**a**) but for the cloud ice water path (IWP). Note that *CAM5* IWP in (**b**) has been multiplied by 0.5 for illustrative purpose (shown as dashed brown line). See text and Table 1 for details of the model configurations.

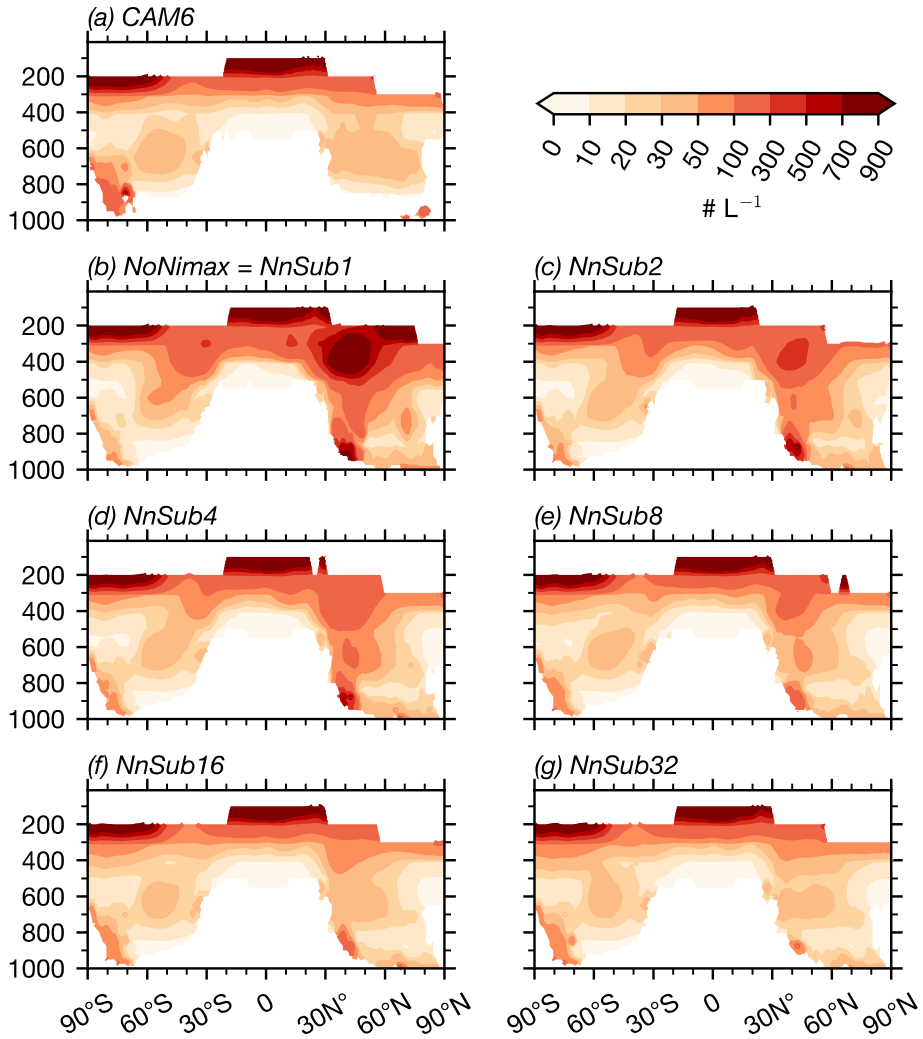


Figure 5. Pressure–latitude section of the zonal mean in-cloud cloud ice number in (**a**) the default *CAM6* simulation and (**b**)–(**g**) simulations with the "nimax" limiter removed and with substep of 1, 2, 4, 8, 16, and 32 in the microphysical scheme, respectively. Pressure is the y-axis with units of hPa. Results are from atmosphere-only simulations forced by the observed present-day sea surface temperature and sea ice (results are similar if coupled simulations are used). The in-cloud ice number is constructed by averaging monthly cloud ice numbers for grid points with a cloud ice mixing ratio greater than 0.01 part-per-million by mass.

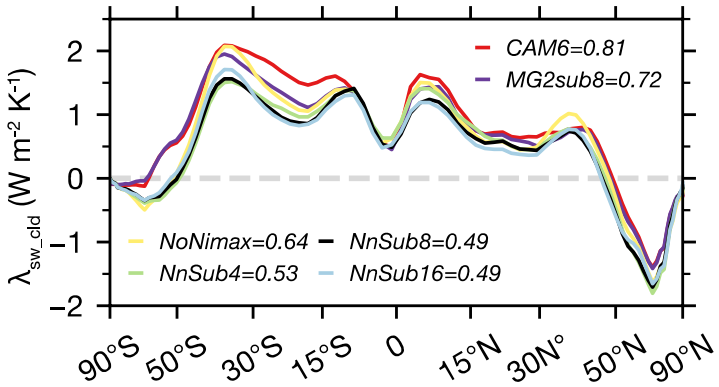
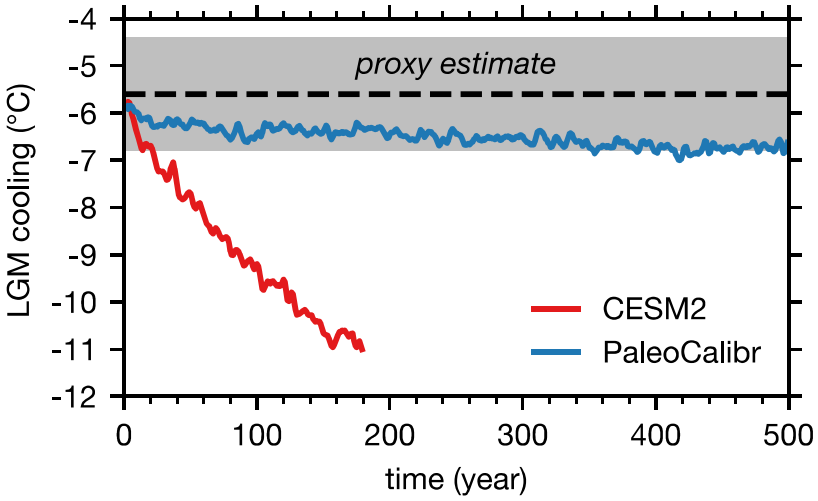


Figure 6. Zonal mean shortwave cloud feedback (λ_{sw_cld}) calculated in the paired preindustrial and Last Glacial Maximum simulations with various configurations. See text and Table 1 for details of the model configurations. *NnSub4* and *NnSub16* are the same as *NnSub8* except for the 4 and 16 substeps in the microphysics, respectively.



910

911 **Figure 7.** Time series of the global mean surface temperature anomaly in the LGM simulations
912 with CESM2 and the paleoclimate-calibrated (PaleoCalibr) configurations. Black dashed line
913 with the gray patch denotes the 95% uncertainty interval from Tierney et al. (2020) for the LGM
914 global cooling.

915

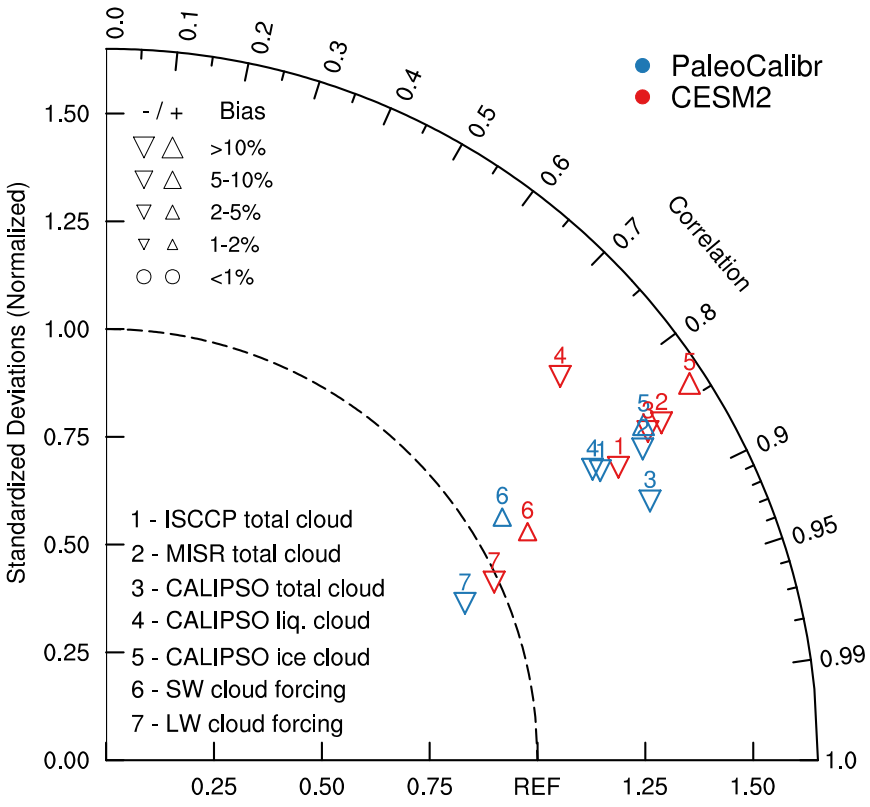
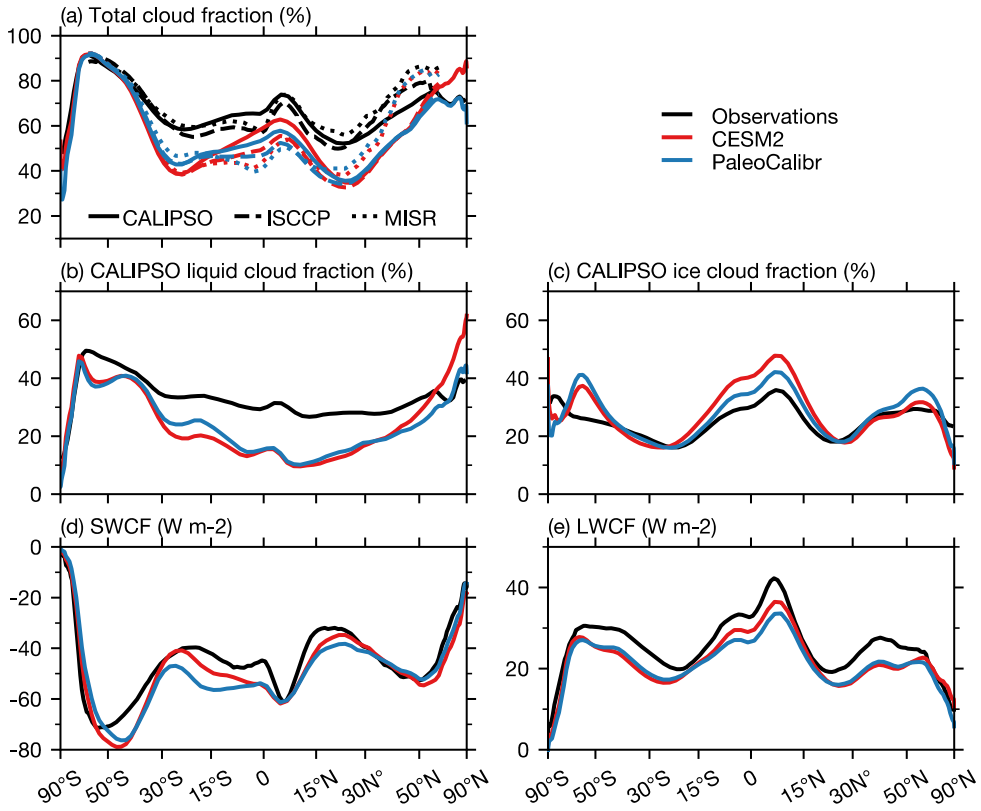


Figure 8. Taylor diagram evaluating key cloud variables in CESM2 and PaleoCalibr with satellite observations. Model variables are from AMIP historical simulations with an active Cloud Feedback Model Intercomparison Project Observational Simulator Package. Cloud observations are the total cloud fraction from ISCCP (60°S–60°N), MISR (60°S–60°N and ocean-only), and CALIPSO, the cloud phase partition between liquid and ice from CALIPSO, and the shortwave and longwave cloud radiative forcing (SWCF and LWCF) from CERES-EBAF. Averages between 2000 and 2014 are used for the model-data comparison, except for the CALIPSO cloud fraction that is averaged between 2008 and 2020.



926

927 **Figure 9.** Comparison of model simulations against (**a**) the zonal mean total cloud fraction in
928 CALIPSO, ISCCP, and MISR, (**b**) the CALIPSO liquid cloud fraction, (**c**) the CALIPSO ice
929 cloud fraction, (**d**) the CERES-EBAF shortwave cloud forcing (SWCF), and (**e**) the CERES-
930 EBAF longwave cloud forcing (LWCF). Model variables are from AMIP historical simulations
931 with an active Cloud Feedback Model Intercomparison Project Observational Simulator Package.
932 Averages between 2000 and 2014 are used for the model-data comparison, except for the
933 CALIPSO cloud fraction and phase partition that are averaged between 2008 and 2020. ISCCP
934 and MISR cloud fraction is plot between 60°S and 60°N. MISR cloud fraction values are over
935 ocean only.

936

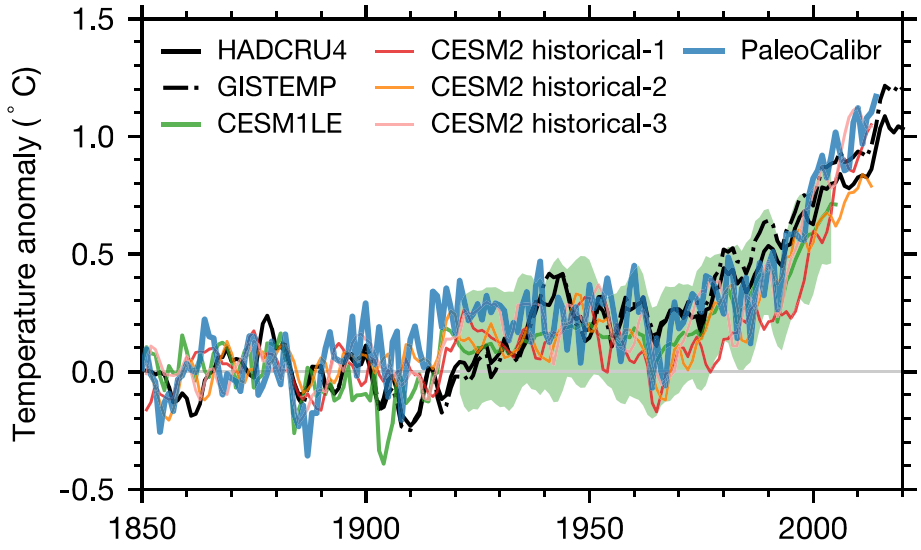


Figure 10. Time series of the global mean surface temperature anomaly during the historical period from observations (black), the CESM2 historical simulations using CAM6 (three members in red, orange, and pink), and the paleoclimate-calibrated configuration (PaleoCalibr; blue). Results from the CESM1 Large Ensemble (CESM1LE; green) are shown as a reference for a possible range of internal variability. The observations are from the HADCRU4 and GISTEMP. Temperature anomalies are calculated from the respective 1850–1900 averages.

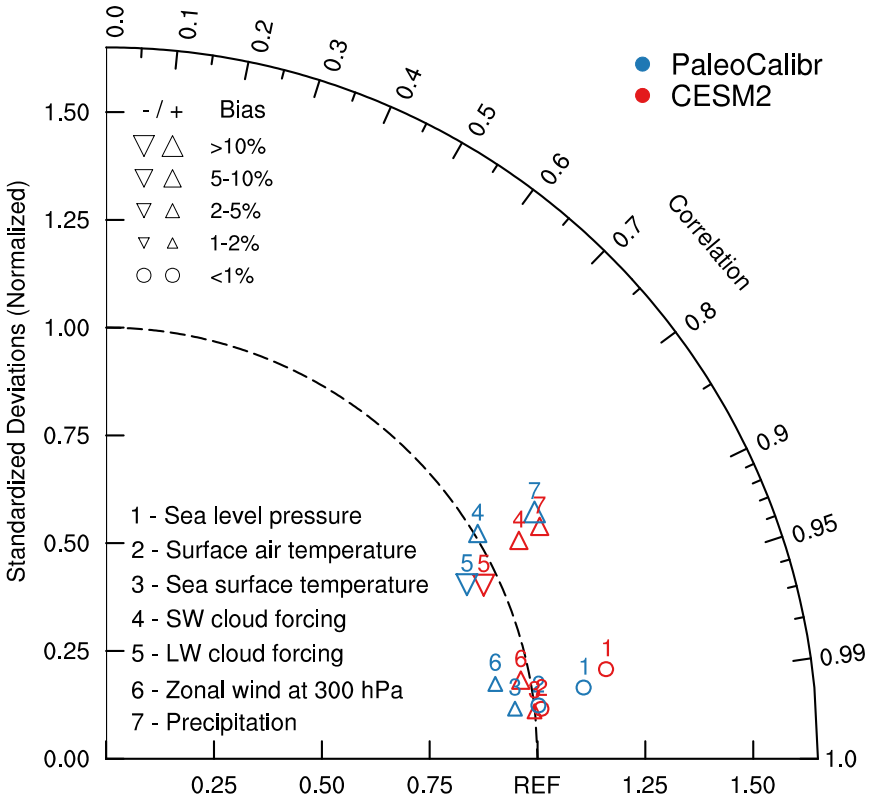


Figure 11. A Taylor diagram evaluating key climate variables in CESM2 and PaleoCalibr coupled historical simulations using observational and reanalysis datasets. Model simulated sea level pressure, surface air temperature, and zonal wind at 300 hPa are compared with averages between 1979 and 2014 from ERA5, shortwave and longwave cloud radiative forcing compared with averages between 2000 and 2014 from CERES-EBAF, precipitation compared with averages between 1979 and 2014 from GPCP, and sea surface temperature compared with averages between 1979 and 2014 from ERSST.

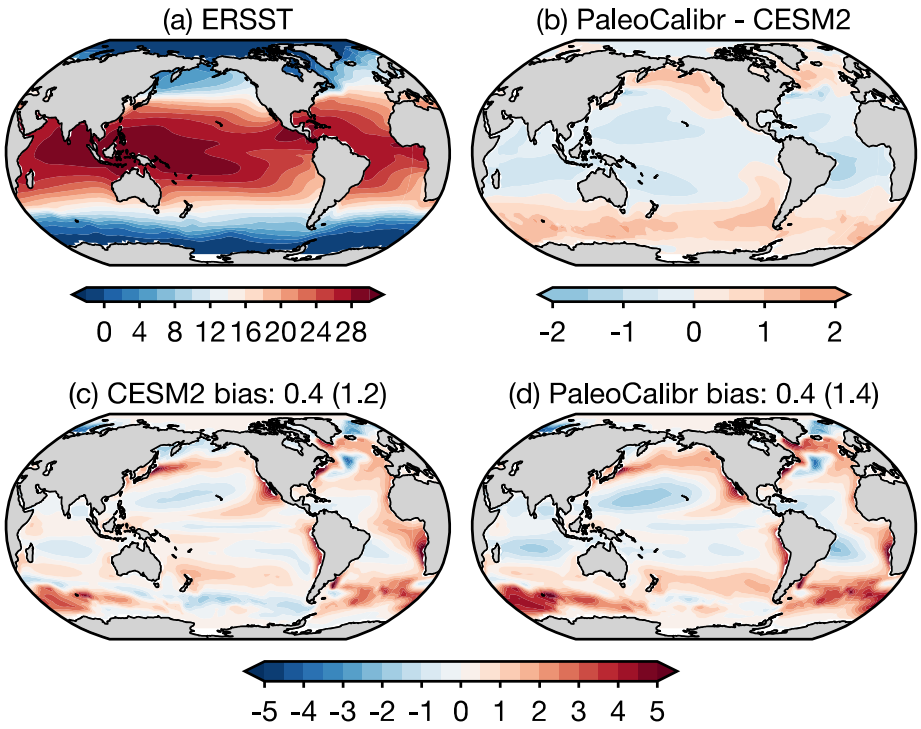
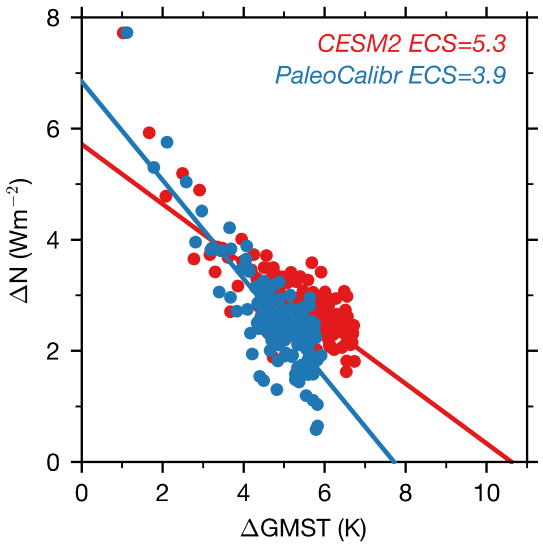


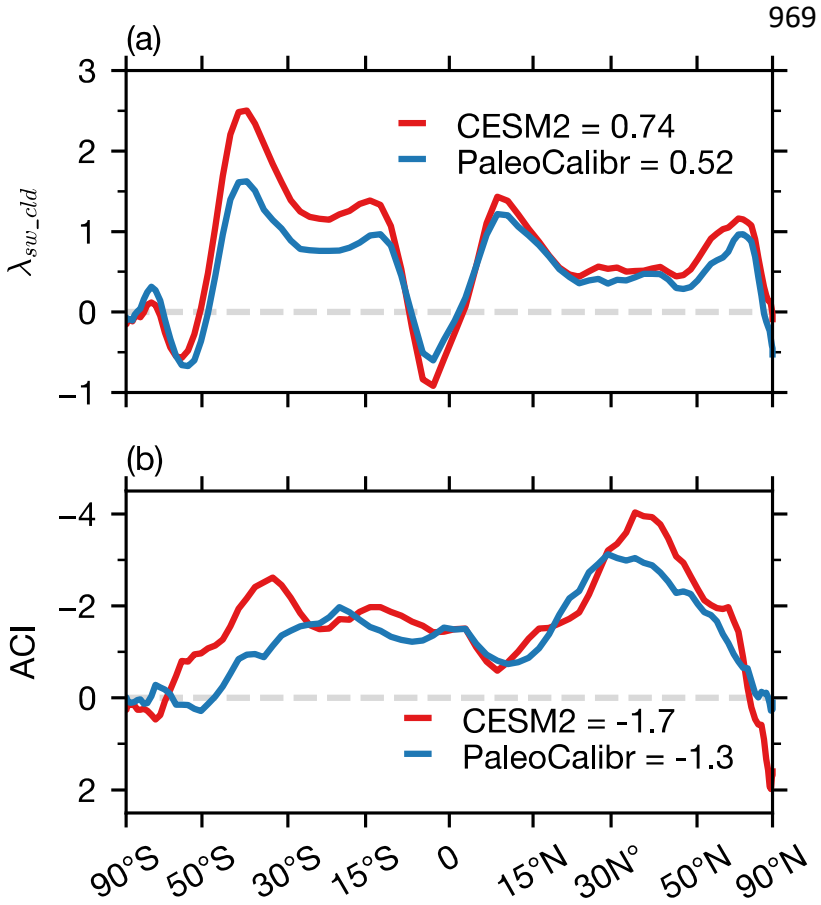
Figure 12. (**a**) Sea surface temperature (SST; units: °C) averaged between 1979 and 2014 CE in observations. (**b**) Difference in SST between the historical simulations using PaleoCalibr and CESM2. (**c**) Difference in SST between the historical simulation using CESM2 and observations. (**d**) as (**c**) but in the historical simulation using *PaleoCalibr*. Numbers in (**c**) and (**d**) are the global mean model bias and the area-weighted root-mean-squared error (in parentheses).



963

964 **Figure 13.** Scatter plot of the global mean surface temperature anomaly (ΔGMST) and the top-
965 of-atmosphere net radiation (ΔN) in the abrupt 4×CO_2 simulations using CESM2 and
966 PaleoCalibr. 4×CO_2 simulations are run for 150 model years. Equilibrium Climate Sensitivity
967 (ECS) is estimated using the regression method, i.e., a half of the x-axis intercept.

968



970

971 **Figure 14.** (**a**) Zonal mean shortwave cloud feedback (λ_{sw_cld}; units: W m^{-2} K^{-1}) calculated in the
972 abrupt 4×CO_2 simulations using CESM2 and PaleoCalibr. (**b**) as (**a**) but for the aerosol cloud
973 interaction (ACI; units: W m^{-2}). The global mean λ_{sw_cld} and ACI are shown with figure legend.
974 Note that the y-axis is revered in (**b**).