

Multiple Equilibria in a Coupled Climate-Carbon Model

Fangze Zhu^a and Brian E. J. Rose^a

^a *Department of Atmospheric and Environmental Sciences, University at Albany, State University
of New York, Albany, New York*

Corresponding author: Fangze Zhu, fzhu@albany.edu

6 ABSTRACT: Multiple stable equilibria are intrinsic to many complex dynamical systems, and
7 have been identified in a hierarchy of climate models. Motivated by the idea that the Quaternary
8 glacial-interglacial cycles could have resulted from orbitally-forced transitions between multiple
9 stable states mediated by internal feedbacks, this study investigates the existence and mechanisms of
10 multiple equilibria in an idealized, energy-conserving atmosphere-ocean-sea ice general circulation
11 model with a fully coupled carbon cycle. Four stable climates are found for identical insolation
12 and global carbon inventory: an ice-free Warm climate, two intermediate climates (Cold and
13 Waterbelt), and a fully ice-covered Snowball climate. A fifth, small ice cap state between Warm
14 and Cold is found to be barely unstable. Using custom radiative kernels and a thorough sampling
15 of the model's internal variability, three equilibria are investigated through the state dependence of
16 radiative feedback processes. For fast feedbacks, the systematic decrease in surface albedo feedback
17 from Cold to Warm states is offset by a similar increase in longwave water vapor feedback. At
18 longer timescales, the key role of the carbon cycle is a dramatic lengthening of the adjustment time
19 comparable to orbital forcings near the Warm state. The dynamics of the coupled climate-carbon
20 system are thus not well separated in time from orbital forcings, raising interesting possibilities for
21 nonlinear triggers for large climate changes.

22 SIGNIFICANCE STATEMENT: How do carbon cycle and other physical processes affect the
23 physical and mathematical properties of the climate system? We use a complex climate model
24 coupled with a carbon cycle to simulate the climate evolution under different initial conditions.
25 Four stable climate states are possible, from the Snowball Earth, in which ice covers the whole
26 planet, to the Warm state, an ice-free world. The carbon cycle drives the global climate change
27 at an extremely slower pace after sea ice retreats. Sea ice and water vapor, on the other hand,
28 constitute the major contributing factors that accelerate faster climate change.

29 **1. Introduction**

30 The climate system is extremely complex, which suggests that the climate response must not
31 be linear to external perturbations. In a recent study, Westerhold et al. (2020) identified four
32 climate states including Hothouse, Warmhouse, Coolhouse and Icehouse in a new, astronomically
33 dated record over the last 66 million years. They found that different climate states associated
34 with different atmospheric CO₂ concentrations and polar ice volumes were closely related to
35 latitude-specific feedback processes. Specifically, warmer climates were dominated by low-latitude
36 feedbacks, while cooler climates involved more complicated high-latitude feedbacks.

37 To understand the past climate changes, such as the glacial-interglacial cycles, the key question
38 is how internal feedbacks could amplify or suppress the climate responses to external forcings
39 (Paillard 1998; Lisiecki 2010; Rial et al. 2013; Abe-Ouchi et al. 2013; Willeit et al. 2019). Important
40 amplifiers may be found among many ocean, atmosphere, and land surface processes. The global
41 carbon cycle in particular is considered to be the most promising candidate (Zachos et al. 2001;
42 Lisiecki 2010; Paillard 2015) due to its long timescales (PALAEOSENS Project Members 2012).
43 However, the lack of consensus on the mechanisms of the CO₂ fluctuations over the past 800,000
44 years makes it a challenge to interpret glacial cycles in terms of CO₂ variations (Sigman and Boyle
45 2000; Sigman et al. 2010; Paillard 2015; Marzocchi and Jansen 2019; Stein et al. 2020).

46 Even the simplest climate model exhibits one of the intrinsic properties in a nonlinear system,
47 multiple equilibria (North 1990). The Budyko-Sellers zero-dimensional energy balance model
48 (EBM) (Budyko 1969; Sellers 1969) depicts two distinct worlds: a warm, habitable world with
49 little amount of ice on the poles and a frozen, “wicked” world where the entire globe is covered by

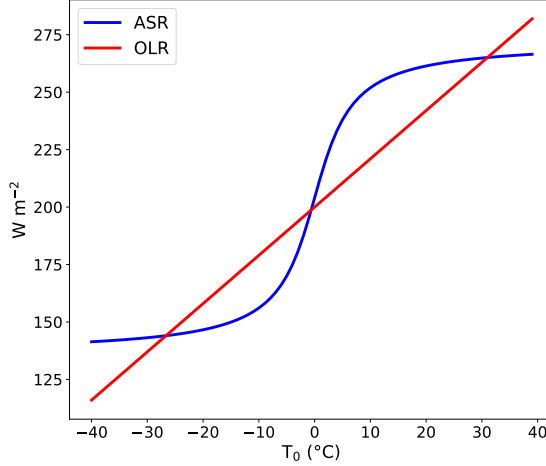


FIG. 1. The relationship of the top of the atmosphere (TOA) radiative fluxes (blue: ASR; red: OLR) versus global mean temperature (T_0) in the framework of the Budyko-Sellers zero-dimensional energy balance model (EBM) in the style of North (1990). The parameterization of the planetary albedo originates from Arnscheidt and Rothman (2020). ASR and OLR are plotted as the first and second terms on the right hand side of Eq. (1), respectively, where $S_0 = 1360 \text{ W m}^{-2}$, $A = 200 \text{ W m}^{-2}$, and $B = 2.1 \text{ W m}^{-2} \text{ } ^\circ \text{C}^{-1}$ following North (1990).

ice, i.e., Snowball Earth. It can be formulated as

$$C \frac{dT_0}{dt} = \frac{S_0}{4} [1 - \alpha(T_0)] - (A + BT_0) \quad (1)$$

where C is the heat capacity of the Earth system, S_0 is the solar constant, α is the planetary albedo, T_0 is the global mean temperature, A and B are constants. The first term on the right hand side of Eq. (1) is the absorbed shortwave radiation (ASR), whereas the second term is a simple linear parameterization of the outgoing longwave radiation (OLR). The term on the left hand side is the gain or loss of energy in the Earth system as a result of radiative imbalance.

Fig. 1 displays the relationship between annual mean ASR/OLR and T_0 in a Budydo-Sellers zero-dimensional EBM [Eq. (1)] styled after North (1990) (though we note that no explanation was given in that paper for the detailed shape of the albedo curve $\alpha(T_0)$). We parameterize $\alpha(T_0)$ as a smoothly varying sigmoid-shaped function of T_0 , which is proportional to ASR in Fig. 1. The specific form of $\alpha(T_0)$ is based on Arnscheidt and Rothman (2020), which is mathematically convenient while capturing the qualitative essence of the surface albedo feedback.

67 Because of the non-constant slope of ASR due to the dependence of surface albedo on T_0 ,
68 there are three intersections between ASR and OLR in Fig. 1 indicating three climatic equilibria:
69 the stable Warm and Snowball states mentioned above, plus an unstable intermediate state near
70 $T_0 = 0^\circ\text{C}$. The stability of the intermediate state was discussed using the potential function in North
71 (1990). Physically, this is because positive feedbacks tend to amplify any small perturbations
72 around this equilibrium point, driving the climate system away from the original state until a
73 stable climate regime is reached. The net feedbacks are related to the slopes of the graphs,
74 respectively $d(\text{ASR})/dT_0$ for the shortwave (SW) feedback and $-d(\text{OLR})/dT_0$ for the longwave
75 (LW) feedback (which we have assumed to be constant in the EBM). Understanding multiple
76 equilibria in the climate system thus requires a thorough investigation into the state dependence of
77 radiative feedback processes.

78 Multiple equilibria also exist in more complex climate models, such as general circulation models
79 (GCMs) (e.g., Popp et al. 2016). Ferreira et al. (2011) reported for the first time that a complex
80 coupled atmosphere-ocean-sea ice model (the Massachusetts Institute of Technology general circu-
81 lation model, or MITgcm, Marshall et al. 1997a,b), in two idealized land configurations, supported
82 three stable climatic equilibria similar to idealized EBM solutions informed by ocean dynamics
83 found in Rose and Marshall (2009). Rose (2015) further discovered that a fourth stable climatic
84 state, the Waterbelt state, in which the ice edge reaches farther equatorward down to the subtropics,
85 also existed in the same coupled climate model with the same idealized configurations. Ferreira
86 et al. (2018) relaxed the constraint of the aquaplanet configuration and used a more Earth-like,
87 although still idealized geometry. They found that three stable states co-existed: a warm state
88 resembling the present-day climate, a cold state similar to the Last Glacial Maximum (LGM) and
89 a snowball. Recently, Brunetti et al. (2019) claimed the existence of up to five climate states in
90 MITgcm in the aquaplanet configuration. The missing fifth state from the previous work was called
91 the “warm state” that lay between the “hot state” and “cold state”. It turns out that their “warm state”
92 corresponds to a barely unstable state in our model (we will refer to this as the “Temperate barely
93 unstable state”) that eventually loses its stability completely, as we will show below. However,
94 there is a common and significant caveat in all these studies: they prescribed the radiatively active
95 atmospheric pCO_2 in their simulations without coupling with the oceanic carbon cycle.

96 Previous studies also explored some mechanisms of multiple equilibria and their transitions in
97 MITgcm. Both Ferreira et al. (2011) and Rose (2015) demonstrated that the ocean heat transport
98 (OHT) convergence played a crucial role in stabilizing the ice edge either in the mid-latitudes or the
99 subtropics. Rose et al. (2013) studied abrupt transitions between warm and cold states, concluding
100 that it was the albedo feedback, rather than the meridional overturning circulation (MOC), that
101 drove the abrupt climate change, but the latter became more active during the warming phase.
102 Ferreira et al. (2018) decomposed the oceanic carbon reservoir and discovered that the air-sea
103 disequilibrium pump played a major role in the glacial atmospheric $p\text{CO}_2$ drawdown, although
104 their model neglected the radiative effects of $p\text{CO}_2$ variations.

105 However, besides the surface albedo feedback, these studies have yet to discuss the relative
106 significance of other feedbacks such as water vapor, lapse rate and cloud feedbacks (i.e. other
107 sources of curvature in the $\text{ASR}(T_0)$ and $\text{OLR}(T_0)$ relationships). In addition, although it was
108 not emphasized by earlier authors, these models did not conserve energy due to the absence of
109 frictional heating. The lack of a closed top of the atmosphere (TOA) energy budget at equilibrium
110 in those earlier simulations prevented a thorough study of the relevant radiative feedback processes
111 (Appendix A). All these limitations motivate us to pursue the following questions: 1) Do multiple
112 equilibria exist in the idealized, energy-conserving MITgcm with a fully-coupled carbon cycle? 2)
113 How does the state dependence of the feedback processes give rise to multiple equilibria? 3)
114 How does the carbon cycle affect the mean state, variability, timescales, and feedbacks in the fully
115 coupled climate system?

116 In the following, we introduce the model configurations and experimental setup in section 2
117 and Appendix A. In section 3, we describe multiple equilibria of the MITgcm using timeseries,
118 climatologies and a bifurcation diagram. In section 4, we introduce the radiative kernels developed
119 specifically for this model based on Soden et al. (2008), and then provide an in-depth discussion
120 on the decomposition of feedback parameters for non-Snowball climate states to quantify the
121 underlying mechanisms. Net feedbacks and the associated climate dynamics for the Warm climates
122 are addressed, and carbon feedbacks near the Warm equilibrium are evaluated. Finally, we have
123 some discussions on possible scenarios of the climate response to orbital forcings.

2. Model and experiments

a. Model descriptions

We use the MITgcm in idealized *Ridge* configuration with a single global ocean basin enclosed by narrow continent extending from pole to pole. The *Ridge* is the simplest possible configuration allowing for wind-driven gyres and planetary-scale meridional overturning circulations in the ocean (Enderton and Marshall 2009), thus capturing some key features of Earth’s climate system without the added complexities of land surface processes. We choose the *Ridge* for explicit comparison to earlier studies of multiple equilibria in this configuration (Ferreira et al. 2011; Rose et al. 2013; Rose 2015), so that we may more clearly see the incremental effects of carbon feedbacks.

The model includes three-dimensional dynamical atmosphere and ocean, and thermodynamic sea ice. The atmosphere includes idealized but physically grounded moist physics and a hydrological cycle. Simplified present-day orbital parameters are prescribed (23.5° obliquity and zero eccentricity), including a seasonal cycle but no diurnal cycle. A highlight of this study is the introduction of a fully coupled carbon cycle (including radiative feedback) into the atmosphere-ocean-sea ice system. The relatively low computational cost of this model relative to more comprehensive Earth-system models enables the very long simulation times needed to study the slow equilibration of oceanic carbon cycle. Details of the model configurations are given in Appendix A.

As mentioned, earlier studies using the coupled MITgcm (Ferreira et al. 2011; Rose et al. 2013; Rose 2015; Ferreira et al. 2018) failed to take into account the frictional heating of the atmosphere, resulting in 2 to 3 W m^{-2} imbalance of the TOA energy flux at equilibrium (Fig. A1). This issue has been resolved in this work, which allows for meaningful feedback analysis through diagnostic decomposition of the TOA energy budget.

b. Equilibrium experiments

The aim of our “equilibrium” experiments is simply to verify if multiple equilibria exist in the model. We take Warm, Cold, and Waterbelt initial conditions from Rose (2015) and adjust initial atmospheric pCO_2 to ensure uniform total carbon inventory. This adjustment is done before starting the experiments, ensuring that all experiments have identical total carbon (Appendix C, bottom panel). The parameter *SOLC* (one quarter of the solar constant) is set to 338 W m^{-2} for

consistency with Rose (2015). Each experiment is then run for many thousands of years out to equilibrium. The timescales will be discussed in more detail in Fig. 3.

c. Forcing experiments

The forcing experiments are initiated from the equilibrium states. Initialized from the last period of the equilibrium experiments, each forcing experiment is forced by a different *SOLC* which is held constant throughout the experiment, in order to search for the thresholds of *SOLC* that cause the transitions between the equilibrium states. Using this information, we generate a bifurcation diagram in Fig. 6.

d. Transient experiments

The transient experiments constitute the main body of simulations in this study. Initial conditions are generated by branching from the forcing experiments, but *SOLC* is set back to 338 W m^{-2} . These model-generated transient initial states are physically self-consistent, which allows for a systematic search for multiple equilibria. These experiments eventually converge toward different equilibrium states depending on where they start.

e. Carbon decoupling experiments

Finally we explore the role of the carbon cycle in organizing the global climate near the Warm equilibrium by disabling its radiative effect. These experiments are described in section 4.

3. Multiple equilibria

a. Identify multiple equilibria in MITgcm

Fig. 2A shows the scatterplot of ASR and OLR versus global mean surface air temperature (TS) for the fully coupled energy-conserving MITgcm, in close analogy with Fig. 1. In this map, we combine all equilibrium and transient experiments (the forcing experiments are not included because they use different values of *SOLC*). The blue and red curves made up of numerous dots, each representing a 20-year-mean snapshot of climate, are almost overlapped. There are seven intersection points, four of which are stable climatic equilibria, defined as Warm, Cold, Waterbelt, and Snowball (terminology following Rose 2015), while the rest of the intersections between the

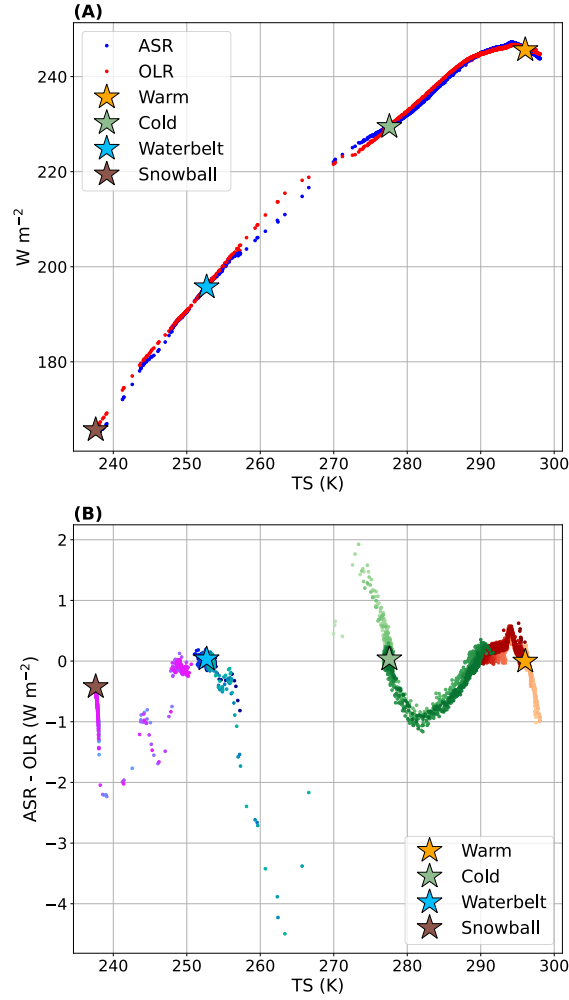


FIG. 2. Scatterplot of (A) global mean TOA radiative fluxes (blue dots: absorbed shortwave radiation (ASR); red dots: outgoing longwave radiation (OLR)) versus global mean surface air temperatures (TS) and (B) global mean net TOA radiative fluxes (ASR – OLR) versus TS from all equilibrium and transient experiments. Different groups of gradually changing colors represent all equilibrium and transient climates attracted by different climate states. Red and orange colors represent Warm branch; Cold branch is displayed from light green to dark green; Waterbelt branch ranges from light blue to deep blue; Snowball branch ranges from magenta to cyan. The first 500 years of the Warm branch of the transient experiments (which all converge to the Warm state) are omitted. Four stable climate states are marked by stars.

stable ones are unstable equilibria. Compared to the EBM (Fig. 1), two additional stable glacial states are identified, in which the polar sea ice extends to either mid-latitudes (Cold) or subtropics (Waterbelt). We omitted the data from the first 500 years of simulations in the Warm branch (which

all converge to the Warm state) to filter out noise induced by the slower adjustment of the climate system to the new *SOLC* forcing. The location of the unstable equilibrium between Cold and Waterbelt is not identified due to the numerical instability occurred near that point, which is likely caused by the very rapid climate change near the unstable equilibrium. The large tendency, as we will show in Fig. 4A, may be resulted from the radical change in the atmospheric $p\text{CO}_2$. We acknowledge that there is a small (less than 0.5 W m^{-2}) negative TOA energy imbalance for the Snowball state because of the similar numerical issue and the lack of geothermal heating (Ferreira et al. 2011). Also, the fact that the unstable equilibrium between Waterbelt and Snowball is too close to the Waterbelt state makes it hard to discern.

To view the relationship from a different perspective, Fig. 2B is the scatterplot of the net TOA radiative fluxes (i.e., $\text{ASR} - \text{OLR}$) versus TS. In this diagram, four stable equilibria as well as some unstable equilibria can be seen by locating the intersection points between climate trajectories and the zero TOA imbalance. A stable equilibrium is identified whenever the adjacent trajectories have a negative slope (i.e., negative net feedback), and by contrast, an unstable equilibrium is associated with a positive slope (i.e., positive net feedback). Note that between the Warm and Cold states, there exists a “barely” unstable equilibrium near 290 K, because the slope is almost flat. After passing the unstable equilibria, the climate trajectories first undergo an accelerating phase characterized by a positive net feedback, and then a decelerating phase dominated by a negative net feedback. Note that there is a huge gap between the Cold and Waterbelt solutions in Fig. 2A (267–270 K) implying a very large TS range between the adjacent transient climates associated with a large positive net feedback.

We next explain how the trajectories in Fig. 2 are generated by the model. Fig. 3 shows the timeseries of TS, equivalent ice-edge latitude (Rose et al. 2013) and atmospheric $p\text{CO}_2$ for all equilibrium and transient experiments. Because the atmospheric $p\text{CO}_2$ is tightly coupled to the ocean carbon reservoir (one of the slowest components in this model), even if both TS and sea ice have approached the equilibrium, the carbon cycle may still drift at a slower pace. For instance, the $p\text{CO}_2$ timeseries in the Warm branch have not yet approached the equilibrium even after 15,000 years.

In Fig. 3A and 3B, between the Warm and Cold branches, the darkest red takes over 8000 years to shift to the Warm regime, while the nearby darkest green takes more than 5000 years to transition

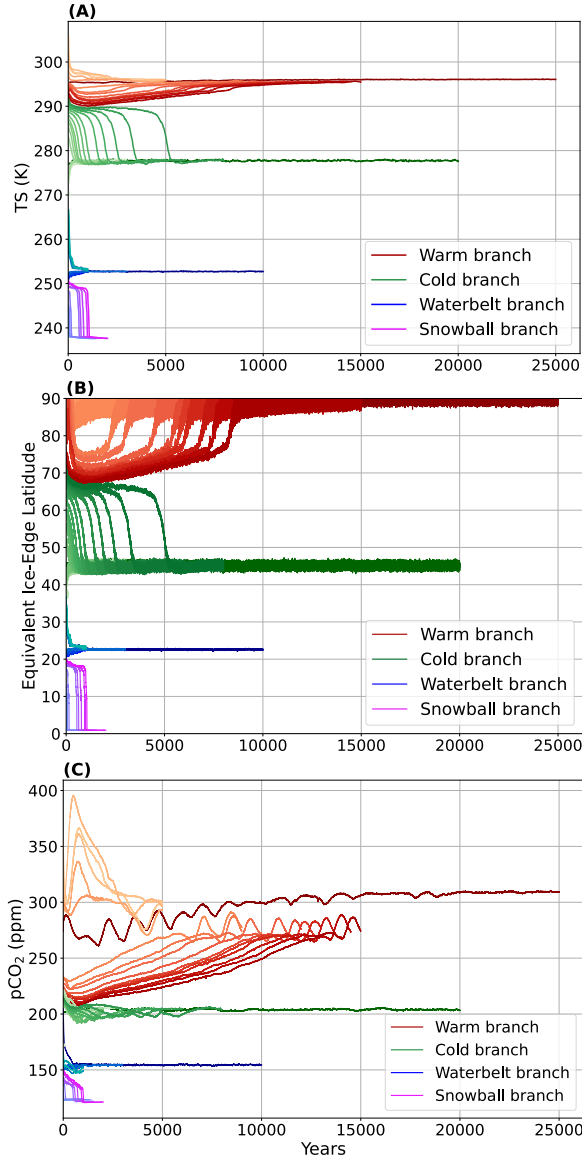


FIG. 3. Timeseries of annual and global mean surface air temperatures (A), monthly equivalent ice-edge latitudes (B) and atmospheric $p\text{CO}_2$ (C) from all equilibrium and transient experiments. Colors are classified into four categories: Warm branch, Cold branch, Waterbelt branch and Snowball branch. The grouping convention is the same as in Fig. 2.

to the Cold state. The long timescales around the “Temperate barely unstable state” seem to be remarkable since they do not appear anywhere else. As we will show below, the net feedbacks are just above zero during this time (Fig. 10 and Appendix B).

226 The global carbon cycle is considerably activated during the warming phase. This can be
227 observed in the timeseries of the atmospheric $p\text{CO}_2$ in the darkest red between the Warm and Cold,
228 and the one reaching the Warm equilibrium (Fig. 3C), in which $p\text{CO}_2$ take many thousands of years
229 to rise by nearly 100 ppm, a magnitude comparable to the atmospheric $p\text{CO}_2$ difference between
230 the peak glacial and interglacial periods (Sigman and Boyle 2000; Past Interglacials Working
231 Group of PAGES 2016). In contrast, our transient experiments converging to the colder states
232 do not exhibit such dramatic $p\text{CO}_2$ changes and prolonged timescales. More intriguingly, some
233 self-sustained millennial-scale oscillations of $p\text{CO}_2$ are superimposed on the long-term increase
234 during the warming phase¹. We also notice, from Fig. 3A and 3B, that close to 75° of the equivalent
235 ice-edge latitude, the small ice cap instability (ice cap smaller than a certain size is unstable, see
236 North (1984)) occurs. This abrupt event coincides with an increase in the growth rate of the
237 atmospheric $p\text{CO}_2$. After the disappearing of sea ice, an extremely long period of adjustment to
238 the Warm state ensues. This is because the climate system is going through the slow manifold to
239 which all fast processes (e.g., albedo, water vapor, clouds, etc) tend to equilibrate (Saltzman 2001)
240 due to the carbon cycle processes as mentioned above.

241 If Fig. 3 mainly provides us with the information about the timescales, then Fig. 4 offers insights
242 on which process may be the main driver of the climate change, using TS as a metric of the
243 global mean climate. The most prominent feature of the relationship between the atmospheric
244 $p\text{CO}_2$ and TS (Fig. 4A) is the state dependence. The Warm branch is characterized by an almost
245 linear and sharp slope (large increase in $p\text{CO}_2$ per degree warming), while the slopes in the Cold
246 and Waterbelt branches are more intricate and flat. Similarly, Fig. 4B displays the relationship
247 between the equivalent ice-edge latitude and TS, and the state dependent slopes reflect the relative
248 rate of change in the ice-edge latitude against global mean climate. Close to 75° , the small ice
249 cap instability shows up more clearly, which appears to be the most abrupt change in the ice-edge
250 latitude. However, the rate of change in the ice fraction at this point is comparable to that near the
251 Snowball (not shown), where the large ice cap instability (Rose and Marshall 2009; Rose 2015)
252 occurs.

¹Preliminary analysis suggests that these oscillations are associated with millennial-scale changes in high latitude ocean stratification and meridional overturning circulation. We speculate that both the initiation and termination of the millennial cycles are triggered by some thresholds of sea ice, which affect deep water formation and surface CO_2 exchange, but we do not investigate further in this study.

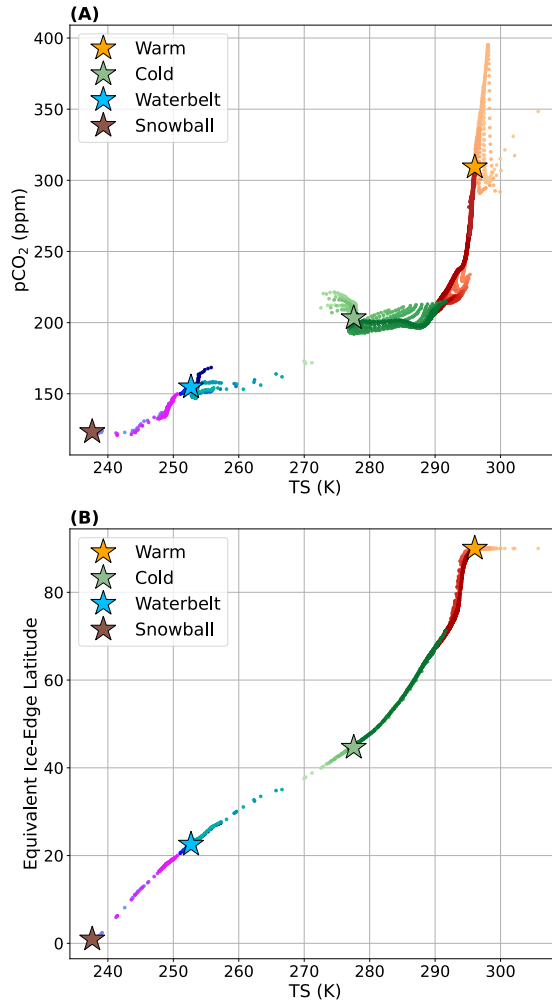


FIG. 4. Scatterplot of atmospheric pCO₂ (A) and equivalent ice-edge latitude (B) versus TS. Colors are classified into four categories: Warm branch, Cold branch, Waterbelt branch and Snowball branch. The grouping convention is the same as that in Fig. 2. Four stable climate states are marked by stars.

Comparing Fig. 4A and 4B, one significant distinction stands out. The slopes related to the ice-edge latitude appears to be much more “linear” than those associated with pCO₂². Actually, the scatters in Fig. 4A fail to display a well-defined relationship between pCO₂ and TS for the transient evolution of climates colder than 294 K. A reasonable speculation for this phenomenon is that pCO₂, and hence the global carbon cycle, may be a somewhat less important or indirect factor of the climate change, because carbon cycle acts on longer timescales than TS does. Climates in those cooler regimes are more susceptible to faster processes such as surface albedo, water

²Note however that neither ice edge latitude nor pCO₂ are expected to be linear with respect to their radiative impacts.

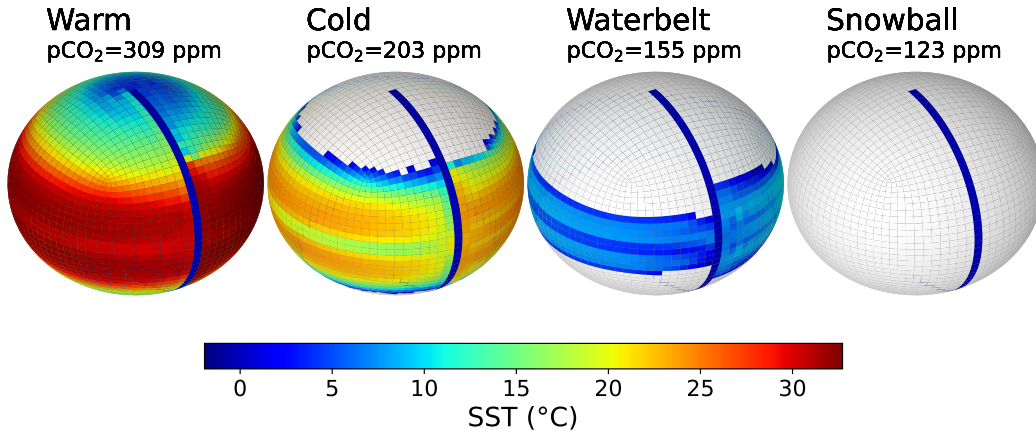


FIG. 5. Spatial patterns of the climatological sea surface temperatures (SSTs) for four stable equilibria shown on the cubed spheres. The SST fields are derived from the last 20-year-mean data from equilibrium experiments. Different states and their equilibrium atmospheric $p\text{CO}_2$ values are labeled above the corresponding spheres. The thin, pole-to-pole lines in the middle of each sphere represent the *Ridge*. The white grids stand for sea ice with the fraction greater than 0.15.

vapor, lapse rate and clouds, making adjustments at shorter timescales (less than 20 years, the temporal resolution of the scatterplots). However, when TS is above 294 K, the more clearly defined relationship between $p\text{CO}_2$ and TS may be indicative of a switch of the main contributing factor of the climate change, as section 4 suggests. At this point, the atmospheric $p\text{CO}_2$ and TS vary on almost the same timescales. Lastly, we spot a discontinuity in the $p\text{CO}_2$ versus TS relationship between the Cold and Waterbelt states near 271 K, in contrast to the smoother appearance in Fig. 4B. This could be the evidence that it is the radical shift in the global carbon cycle, rather than the ice albedo, that substantially disturbs the climate system and thus causes the numerical instability.

b. A climatology of multiple equilibria

Now, let us take a look at the spatial distributions of the climatological sea surface temperatures (SSTs) in four stable equilibria (Fig. 5). The SSTs are displayed on their native cubed spheres so that their three-dimensional structures can be observed. The *Ridge* is depicted by a line of grid cells in deep blue connecting North and South poles. From left to right, the equilibrium climate is getting colder, accompanied by an increase in the sea ice extent and a declined atmospheric $p\text{CO}_2$.

283 The annual mean Warm state is almost ice-free. A tiny fraction of sea ice only appears during the
 284 cold season (Fig. 3B). The polar SST is above the freezing point, and most of the tropical SST
 285 exceeds 30° C. The pCO₂ value, however, is only 309 ppm, close to the preindustrial level. In the
 286 Cold state, the ice extends to 45° and the pCO₂ drops to 203 ppm, comparable but slightly higher
 287 than the Last Glacial Maximum (180-190 ppm, Petit et al. 1999). The Waterbelt state has the ice
 288 edge stabilized in the subtropics (23°) with the pCO₂ of 155 ppm. The last one is the Snowball,
 289 a completely frozen Earth with only 123 ppm of pCO₂. Compared with Rose (2015), the Warm
 290 state found here is slightly warmer, especially in the polar regions. The Cold state is also warmer,
 291 in which the equivalent ice-edge latitude is approximately 0.7° closer to the poles. The Waterbelt,
 292 however, appears to be more vulnerable to the large ice cap instability due to the radiative effect of
 293 lower pCO₂ value compared to Rose (2015), making this state colder.

294 We also examine the bifurcation diagram in Fig. 6A, which provides us with all possible stable
 295 climate solutions given a range of external forcing parameters. To produce the bifurcation diagram
 296 for this model configuration, we run all forcing experiments displayed in Fig. 6B. Because each
 297 simulation could take up to tens of thousands of years, we only consider the higher and lower
 298 bounds of each set of stable climate solutions. Between higher and lower end points, a linear
 299 extrapolation is drawn (shown as a line segment) assuming the model exhibits similar properties as
 300 in Rose (2015) (compared to their Figure 7). We vary the radiative forcing by 1 W m⁻² of *SOLC*
 301 or 4 W m⁻² of solar constant (*S*₀). The relatively coarse resolution of external forcing parameters
 302 results in the “identical” threshold for the Warm and Cold at *SOLC* = 336 W m⁻². We do not
 303 dig for more precise bifurcation thresholds up to one decimal place due to the exceptionally long
 304 computational time cost by the Warm simulations (Fig. 6B). The higher bound of the stable Warm
 305 solutions and that of the Snowball have not been identified due to some numerical issues. For
 306 example, when perturbed by a very high *SOLC*, e.g., 356 W m⁻² (the dark red line in Fig. 6B), the
 307 Warm climate seems to be out of control at year 6200. There are two possible scenarios including
 308 a hotter but stable additional equilibrium or a runaway greenhouse, but the model blows up before
 309 the future scenario unfolds. Similar numerical instability also occurs when forcing the Snowball
 310 with a very high solar constant.

322 At least four important traits stand out in the bifurcation diagram. First of all, the Cold and
 323 Waterbelt states would be inaccessible via a hysteresis loop in *SOLC*, which means they cannot

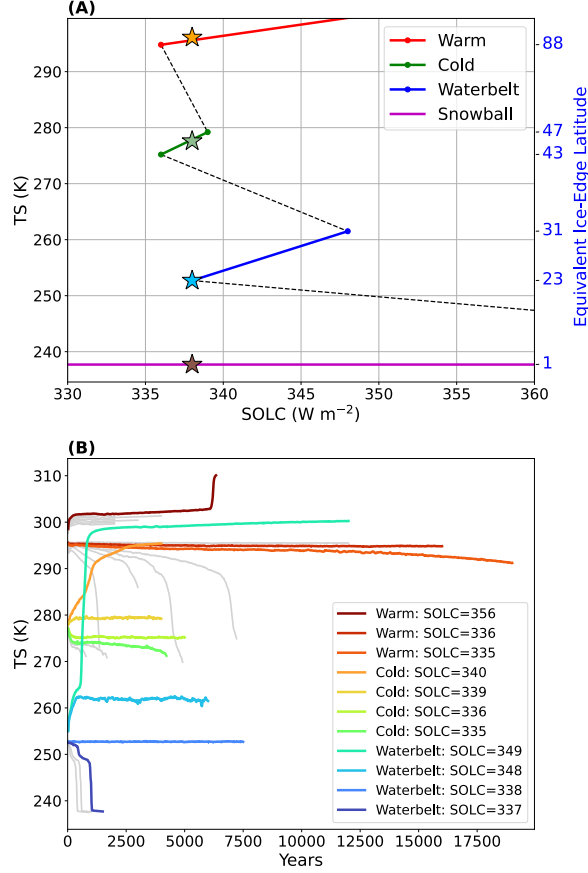


FIG. 6. Bifurcation diagram (A) and timeseries of TS in the forcing experiments (B). The solid line segments in (A) represent stable solutions given a solar forcing $SOLC$, while the dashed line segments stand for unstable solutions. Stable Warm solutions can be found for $SOLC$ (or $S_0/4$) being at least 336 W m^{-2} ; stable Cold solutions survive between 336 and 339 W m^{-2} ; stable Waterbelt solutions exist between 338 and 348 W m^{-2} . The higher bound of the stable Warm solutions and the threshold for the Snowball to transition to the Warm state have not been quantified. Four climatic equilibria found in this study ($SOLC=338 \text{ W m}^{-2}$) are marked by stars. The timeseries of TS in (B) show how the bifurcation diagram is produced. Only the experiments that determine the thresholds in the bifurcation diagram are labeled. These forcing experiments all start from the equilibrium climates (Warm, Cold and Waterbelt) with different $SOLC$ (the first points plotted in each curve do not exactly overlap because they are the output of the first 20 years of simulation). Each simulation either stays at the same state or transitions into another stable state. The unit of $SOLC$ in (B) is W m^{-2} .

be accessed from warmer states merely by lowering the radiative forcing. In fact, any transitions exerted by slow-varying (compared to the timescales of internal climate variability) external forcings

would fall on either Warm or Snowball. Secondly, the Cold state seems to be more vulnerable compared to Warm, Waterbelt and Snowball because the range of the stable Cold solutions is much narrower, and the Warm state is the most stable non-Snowball state for the wider range of solar radiation it walks through. Thirdly, the timescales of the Warm to Snowball transitions (cooling phase) are significantly longer than those of the warming phase (before the slow carbon feedback takes over), due to the prolonged adjustment of the carbon cycle near the Warm state. And lastly, four equilibria could only exist when *SOLC* ranges from 338 to 339 W m⁻² (or solar constant in a range of 1352–1356 W m⁻²).

As Rose (2015) mentioned, the specific details of the bifurcation diagram vary with different models and configurations. The diagram shown in Fig. 6 differs with that in Rose (2015) in several important aspects. Two glacial states from that paper can be accessed through transitioning from a warmer state under slow-varying forcings. The Warm and Cold solutions in our study are shifted slightly towards low *SOLC* (lower solar forcing is required for the same equilibria) compared to Rose (2015), indicating the role of frictional heating. The Waterbelt, however, is shifted considerably towards high *SOLC* as a result of both frictional heating (warming effect) and the radiative effect of lower pCO₂ (cooling effect). The latter effect is undoubtedly dominating the overall changes, making the Waterbelt here more sensitive to the large ice cap instability.

To summarize, the introduction of the carbon cycle and frictional heating substantially decreases the overlap between stable climate ranges (especially between the Cold and Waterbelt) that inevitably narrows down the range of multiple stable solutions. Moreover, the slow response of the carbon cycle considerably lengthens the timescales of transitioning from/to the Warm equilibrium.

4. Feedback analysis

After qualitatively discussing possible feedback mechanisms in section 3, in this section, we perform a quantitative feedback analysis using the radiative kernel method. A brief introduction of the technical details regarding the radiative kernels and feedback parameter calculations are presented in part a, followed by the results in part b.

352 *a. Methods*

353 We apply the radiative kernel method (Soden et al. 2008) to calculate feedback parameters. For
 354 a variable x with radiative effects, the feedback parameter is defined as

$$\lambda_x = \frac{\partial R}{\partial x} \frac{dx}{dT_S} \quad (2)$$

355 with the unit $\text{W m}^{-2}\text{K}^{-1}$, where R is the radiative flux at TOA (defined as positive down), and T_S
 356 is surface temperature.

357 The first term on the right hand side of Eq. (2) is the radiative kernel. It is a partial derivative
 358 of the TOA radiative flux versus a specific climate field (e.g., temperature, specific humidity,
 359 etc.). We develop the radiative kernels for air temperature, surface temperature, water vapor,
 360 surface albedo and CO_2 at Warm, Cold and Waterbelt states. To calculate the kernels, we first
 361 choose a reference state, and then perturb it in the radiation code by 1 K for temperature, 6%
 362 of specific humidity for water vapor³, 0.01 for surface albedo, and 1% for atmospheric pCO_2 ,
 363 respectively. The corresponding instantaneous radiative flux change relative to the control climate
 364 can be measured for each perturbation. We perform these offline perturbations once a day for a year.
 365 Air temperature and water vapor kernels are computed at different model levels. Finally, the zonal
 366 and annually averaged radiative kernels are computed. To avoid the uncertainty induced by the
 367 internal variability of a single reference climate, we repeat the same procedure but for another four
 368 different reference climates within the same climate regime (i.e., near the same equilibrium), evenly
 369 sampled from the last 5000 years of simulations. Fig. 7 shows the zonal and annual mean spatial
 370 structures of the radiative kernels for Warm, Cold and Waterbelt states. Because the kernels based
 371 on five different reference climates look strikingly similar, here we only present one realization of
 372 the radiative kernels for convenience.

377 To complete the feedback analysis we also need to calculate the second term in Eq. (2), i.e.,
 378 the climate response, or the climate change patterns. This is the derivative of a climate variable
 379 with respect to the global averaged T_S for the same period. Here we use a random sampling
 380 approach to minimize any biases associated with unforced internal climate variability. For each of
 381 the three equilibrium states, we select two periods of 1980 consecutive years each, that stand for

³There is an additional complication due to the fact that this model computes cloud amounts diagnostically based on relative and specific humidity thresholds (Molteni 2003), such that perturbations to air temperature and specific humidity cause instantaneous perturbations to cloud amounts. To alleviate this issue, we modify the source code to hold the cloud amount fixed during these calculations.

“pre-equilibrium” and “equilibrium” stages. Then, we randomly select 9801 individual pairs of time slices from these periods to calculate a large pool of feedback parameters. The results indicate that they roughly follow the Gaussian distribution (not shown).

For the cloud feedback, we follow Soden et al. (2008) by combining the cloud radiative forcing (CRF), the difference between total-sky and clear-sky TOA radiative fluxes, and the difference between total-sky and clear-sky feedback parameters (air temperature, water vapor, etc.) in each LW/SW band.

b. Results

1) FEEDBACKS IN MULTIPLE EQUILIBRIA

Let us first look at the global, annual mean feedback parameters for Warm, Cold and Waterbelt states (Fig. 8 and Table 1). The temperature feedback is decomposed into the Planck feedback and lapse rate feedback. We also show the net LW/SW as well as the net feedbacks, i.e., the sum of individual feedbacks. A curious result is that the relative roles of net LW and SW feedbacks are reversed in the Warm state relative to the Cold and Waterbelt states. This is consistent with the reversal of the slopes of the ASR and OLR curves in Fig. 2 for very warm climates. Specifically, the net LW feedback in the Warm state is weakly positive (destabilizing) and the net SW feedback is negative (stabilizing), while the reverse is true in the Cold and Waterbelt states.

Digging into this result, we find that the change of sign of the net LW feedback is mainly due to the combined water vapor plus lapse rate feedback, which exhibits a strong state dependence (more positive in warmer climates). Meanwhile the dominant reason for the change of sign on the SW side is the surface albedo feedback, which is strongly positive in colder climates but near zero in the Warm state. These compensating state dependencies in LW and SW bands between the Warm and colder climates result in net feedbacks that vary less than their individual components.

The kernel analysis allows us to separate out the radiative effects of clouds. This is of interest because the cloud processes in our model are highly parameterized and less well constrained by basic physics than the clear-sky processes, thus might be considered less robust. In all three states, the SW cloud feedbacks are negative and partially mitigated by positive LW cloud feedbacks, leading to negative net cloud feedbacks. Thus clouds act to stabilize the climate system in all cases. Would any of the three equilibrium states become unstable in the absence of cloud changes? Our

TABLE 1. Global and annual mean feedback parameters and standard deviations computed from two realizations of radiative kernels, in units of $\text{W m}^{-2} \text{K}^{-1}$. LR, lapse rate; WV, water vapor; LW, longwave; SW, shortwave.

	Planck	LR	WV (LW)	LR+WV (LW)	Carbon	Cloud (LW)	Cloud (SW)	WV (SW)	Albedo
Warm	-3.50 ± 0.26	-0.13 ± 1.14	2.77 ± 0.92	2.66 ± 0.51	0.35 ± 0.36	0.82 ± 0.42	-1.05 ± 1.32	0.13 ± 0.03	0.01 ± 0.03
Cold	-3.23 ± 0.32	0.75 ± 0.32	0.92 ± 0.29	1.67 ± 0.27	0.10 ± 0.06	0.37 ± 0.11	-0.43 ± 0.18	0.11 ± 0.02	1.22 ± 0.34
Waterbelt	-3.26 ± 1.14	0.90 ± 0.76	0.37 ± 0.66	1.26 ± 0.74	0.04 ± 0.16	0.08 ± 0.09	-0.32 ± 0.64	0.07 ± 0.03	1.63 ± 0.93

results suggest not, because the negative net cloud feedback is smaller in absolute value than the overall net feedback in all cases. We conclude that state dependence of cloud feedbacks is not a primary factor in shaping the multiple equilibria.

The “error bars” for the feedback parameters shown in Fig. 8 and Table 1 are standard deviations derived from our random sampling of time slices. They indicate the typical range of feedback parameters spanned by the sampled internal variability, or in other words, a measure of the error that might be encountered if the feedback parameters were under-sampled. It is likely that the feedback processes governing short-term variability differ systematically from those governing the long term drift toward equilibrium due to differing spatial pattern effects (e.g., Andrews et al. 2018). Our sampling strategy was designed to measure the feedbacks operating on the longest timescales. The Cold state exhibits the smallest feedback range because the samples are drawn from simulations that undergo larger amplitude climate changes, so the “signal” (the adjustment to equilibrium) is larger than the “noise” (the internal variability).

To make sense of the net feedbacks just calculated, we relate these numbers to Fig. 6A. Consistent with previous discussions, the Warm state is the most stable non-Snowball state and the Cold state is relatively unstable, because the net feedback of the Warm state is more negative than that of the Cold state. To validate our kernel-derived net feedback parameters, we compare them with dR/dTS near the equilibria in Fig. 2. See Table 2 for the details. The good agreement between dR/dTS and the kernel-derived feedback parameters confirms the reliability of our calculations.

439 TABLE 2. dR/dTS (LW, SW and Net) by linear regressions and the global, annual mean LW, SW and
440 Net feedback parameters and standard deviations computed from radiative kernels, in units of $\text{W m}^{-2}/\text{K}$. LW,
441 longwave; SW, shortwave.

	Slope (LW)	Feedback (LW)	Slope (SW)	Feedback (SW)	Slope (Net)	Feedback (Net)
Warm	0.26	0.31 ± 0.77	-1.07	-0.91 ± 1.32	-0.81	-0.60 ± 1.15
Cold	-1.06	-1.10 ± 0.22	0.89	0.90 ± 0.45	-0.17	-0.20 ± 0.44
Waterbelt	-1.86	-1.87 ± 0.83	1.48	1.40 ± 1.09	-0.38	-0.46 ± 1.28

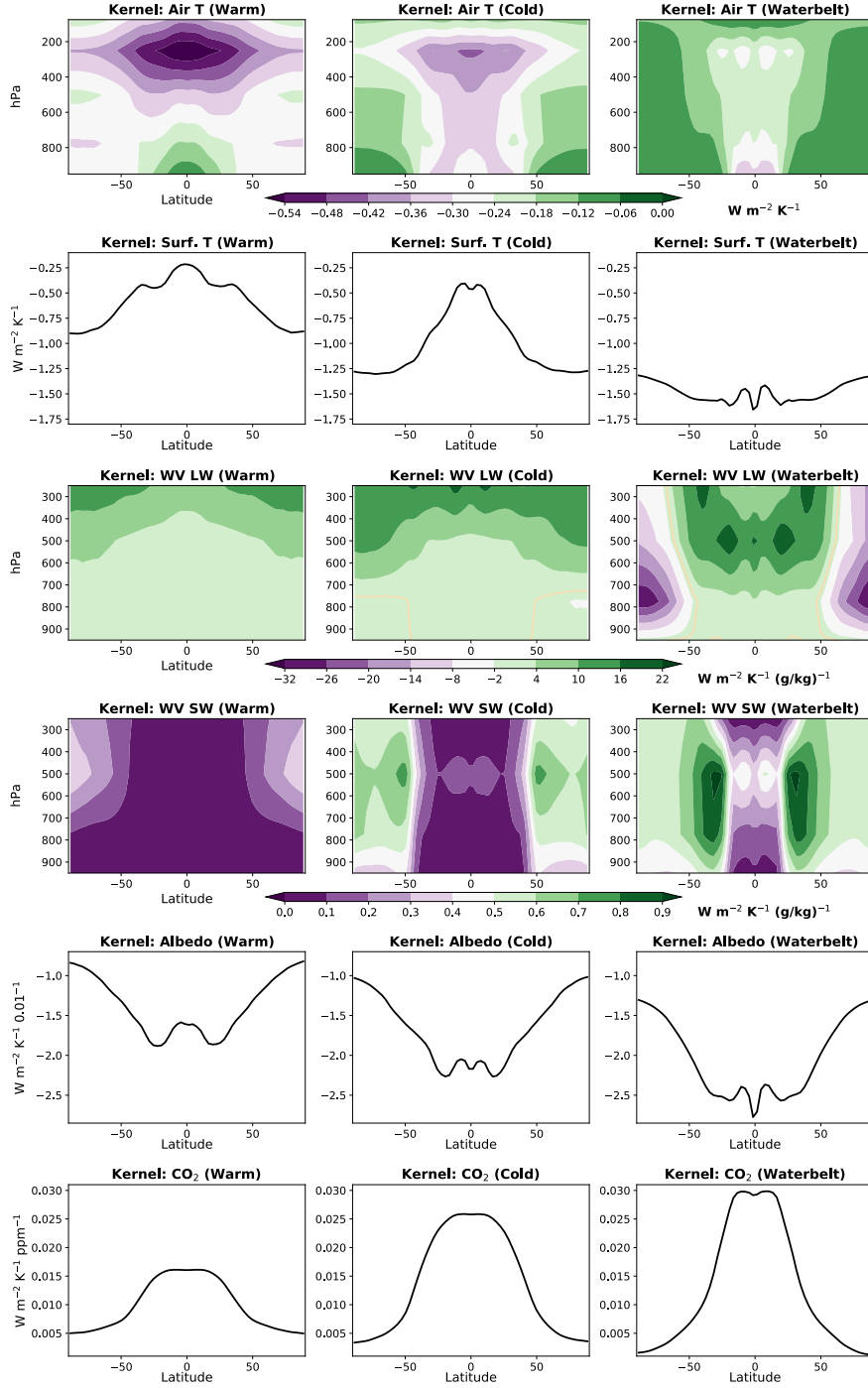


FIG. 7. Zonal averaged and annual mean radiative kernels of air temperature, surface temperature, longwave and shortwave components of water vapor, surface albedo and CO₂ for Warm (left), Cold (middle) and Waterbelt (right) states. The zero contours are highlighted as yellow, solid lines. These kernels are calculated based on one reference climate for each climate regime.

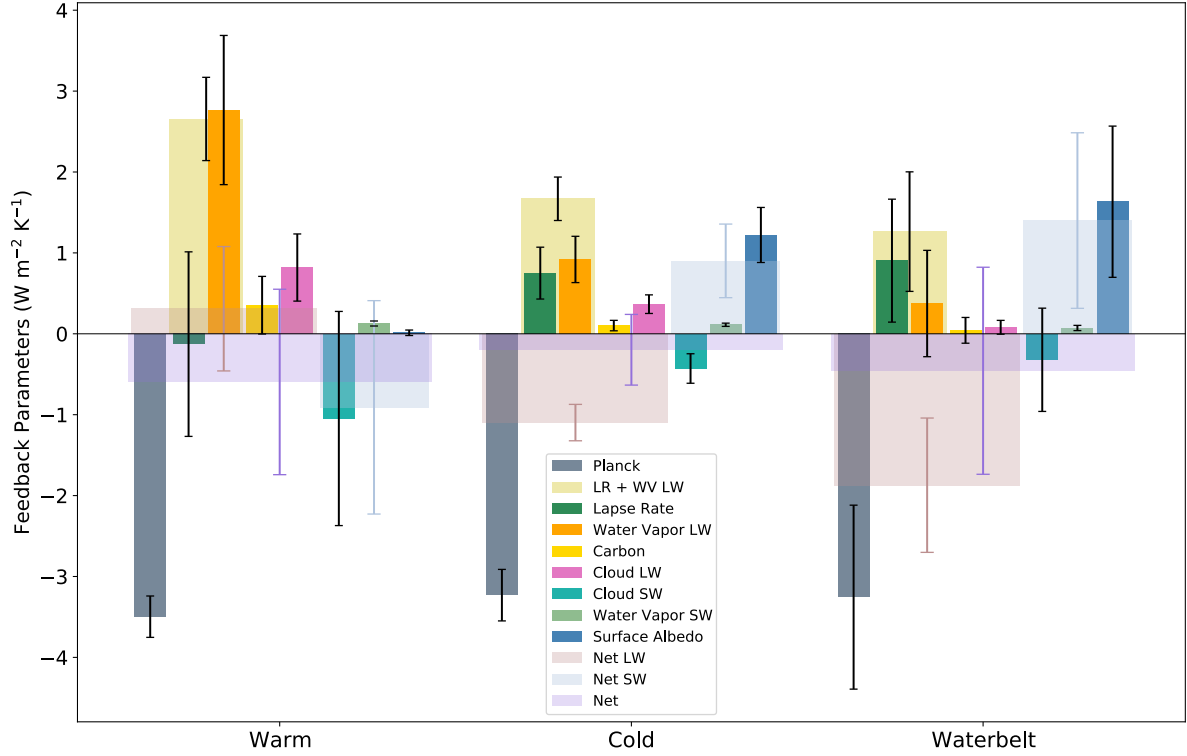


FIG. 8. Global averaged and annual mean linear decomposition of feedback parameters for Warm, Cold and Waterbelt states. The Planck, lapse rate, water vapor, carbon, cloud and surface albedo feedbacks are displayed and grouped into either longwave or shortwave bins. The net longwave, shortwave and the net feedbacks (all in transparent colors) are overlaid on the corresponding bins. The error bars which indicate one standard deviation are plotted in back except for the net longwave/shortwave and net feedbacks. The results are based on the average of two realizations of radiative kernels. The feedback parameters are derived from the average on the 10th to 90th percentiles of the 9801 samples.

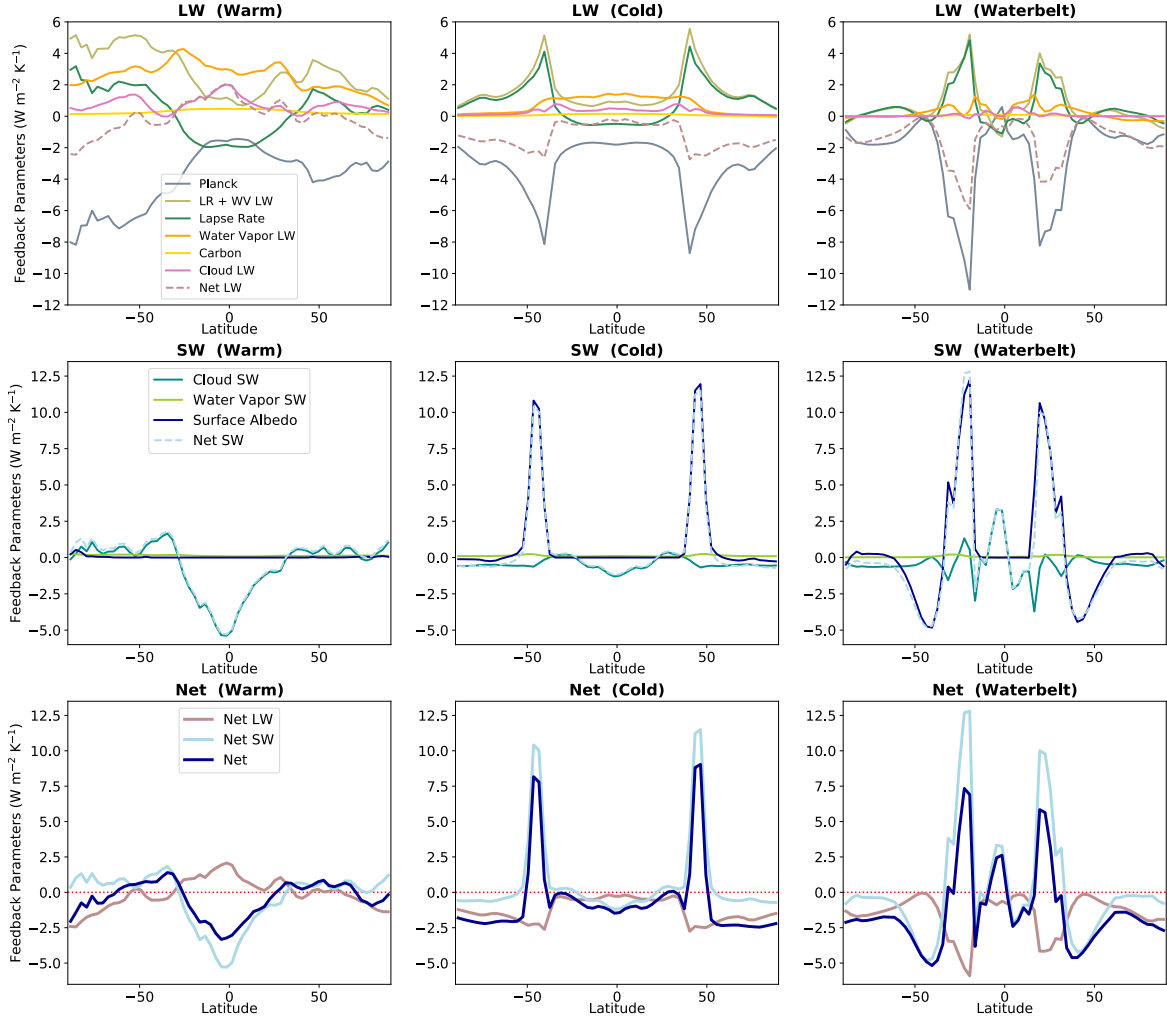


FIG. 9. Zonal averaged, vertical integrated, and annual mean feedback parameters in LW component (top), SW component (middle) and net feedbacks (bottom) for Warm (left), Cold (middle) and Waterbelt (right) states. The results are based on one realization of radiative kernels, the same as Fig. 7. The feedback parameters are derived from the median of the 9801 samples.

Next, a closer look at the spatial structure of (zonally averaged and vertically integrated) feedback parameters (Fig. 9). These figures show regional contributions to the global mean feedback parameters (Feldl and Roe 2013). Many more details are revealed, especially those processes associated with specific latitudes. For example, the Warm state exhibits asymmetry in the temperature (Planck and lapse rate) and LW water vapor feedbacks. Given our symmetric model boundary conditions, all asymmetries arise from internal variability. In this case the asymmetry results from a small amount of sea ice that was initially present in the Southern Hemisphere which resulted in some bottom amplified high-latitude warming (not shown). In the Cold and Waterbelt states, the feedbacks are much more symmetric. Poleward of the ice edge, temperature feedbacks dominate the LW feedbacks, while surface albedo feedback leads the SW feedbacks.

To summarize, the latitudinal feedback patterns are clear: for Warm, low-latitude processes dominate, especially LW water vapor feedback and SW cloud feedback; on the other hand, the Cold and Waterbelt are more dependent on the mid- and high-latitude processes, such as temperature feedbacks for LW and surface albedo feedback for SW. This conclusion is consistent with Westerhold et al. (2020).

2) FEEDBACKS IN THE WARM BRANCH

Now let us zoom in to focus on the Warm branch and try to understand three questions left over from Fig. 3: why does the “Temperate barely unstable state” last nearly 6000 years? What happens when the small ice cap instability occurs? And why does it take another 25,000 years for climate to settle down to equilibrium even if the radiative imbalance ($ASR - OLR$) is so close to zero?

Here in Fig. 10, we discuss the net feedbacks associated with a typical trajectory in the Warm branch (the darkest red trajectory at the Warm/Cold boundary in Fig. 3) as well as the trajectory of the Warm equilibrium experiment (Figs. 3 and 11). The net feedbacks are derived from dR/dTS . Because the starting point of the Warm equilibrium experiment (bottom panel of Fig. 10, after 2000 years) is in a very similar state to the end of the transient experiment (top panel of Fig. 10), the two timeseries together approximate a complete picture of the climate evolution throughout the Warm branch.

In the top panel, after some rapid adjustments in the first 2000 years, the climate system enters a barely unstable regime, with net feedbacks between 0 and $+0.1 \text{ W m}^{-2} \text{ K}^{-1}$ over 5000 years.

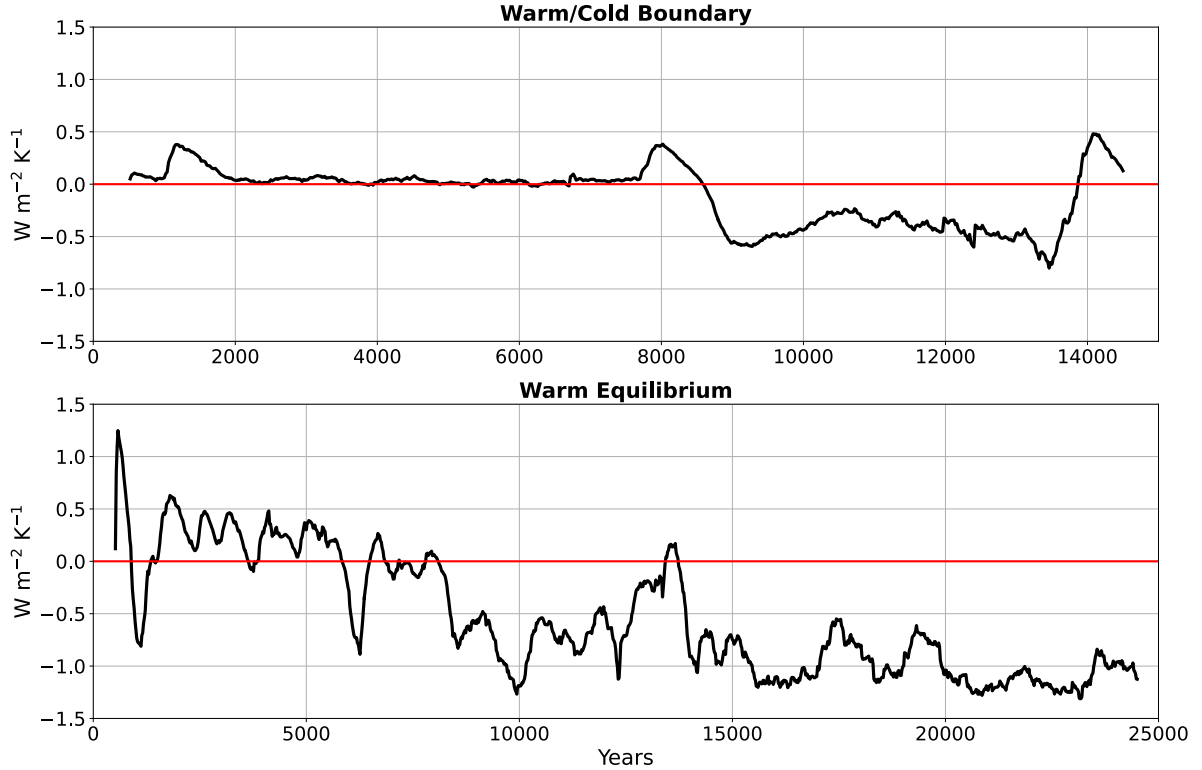


FIG. 10. Net feedbacks of the Warm branch derived from dR/dTS in the experiment at the Warm/Cold boundary (top), and the Warm equilibrium experiment (bottom). The slopes are the results of linear regression of radiative imbalance at TOA ($ASR - OLR$) versus TS at 1000-year interval. The red solid lines indicate zero net feedback.

We demonstrate that the stronger surface albedo feedback and weaker carbon feedback relative to the Warm equilibrium contribute to the weakly positive net feedback (Fig B1). As the climate slowly warms, sea ice retreats gradually (Fig 3), while carbon feedback slowly strengthens probably because the entire ocean is warming, and more ocean surface is exposed which enhances air-sea gas exchange. The net fast feedback (by which we mean the total feedback excluding the slow carbon feedback, following the definitions used by PALAEOSSENS Project Members (2012)) is slightly negative (Fig. B1), suggesting a very weakly stable condition. It seems that the fast processes almost equilibrate, while the weakly positive, slower processes (e.g., carbon cycle) nudge the system away from the equilibrium. Then, the rapid sea ice retreat, probably driven by the ocean processes (Rose et al. 2013), causes a ramp of surface albedo feedback, which corresponds to the spike of net feedback near year 8000. After a short period of rapid changes, the climate is back to

490 a temporarily stable condition at about year 8500 in the top panel, when polar sea ice almost melts
491 away (Fig. 3B). For the next thousands of years, despite being relatively stable overall, the carbon
492 feedback further strengthens (Fig. B2), which happens as CO₂ outgassing accelerates (Fig. 3C) and
493 ocean heat content increases (not shown).

494 The fast feedbacks almost equilibrate again until year 14,000 (top panel), when the carbon
495 feedback further intensifies (Fig. B2) to the extent that it pushes the net feedback back to the
496 positive territory accompanied by the self-sustained, millennial-scale oscillations (Figs. 3C and
497 11). These oscillations continue in the bottom panel from year 1500 to 6000, which consist of an
498 accelerating phase (e.g., net feedbacks from 0 to 0.6 W m⁻² K⁻¹ from year 1500 to 1800 in the
499 bottom panel of Fig. 10) dominated by positive feedback processes, and a decelerating phase (e.g.,
500 net feedbacks from 0.6 to 0.1 W m⁻² K⁻¹ from year 1800 to 2400 in the bottom panel of Fig. 10)
501 dictated by certain negative feedback mechanisms. The overall positive feedback associated with
502 the oscillations leads to 0.8 K warming of TS (Figs. 3 and 11). After year 8000, the bottom panel
503 of Fig. 10 shows a slow relaxation of the climate to the equilibrium as the oscillations gradually
504 subside. The overall declining trend of the carbon feedback (Fig. B2) confirms the result. We
505 speculate that the breakdown of the oscillatory behavior may be associated with a key component
506 of the oscillator reaching a threshold as the climate warms.

507 As an attempt to quantify the physical processes governing the carbon feedbacks, we further
508 analyzed the temperature-dependent solubility pump (Weiss 1974) for the experiments shown in
509 Fig. 10. Preliminary results suggest that the temperature-dependent solubility pump accounts
510 for about 33% of the total atmospheric pCO₂ change for the experiment in the top panel, while
511 it only contributes to a relatively small fraction (8%) of atmospheric pCO₂ change in the Warm
512 experiment (bottom panel). Although a complete ocean carbon pump decomposition (e.g., Ferreira
513 et al. 2018; Marzocchi and Jansen 2019) is required to provide a full explanation (we will investigate
514 in future work), current analysis indicates that processes such as sea ice, ocean circulation, and
515 biogeochemical activities predominantly contribute to the carbon feedback and the associated
516 nonlinear behaviors in the Warm experiment.

517 3) CARBON FEEDBACKS NEAR THE WARM EQUILIBRIUM

518 Since the carbon feedback plays a crucial role in the evolution of climate in terms of the timescale
519 (Figs. 3C and 10), climate variability (Fig. 3C) and mean state (Fig. 4A) near the Warm equilibrium,
520 we devote this section to evaluating the carbon feedbacks in the Warm equilibrium experiment and
521 ask the following questions: is the slow evolution of climate caused by the carbon feedback? Does
522 the carbon cycle dictate the climate variability? Is the carbon feedback decreasing over time?

523 We use the term “carbon feedback” rather than CO₂ feedback, because we aim to include the
524 global carbon cycle in our analysis. The global carbon cycle in this model refers to the atmospheric
525 CO₂ and the ocean carbon inventory, and the latter is almost 60 times as large as the former reservoir
526 (Williams and Follows 2011). Although we perturb the atmospheric pCO₂ by 1% and measure the
527 instantaneous response of the climate system, similar to what we did for other fast feedbacks, we
528 also account for the slow ocean components coupled to CO₂ by measuring the feedbacks operating
529 on the longest timescales using the random sampling strategy.

536 Here we conduct two carbon decoupling experiments, shown in Fig. 11. The first experiment
537 (EXP1) is initiated from the beginning of the Warm simulation but with the atmospheric pCO₂ fixed
538 at its initial value. Comparing EXP1 to the Warm simulation, it is found that: 1) the shortening of
539 timescale in EXP1 suggests that carbon feedback substantially lengthens the equilibration time of
540 the Warm climate; 2) the strong millennial oscillations before year 15,000 are originated from the
541 carbon cycle; 3) the equilibrium temperature in EXP1 is lower (about 0.6 K) than that in the Warm
542 simulation, suggesting that carbon feedback contributes to warming.

543 The second experiment (EXP2) from year 15,000 to 25,000 (the same period for calculating
544 the feedback parameters) has its atmospheric pCO₂ fixed to the year 15,000 value. As discussed
545 in Fig. 10, the net feedbacks can be derived from dR/dTS . The difference between the net
546 LW feedback of the Warm simulation and that of EXP1/EXP2 represents the carbon feedback
547 and contributions from the rest of LW feedbacks influenced by it. By comparing dR/dTS from
548 EXP1/EXP2 with Warm simulations, it is shown that the increase in the net LW feedback in EXP1
549 is $1.28 \text{ W m}^{-2} \text{ K}^{-1}$, while it is only $0.35 \text{ W m}^{-2} \text{ K}^{-1}$ for EXP2, which happens to be the same as the
550 calculated carbon feedback in the Warm state. This demonstrates that the carbon feedback gradually
551 subsides after year 15,000, consistent with Fig. B2 which is generated by radiative kernels.

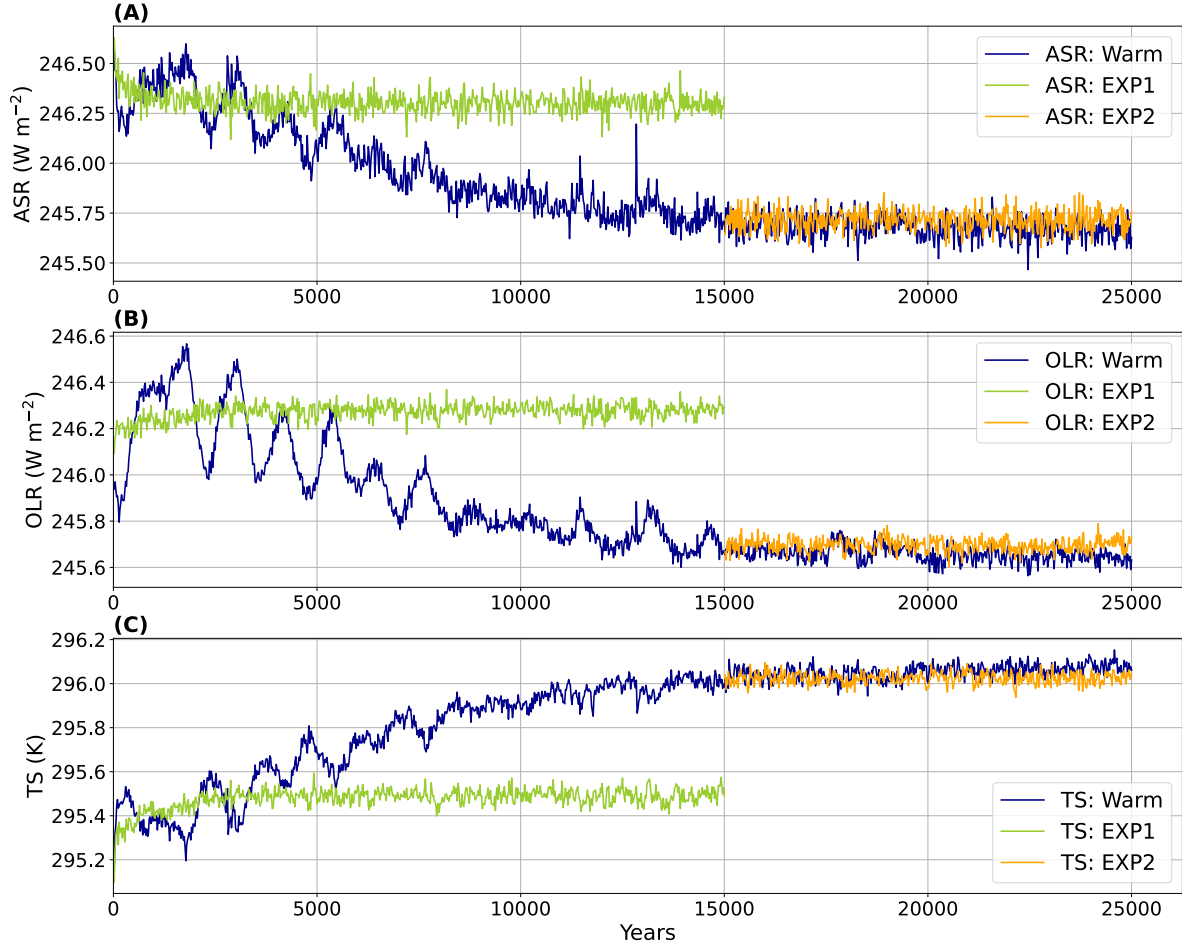


FIG. 11. Timeseries of two sets of carbon feedback experiments and Warm equilibrium experiment. Global averaged and annual mean timeseries of ASR (A), OLR (B) and TS (C) for the Warm equilibrium experiment are shown in dark blue. The first experiment (EXP1) timeseries are displayed in light green, which is initiated from the beginning of the Warm simulation but with an uncoupled carbon cycle. Similar to EXP1, the second experiment (EXP2) timeseries are shown in orange, which starts from year 15,000 and ends at year 25,000 without a coupled carbon cycle.

5. Discussions and Conclusions

Four stable climatic equilibria have been discovered in a fully coupled, energy-conserving, atmosphere-ocean-sea ice model (MITgcm) with an interactive global carbon cycle. The climatology of these states do not look much different from those in Rose (2015), but the underlying mechanisms distinguish our work from others largely due to the carbon cycle. Furthermore, the

557 closing of the TOA energy budget through frictional heating in our simulations has allowed a thor-
558 ough and detailed diagnostic study of the radiative feedback processes and their state dependence.

559 The overarching goal for this study is to understand the essential physical processes in the
560 current GCM, especially in the Warm climates, that are absent from the simplest EBM as well
561 as models without a carbon cycle. The main approach is to linearly decompose net radiative
562 feedbacks into individual processes using radiative kernels. The longwave water vapor feedback
563 (low-latitude processes) dominates the fast positive feedbacks in the Warm state, while the surface
564 albedo feedback (mid- and high-latitude processes) largely dictates the evolution of Cold and
565 Waterbelt states. This latitudinal feedback patterns associated with different states are consistent
566 with Westerhold et al. (2020). One way of understanding the existence of additional equilibria
567 in the MITgcm versus the EBM is that the more complex state-dependence of both longwave
568 and shortwave feedback processes in the GCM give rise to a larger number of intersection points
569 between the ASR(TS) and OLR(TS) curves (comparing Figs. 1 and 2).

570 Despite being relatively small in magnitude except for certain periods in the Warm climates,
571 the carbon feedback is crucial to the timescales, climate variability and mean climate near the
572 Warm equilibrium. The exceptionally long timescales toward the Warm state after temporarily
573 achieving the radiative balance by faster feedbacks suggests the entering into the slow manifold of
574 the dynamical system, which is caused by the carbon cycle. Upon reaching its climax, the carbon
575 feedback is in control of the global climate change, driving millennial-scale oscillations all the way
576 to the equilibrium. The rapid sea ice loss during the warming phase (which we interpret as the
577 small ice cap instability) is tightly coupled to the carbon cycle. During the course of this abrupt
578 change, the carbon feedback strengthens steadily.

579 The very long timescales associated with the carbon cycle and transitions to/from the Warm
580 equilibrium raise interesting questions about nonlinear and transient triggers for climate change.
581 Rose et al. (2013) and Rose (2015) considered hysteresis loops between Warm and icy (Cold or
582 Waterbelt) states in response to slow-varying *SOLC* in the absence of carbon feedbacks, where
583 “slow” meant significantly slower than the internal adjustment processes of the model. We have
584 shown that our energy- and carbon-conserving version of the model does not permit such a
585 hysteresis: a Warm climate subject to a very slow reduction in radiative forcing will inevitably go
586 through the Snowball catastrophe.

587 On the other hand, we must redefine what we mean by “slow” in the presence of carbon feedbacks,
588 as we have shown that even 15,000 years is not sufficient to reach the Warm equilibrium from an
589 almost energy-balanced condition. We speculate that transitions from Warm to Cold states could be
590 driven by temporary but long-lived radiative forcings that evolve over timescales comparable to this
591 slow carbon adjustment, as the climate system will neither have time to cool down enough to trigger
592 the Snowball catastrophe, nor will it get trapped in the slow manifold. This possibility is of great
593 interest due to the fact that orbital variations (not accounted for in our simulations) indeed cause
594 periodic radiative forcings on timescales of order 10^4 – 10^5 years. However, there are some caveats to
595 this speculation, notably that orbital variations principally cause seasonal and meridional rather than
596 global mean variations in insolation, and that our idealized *Ridge* model does not represent many
597 of the key zonal and interhemispheric asymmetries relevant to the Quaternary glacial-interglacial
598 cycles. Nevertheless, we think the possibility of such “rate-induced” transitions (Arnscheidt and
599 Rothman 2020) between multiple stable states paced by orbital variations and carbon feedbacks is
600 well worth exploring in future work.

601 Any lessons drawn from these simulations for the real Earth system are necessarily indirect due to
602 the reduced complexity and geometric simplicity of the model, including the relatively simple cloud
603 parameterizations. On the other hand, the many hundreds of thousands of years of simulations that
604 were required to study the equilibration of the coupled climate-carbon system would have been
605 prohibitively expensive in a more comprehensive model. With this trade-off in mind, we intend to
606 pursue stepwise increases in complexity up the model hierarchy in future studies of the response
607 of the coupled climate-carbon system to orbital forcings, for example, by adopting more realistic
608 asymmetries in land-ocean distribution, and coupling with an idealized dynamic ice sheet model.

609 *Acknowledgments.* We thank the MITgcm developer community for maintaining and supporting
610 such a rich research tool. This work was supported by NSF awards AGS-1455071 and AGS-
611 1952745, and internal seed funding from the University at Albany.

612 *Data availability statement.* The authors will post all MITgcm configuration files in a publicly
613 available code repository.

614 APPENDIX A

615 **MITgcm Configurations and Improvements in the TOA Energy Budget**

616 All components of our MITgcm simulations including atmosphere, ocean, and sea ice use the
617 cubed-sphere grid at a low C24 resolution (3.75° at the equator).

618 The atmosphere is a 5-level model using the physical parameterization based on SPEEDY
619 (Molteni 2003), including a moist convection scheme, diagnostic clouds, short- and longwave
620 radiation schemes with full seasonal cycle but diurnally averaged insolation, surface fluxes and
621 vertical diffusion. Pressure coordinates are used, with the bottom level in the boundary layer and
622 the top level in the stratosphere.

623 The ocean is 3 km deep with a flat bottom, using the rescaled height coordinate Z^* (Adcroft
624 and Campin 2004) with 15 levels. Advective mesoscale eddy transport is parameterized (Gent and
625 McWilliams 1990), along with a parameterization of isopycnal diffusion (Redi 1982) and convective
626 adjustment (Klinger et al. 1996). The sea ice is a 3-layer thermodynamic model based on Winton
627 (2000). Ice surface albedo depends on ice thickness, surface temperature, snow depth and snow
628 age. Diffusion of ice thickness is used as a proxy for ice dynamics.

629 The carbon cycle is represented through an ocean biogeochemistry model consisting of five
630 tracers: dissolved inorganic carbon (DIC), alkalinity, phosphate, dissolved organic phosphorus
631 (DOP), and oxygen (Dutkiewicz et al. 2006). Net primary productivity is calculated based on
632 phosphate and light availability. Sinking and remineralization of particulate organic matter are
633 represented. The ocean carbon cycle is coupled to the atmosphere through air-sea gas exchange
634 and a globally well-mixed atmospheric CO_2 reservoir. Atmospheric CO_2 exerts a radiative effect
635 through the idealized longwave radiation scheme.

636 Earlier papers including Ferreira et al. (2011), Rose et al. (2013), and Rose (2015) used the
637 same coupled model in a very similar *Ridge* configuration, except that the model does not conserve

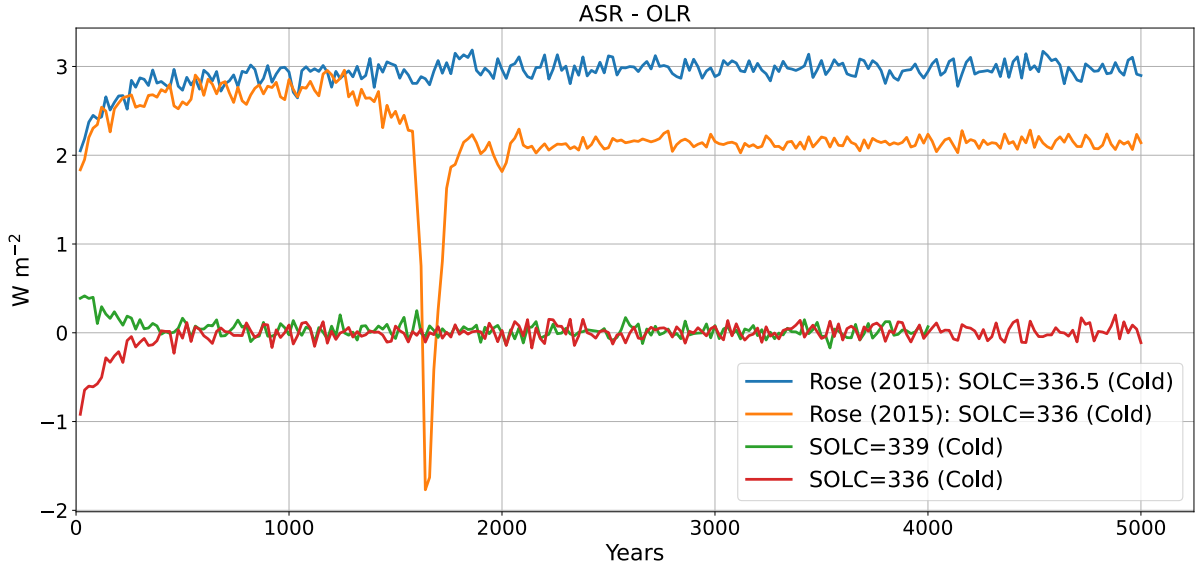


FIG. A1. TOA radiative imbalance (ASR – OLR) in this study (green and red lines) and in Rose (2015) (blue and orange lines). These simulations are all initiated from the Cold state, forced by *SOLC* indicated in the figure. The unit of *SOLC* is $W m^{-2}$.

energy at TOA and an active carbon cycle is not involved. By introducing the frictional heating to the climate system, here we demonstrate that our current model satisfies energy conservation.

Fig. A1 shows TOA imbalance in simulations previously reported by Rose (2015). Both experiments were initialized in the Cold state. One stabilized in the Cold state after some initial cooling, with $3 W m^{-2}$ net imbalance at equilibrium (blue). The other underwent a transition to Waterbelt, but with about $2 W m^{-2}$ net imbalance at equilibrium (orange). These imbalances are positive due to frictional dissipation of kinetic energy in the model atmosphere that is not returned to the fluid as internal energy. The imbalance is state dependent due to varying magnitudes of frictional dissipation associated with changes in atmospheric storm tracks. Fig. A1 illustrates the difficulty with using the imbalanced TOA radiative budget as a diagnostic, since it mixes physical signals (e.g., the large negative excursion near year 1700 associated with rapid cooling) with spurious unphysical feedbacks (the state dependence of the missing heat source).

In our new simulations including frictional heating, on the other hand, all trajectories converge toward zero net TOA radiative imbalance, regardless of the initial condition and the final equilibrium climate regime. Here, only two examples are shown (green and red lines in Fig. A1), but other

656 examples of closed energy budgets were shown in Fig. 2. Our diagnostic study of the TOA radiative
657 budget through the use of radiative kernel analysis would simply not be possible without a closed
658 energy budget.

659 APPENDIX B

660 Feedback Analysis of the Warm Branch

661 *a. Temperate barely unstable state*

662 Both Figs. 2 and 3 suggest a barely unstable state between Warm and Cold, which we refer to as
663 “Temperate”. This state features an extremely slow evolution to the Warm state (also to Cold, but
664 with a faster rate). The TS (290 K) and ice extent (edge near 70°) are comparable to the present-day
665 climate, while the pCO₂ (210 ppm) is significantly lower.

666 For insight into the underlying physical processes for the barely unstable state, we use radiative
667 kernels to decompose the net feedback (Fig. B1). Unfortunately, we do not have the necessary clear-
668 sky diagnostics to compute cloud feedbacks (Fig. 8, Fig. 9 and Table 1), so we instead estimate
669 them as residuals between net feedbacks derived from dR/dTS and the sum of kernel-derived
670 temperature, water vapor, surface albedo, and carbon feedbacks.

671 As expected, dR/dTS (net feedback) is just above zero, which indicates a very small positive net
672 feedback ($+0.06 \text{ W m}^{-2}$). This causes the climate system to spend a longer-than-usual time near this
673 unstable state. While lapse rate and water vapor feedbacks are similar to the Warm state, surface
674 albedo feedback ($+0.76 \text{ W m}^{-2}$) is closer to the Cold state. This additional positive contribution
675 from surface albedo feedback as well as the relatively small carbon feedback compared to Warm
676 are the primary reasons for the slightly positive net feedback.

683 *b. Carbon feedbacks in the Warm branch*

684 The carbon feedback plays an essential role in the evolution of Warm climate, as we discussed
685 in section 4. In Fig. B2, we calculate carbon feedbacks for different periods during the adjustment
686 toward Warm equilibrium.

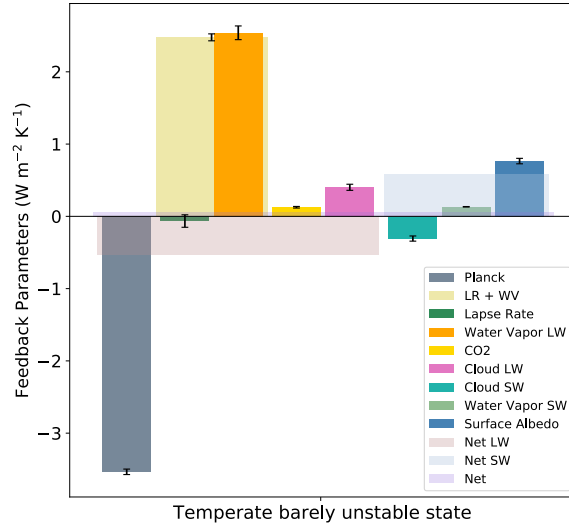


FIG. B1. Global averaged and annual mean linear decomposition of feedback parameters for the “Temperate barely unstable state” between the Warm and Cold using radiative kernels. The net LW/SW and net feedbacks are derived from dR/dTS . The cloud (LW/SW) feedbacks are calculated as the residuals of the net LW/SW feedbacks. The error bars which indicate one standard deviation are plotted except for the net LW/SW and net feedbacks. The feedback parameters are derived from the average on the 10th to 90th percentiles of the 9801 samples.

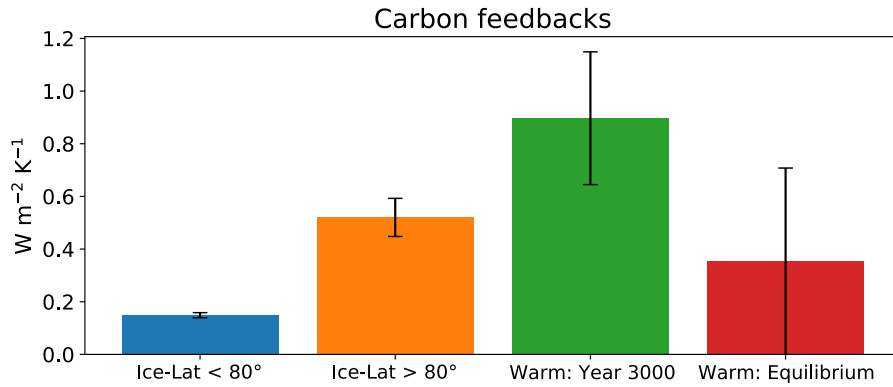


FIG. B2. Global averaged and annual mean carbon feedbacks using radiative kernels during different periods of the Warm branch simulations. The left two bars indicate carbon feedbacks equatorward/poleward of 80° equivalent ice-edge latitude for the experiment in the top panel of Fig. 10. The right two bars show the carbon feedbacks for the Warm equilibrium experiment (green bar shows the feedback around year 3000 when millennial oscillations occur, see Fig. 11; red bar is a replication of the carbon feedback in the Warm state, see Fig. 8). The error bars stand for one standard deviation. The feedback parameters are derived from the average on the 10th to 90th percentiles of the 9801 samples.

At the boundary of the Warm and Cold states when the ice edge is equatorward of 80° , the carbon feedback is about $+0.15 \text{ W m}^{-2} \text{ K}^{-1}$, just slightly larger than that in the Cold state; it then becomes significantly larger ($+0.52 \text{ W m}^{-2} \text{ K}^{-1}$) during the disappearance of the polar sea ice. As the millennial oscillation in the carbon cycle initiates, the carbon feedback peaks at approximately $+0.9 \text{ W m}^{-2} \text{ K}^{-1}$, estimated with the kernel at year 3000 in Fig. 11. And finally, it gradually gets weaker ($+0.35 \text{ W m}^{-2} \text{ K}^{-1}$) close to the equilibrium.

APPENDIX C

Total carbon inventory

The model's total carbon inventory C_{total} (in units of moles) can be decomposed as:

$$C_{total} = C_{atm} + \text{DIC} + C_{org} \quad (\text{C1})$$

where C_{atm} is the atmospheric carbon reservoir, and DIC and C_{org} (organic carbon) together comprise the oceanic carbon reservoir. C_{atm} is proportional to atmospheric $p\text{CO}_2$:

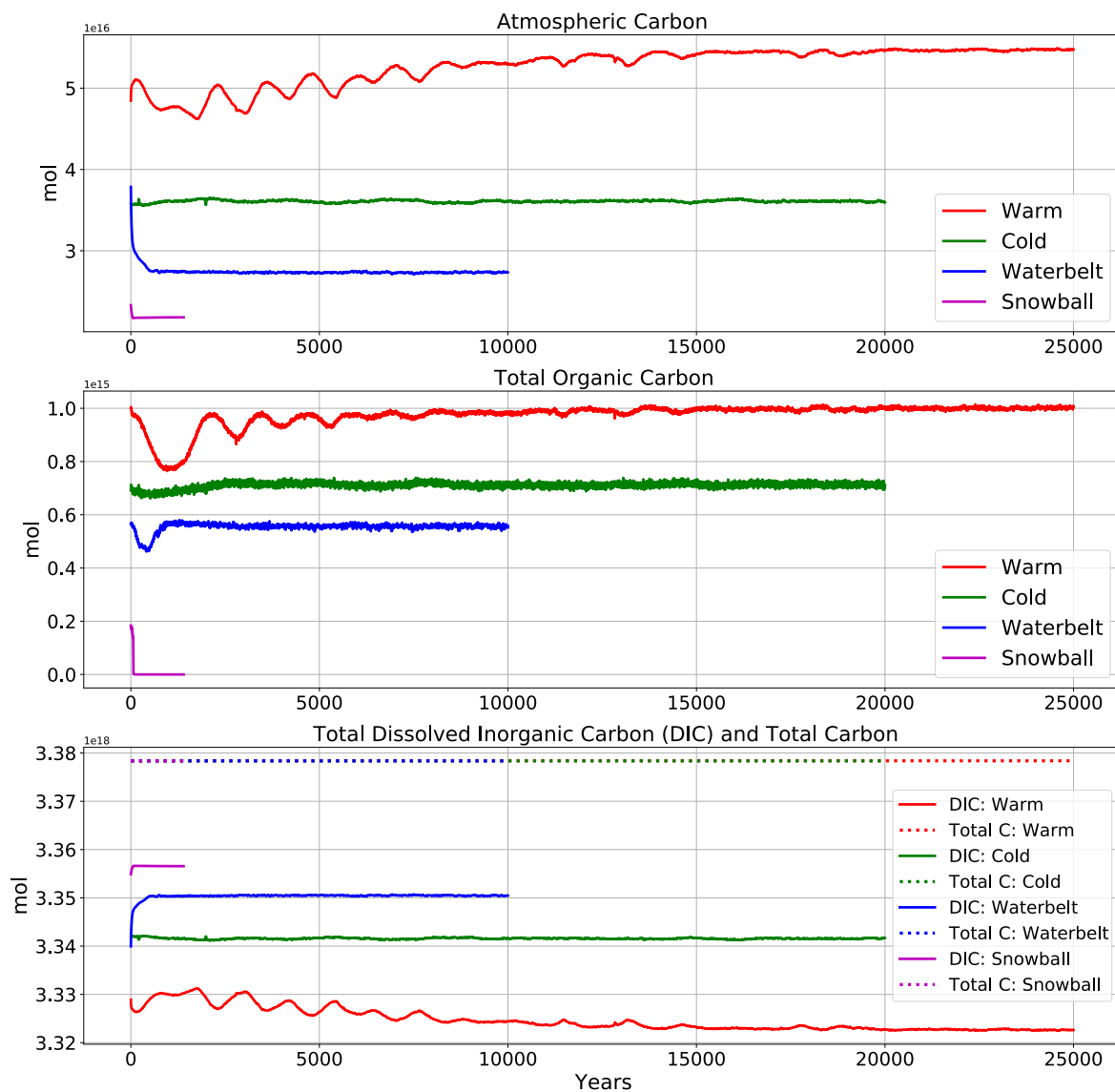
$$C_{atm} = p\text{CO}_2 \times 10^{-6} \times M \quad (\text{C2})$$

with $p\text{CO}_2$ in ppm, and M is the number of moles of air (about 1.77×10^{20} mol). C_{org} is computed from

$$C_{org} = R_{cp} \times \text{DOP} \quad (\text{C3})$$

where $R_{cp} = 117$ is the prescribed Redfield ratio, and DOP is the dissolved organic phosphorus.

The atmospheric and oceanic carbon inventory for four stable climatic equilibria is described in Fig. C1. C_{org} is at least one order of magnitude smaller than C_{atm} and three orders of magnitude less than DIC, which is the largest carbon reservoir regardless of climate state.



711 FIG. C1. Timeseries of (top) atmospheric carbon reservoir, (middle) total organic carbon, and (bottom) total
 712 dissolved inorganic carbon (DIC) and total carbon reservoir for four stable climatic equilibria.

References

- Abe-Ouchi, A., F. Saito, K. Kawamura, M. E. Raymo, J. Okuno, K. Takahashi, and H. Blatter, 2013: Insolation-driven 100,000-year glacial cycles and hysteresis of ice-sheet volume. *nature*, **500 (7461)**, 190–193.
- Adcroft, A., and J.-M. Campin, 2004: Rescaled height coordinates for accurate representation of free-surface flows in ocean circulation models. *Ocean Modelling*, **7 (3-4)**, 269–284.
- Andrews, T., and Coauthors, 2018: Accounting for changing temperature patterns increases historical estimates of climate sensitivity. *Geophys. Res. Lett.*, **45**, <https://doi.org/10.1029/2018GL078887>.
- Arnscheidt, C. W., and D. H. Rothman, 2020: Routes to global glaciation. *Proceedings of the Royal Society A*, **476 (2239)**, 20200 303.
- Brunetti, M., J. Kasparian, and C. V  rard, 2019: Co-existing climate attractors in a coupled aquaplanet. *Climate Dynamics*, **53 (9)**, 6293–6308.
- Budyko, M. I., 1969: The effect of solar radiation variations on the climate of the earth. *tellus*, **21 (5)**, 611–619.
- Dutkiewicz, S., M. J. Follows, P. Heimbach, and J. Marshall, 2006: Controls on ocean productivity and air-sea carbon flux: An adjoint model sensitivity study. *Geophys. Res. Lett.*, **33 (L02603)**, <https://doi.org/10.1029/2005GL024987>.
- Enderton, D., and J. Marshall, 2009: Explorations of atmosphere-ocean-ice climates on an aquaplanet and their meridional energy transports. *J. Atmos. Sci.*, **66**, 1593–1611.
- Feldl, N., and G. H. Roe, 2013: Four perspectives on climate feedbacks. *Geophys. Res. Lett.*, **40**, <https://doi.org/10.1002/grl.50711>.
- Ferreira, D., J. Marshall, T. Ito, and D. McGee, 2018: Linking glacial-interglacial states to multiple equilibria of climate. *Geophysical Research Letters*, **45 (17)**, 9160–9170.
- Ferreira, D., J. Marshall, and B. E. J. Rose, 2011: Climate determinism revisited: Multiple equilibria in a complex climate model. *Journal of Climate*, **24 (4)**, 992–1012.

- Gent, P. R., and J. C. McWilliams, 1990: Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, **20** (1), 150–155.
- Klinger, B. A., J. Marshall, and U. Send, 1996: Representation of convective plumes by vertical adjustment. *Journal of Geophysical Research: Oceans*, **101** (C8), 18 175–18 182.
- Lisiecki, L. E., 2010: Links between eccentricity forcing and the 100,000-year glacial cycle. *Nature geoscience*, **3** (5), 349–352.
- Marshall, J., A. Adcroft, C. Hill, L. Perelman, and C. Heisey, 1997a: A finite-volume, incompressible navier stokes model for studies of the ocean on parallel computers. *Journal of Geophysical Research: Oceans*, **102** (C3), 5753–5766.
- Marshall, J., C. Hill, L. Perelman, and A. Adcroft, 1997b: Hydrostatic, quasi-hydrostatic, and nonhydrostatic ocean modeling. *Journal of Geophysical Research: Oceans*, **102** (C3), 5733–5752.
- Marzocchi, A., and M. F. Jansen, 2019: Global cooling linked to increased glacial carbon storage via changes in antarctic sea ice. *Nature Geoscience*, **12** (12), 1001–1005.
- Molteni, F., 2003: Atmospheric simulations using a gcm with simplified physical parametrizations. i: Model climatology and variability in multi-decadal experiments. *Climate Dynamics*, **20** (2), 175–191.
- North, G. R., 1984: The small ice cap instability in diffusive climate models. *Journal of Atmospheric Sciences*, **41** (23), 3390–3395.
- North, G. R., 1990: Multiple solutions in energy balance climate models. *Global and Planetary Change*, **2** (3-4), 225–235.
- Paillard, D., 1998: The timing of pleistocene glaciations from a simple multiple-state climate model. *Nature*, **391** (6665), 378–381.
- Paillard, D., 2015: Quaternary glaciations: from observations to theories. *Quaternary Science Reviews*, **107**, 11–24.
- PALAEOSSENS Project Members, 2012: Making sense of palaeoclimate sensitivity. *Nature*, **491**, 683–691, <https://doi.org/10.1038/nature11574>.

766 Past Interglacials Working Group of PAGES, 2016: Interglacials of the last 800,000 years. *Reviews*
767 *of Geophysics*, **54** (1), 162–219.

768 Petit, J.-R., and Coauthors, 1999: Climate and atmospheric history of the past 420,000 years from
769 the Vostok ice core, Antarctica. *Nature*, **399** (6735), 429–436.

770 Popp, M., H. Schmidt, and J. Marotzke, 2016: Transition to a moist greenhouse with CO₂ and solar
771 forcing. *Nature communications*, **7** (1), 1–10.

772 Redi, M. H., 1982: Oceanic isopycnal mixing by coordinate rotation. *Journal of Physical Oceanog-*
773 *raphy*, **12** (10), 1154–1158.

774 Rial, J. A., J. Oh, and E. Reischmann, 2013: Synchronization of the climate system to eccentricity
775 forcing and the 100,000-year problem. *Nature Geoscience*, **6** (4), 289–293.

776 Rose, B. E. J., 2015: Stable “waterbelt” climates controlled by tropical ocean heat transport: A
777 nonlinear coupled climate mechanism of relevance to snowball Earth. *Journal of Geophysical*
778 *Research: Atmospheres*, **120** (4), 1404–1423.

779 Rose, B. E. J., D. Ferreira, and J. Marshall, 2013: The role of oceans and sea ice in abrupt
780 transitions between multiple climate states. *Journal of climate*, **26** (9), 2862–2879.

781 Rose, B. E. J., and J. Marshall, 2009: Ocean heat transport, sea ice, and multiple climate states:
782 Insights from energy balance models. *Journal of the atmospheric sciences*, **66** (9), 2828–2843.

783 Saltzman, B., 2001: *Dynamical paleoclimatology: generalized theory of global climate change*.
784 Elsevier.

785 Sellers, W. D., 1969: A global climatic model based on the energy balance of the earth-atmosphere
786 system. *Journal of Applied Meteorology and Climatology*, **8** (3), 392–400.

787 Sigman, D. M., and E. A. Boyle, 2000: Glacial/interglacial variations in atmospheric carbon
788 dioxide. *Nature*, **407** (6806), 859–869.

789 Sigman, D. M., M. P. Hain, and G. H. Haug, 2010: The polar ocean and glacial cycles in
790 atmospheric CO₂ concentration. *Nature*, **466** (7302), 47–55.

791 Soden, B. J., I. M. Held, R. Colman, K. M. Shell, J. T. Kiehl, and C. A. Shields, 2008: Quantifying
792 climate feedbacks using radiative kernels. *Journal of Climate*, **21** (14), 3504–3520.

- 793 Stein, K., A. Timmermann, E. Y. Kwon, and T. Friedrich, 2020: Timing and magnitude of southern
794 ocean sea ice/carbon cycle feedbacks. *Proceedings of the National Academy of Sciences*, **117** (9),
795 4498–4504.
- 796 Weiss, R., 1974: Carbon dioxide in water and seawater: the solubility of a non-ideal gas. *Marine*
797 *chemistry*, **2** (3), 203–215.
- 798 Westerhold, T., and Coauthors, 2020: An astronomically dated record of earth’s climate and its
799 predictability over the last 66 million years. *Science*, **369** (6509), 1383–1387.
- 800 Willeit, M., A. Ganopolski, R. Calov, and V. Brovkin, 2019: Mid-pleistocene transition in glacial
801 cycles explained by declining co2 and regolith removal. *Science Advances*, **5** (4), eaav7337.
- 802 Williams, R. G., and M. J. Follows, 2011: *Ocean dynamics and the carbon cycle: Principles and*
803 *mechanisms*. Cambridge University Press.
- 804 Winton, M., 2000: A reformulated three-layer sea ice model. *Journal of atmospheric and oceanic*
805 *technology*, **17** (4), 525–531.
- 806 Zachos, J., M. Pagani, L. Sloan, E. Thomas, and K. Billups, 2001: Trends, rhythms, and aberrations
807 in global climate 65 ma to present. *science*, **292** (5517), 686–693.