The varying Earth's radiative feedback connected to the ocean energy uptake: a theoretical perspective from conceptual frameworks

Diego Jiménez de la Cuesta[1,1,1] and Diego Jiménez de la Cuesta[1]

[1]Max-Planck-Institut für Meteorologie

December 7, 2022

Abstract

When quadrupling the atmospheric CO$_{2}$ concentration in relation to pre-industrial levels, most global climate models show an initially strong net radiative feedback that significantly reduces the energy imbalance during the first two decades after the quadrupling. Afterwards, the net radiative feedback weakens, needing more surface warming than before to reduce the remaining energy imbalance. Such weakening radiative feedback has its origin in the tropical oceanic stratiform cloud cover, linked to an evolving spatial warming pattern. In the classical linearized energy balance framework, such variation is represented by an additional term in the planetary budget equation. This additional term is usually interpreted as an ad-hoc emulation of the cloud feedback change, leaving unexplained the relationship between this term and the spatial warming pattern. I use a simple non-linearized energy balance framework to justify that there is a physical interpretation of this term: the evolution of the spatial pattern of warming is explained by changes in the ocean's circulation and energy uptake. Therefore, the global effective thermal capacity of the system also changes, leading to the additional term. In reality, the clouds respond to what occurs in the ocean, changing their radiative effect. In the equation, the term is now a concrete representation of the ocean's role. Additionally, I derive for the first time an explicit mathematical expression of the net radiative feedback and its temporal evolution in the linearized energy balance framework. This mathematical expression supports the new proposed interpretation. As a corollary, it justifies the twenty-year time scale used to study the variation of the net radiative feedback.

The varying Earth's radiative feedback connected to the ocean energy uptake: a theoretical perspective from conceptual frameworks

Diego Jiménez-de-la-Cuesta

Max-Planck-Institut für Meteorologie, Hamburg, Deutschland

Corresponding author: Diego Jiménez-de-la-Cuesta, diego.jimenez@mpimet.mpg.de

ABSTRACT: When quadrupling the atmospheric CO_2 concentration in relation to pre-industrial levels, most global climate models show an initially strong net radiative feedback that significantly reduces the energy imbalance during the first two decades after the quadrupling. Afterwards, the net radiative feedback weakens, needing more surface warming than before to reduce the remaining energy imbalance. Such weakening radiative feedback has its origin in the tropical oceanic stratiform cloud cover, linked to an evolving spatial warming pattern. In the classical linearized energy balance framework, such variation is represented by an additional term in the planetary budget equation. This additional term is usually interpreted as an ad-hoc emulation of the cloud feedback change, leaving unexplained the relationship between this term and the spatial warming pattern. I use a simple non-linearized energy balance framework to justify that there is a physical interpretation of this term: the evolution of the spatial pattern of warming is explained by changes in the ocean's circulation and energy uptake. Therefore, the global effective thermal capacity of the system also changes, leading to the additional term. In reality, the clouds respond to what occurs in the ocean, changing their radiative effect. In the equation, the term is now a concrete representation of the ocean's role. Additionally, I derive for the first time an explicit mathematical expression of the net radiative feedback and its temporal evolution in the linearized energy balance framework. This mathematical expression supports the new proposed interpretation. As a corollary, it justifies the twenty-year time scale used to study the variation of the net radiative feedback.

SIGNIFICANCE STATEMENT: Linearized energy balance models have helped the study of Earth's radiative response. However, the present linear models are at the edge of usefulness to get more insights. In this work, I justify that part of the non-linearity in the radiative response can be explained without peculiar atmospheric radiative feedback mechanisms or a non-linearity in the radiative response. Instead, the concept of an evolving thermal capacity recovers the ocean's role in redistributing the energy, changing the spatial warming pattern, and, finally, altering the atmospheric feedback mechanisms. This work also justifies the timescales used in the field for studying the variation of the net radiative feedback.

1. Introduction

The principle of conservation of energy has provided an important tool to study Earth's climate (e.g., Fourier 1827; Arrhenius 1896; Callendar 1938; Budyko 1969; Hansen et al. 1985; Senior and Mitchell 2000; Gregory et al. 2004; Hansen et al. 2010). At the top of the atmosphere (TOA), the incoming radiative flux should balance the outgoing radiative flux, leading to a zero net change of the Earth's energy content (E). If we perturb the radiative balance, the Earth system will change its energy content: this is the radiative forcing (F). Consequently, the surface temperature (T_{u}) will also change, reducing the imbalance. Other variables that define the state of the Earth system also adjust after a surface temperature change, leading to variations in the planetary albedo or the outgoing long-wave radiation, and further altering the TOA net radiative flux. These are the radiative feedback mechanisms that generate the radiative response (R) to the forcing. The balance just described can be summarized in an equation

$$\dot{E} = F + R, \tag{1}$$

where one usually considers that $\dot{E}$ is equivalent to the change in the TOA net radiative flux N. This quantity is also called the TOA net radiative imbalance. A radiative feedback mechanism is negative if it reduces the radiative imbalance. Present Earth's climate has a negative net radiative feedback. Therefore, the radiative response stabilizes the system under forcing at the expense of surface temperature changes: the climate sensitivity. Thus, the more negative the net radiative feedback is, the smaller the surface temperature change is. We can visualize how negative is the net radiative feedback with a $NT-$diagram (Gregory et al. 2004): a plot of N versus ΔT_{u} (Figure

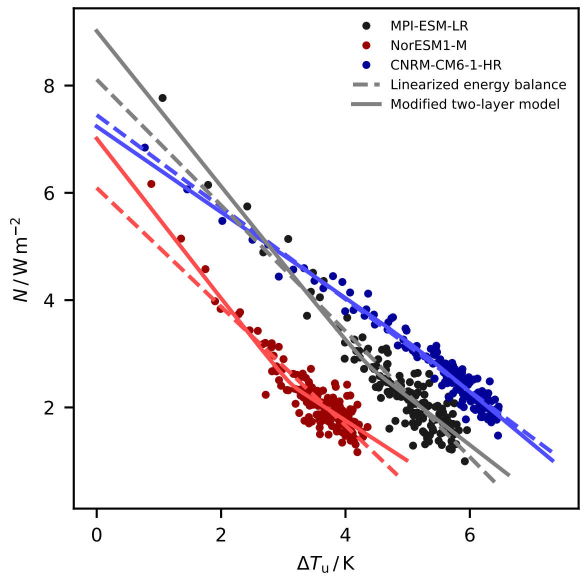

FIG. 1. $NT-$diagram for three GCMs forced with a quadrupling of the atmospheric CO_2 concentration. Dots, annually- and globally-averaged TOA radiative imbalance plotted versus the surface temperature change in relation to the pre-industrial control state. Dashed lines, a linear regression estimate for the relationship between N and ΔT_u. Continuous lines, fit using the modified two-layer model. The model in red presents a large variation in the net radiative feedback, as shown by the strong curvature of the relationship between N and ΔT_u. The net radiative feedback weakens as the system evolves. The model in grey shows a slight curvature. The model shown in blue has a reversed curvature, which means that the net radiative feedback strengthens as the system evolves.

1). The slope of the diagram is the magnitude of the net radiative feedback. The problem is giving R a functional form in terms of variables that describe the system.

Several studies have used equation (1) together with $NT-$diagrams for successfully studying the radiative response and the equilibrium climate sensitivity to CO_2 forcing (ECS) as shown in global climate models (GCMs) and the historical record (e.g., Senior and Mitchell 2000; Gregory et al. 2002, 2004; Andrews et al. 2012; Otto et al. 2013; Armour et al. 2013; Armour 2017). Given the quasi-linearity found in the $NT-$diagrams of GCMs forced with a quadrupling of the pre-industrial atmospheric CO_2 and assuming that R is only a function of T_u, most of these studies implicitly used a Taylor series of R truncated at its first-order term (popularized by Gregory et al. 2004). Consequently, they assumed that (1) the climate state used as the basis for the Taylor series is in balance, and (2) the changes in T_u due to the CO_2 forcing are small enough to neglect higher-order terms of the series. The result is

$$\dot{E} = N \approx F + \left.\frac{dR}{dT_\mathrm{u}}\right|_{T_\mathrm{u}=T_\mathrm{u}^*} \Delta T_\mathrm{u}, \tag{2}$$

where T_u^* is the surface temperature in the reference climate state, and ΔT_u are the anomalies around this reference state. The evaluated derivative is usually called the climate feedback parameter λ, representing an approximation to the magnitude of the net radiative feedback and leading to the more clean equation

$$N \approx F + \lambda \Delta T_\mathrm{u}. \tag{3}$$

Under these strong assumptions, one obtains λ and F estimates from the $NT-$diagrams or observations. Afterwards, using equation (3) one then estimates ECS. This estimate is important in GCMs, as models usually are not run to the equilibrium. However, the linearity assumptions break: in most GCMs, the net radiative feedback becomes less negative as the system warms in timescales of around twenty years. Thus, the ECS is underestimated when using such linearization (Rugenstein and Armour 2021). More importantly, this variation indicates two options: (a) the non-linearity in $R(T_\mathrm{u})$ is important and one should take more terms of the Taylor series, and (b) state variables other than T_u are also important for calculating R.

Some authors extended the framework of equation (3) to accommodate this effect (Held et al. 2010; Winton et al. 2010; Geoffroy et al. 2013b,a). First they introduced two layers: a) the upper layer that includes the atmosphere and the ocean's mixed layer, and b) the deep ocean's layer. Therefore, the state variables are now the surface (T_u) and the deep-ocean (T_d) temperatures. These two layers greatly differ on thermal capacities, introducing two timescales: the fast upper layer and the slow deep layer. They connected both layers with the deep ocean's energy uptake (H), which should depend on $T_\mathrm{u} - T_\mathrm{d}$. However, they also introduced a perturbed energy uptake in the upper layer H' to account for the change in the radiative response

$$\begin{cases} N_\mathrm{u} \approx F + \lambda \Delta T_\mathrm{u} - H' \\ N_\mathrm{d} \approx \qquad\qquad H \end{cases} \tag{4}$$

$$N = N_\mathrm{u} + N_\mathrm{d} \approx F + \lambda \Delta T_\mathrm{u} - (H' - H), \tag{5}$$

where the term $H' - H$ translates the concept of the varying feedback to a problem of variation of the deep ocean's energy uptake. Equations (4) and the corresponding planetary budget correctly represent a varying climate feedback parameter. However, some interpreted $H' - H$ as an additional radiative feedback mechanism from equation (5). Nonetheless, this perspective presents the new term $H' - H$ as devoid of any physical meaning, leading to energy conservation issues and, apparently, rendering the conceptual framework as flawed.

Observations suggest that the net radiative feedback changes in response to an evolving spatial pattern of surface warming (Zhou et al. 2016; Mauritsen 2016; Ceppi and Gregory 2017). The pattern alters the atmospheric stability in decadal timescales, modifying the tropical stratiform clouds' contribution to the short-wave radiative response. In the early decades after the forcing in GCMs, the surface mildly warms in subsidence regions, whereas the deep convection warms the free troposphere. More warming aloft than below enhances the boundary-layer inversion, leading to more stratiform cloud cover and reflected short-wave radiation. After the first decades, there is more warming below than aloft, leading to a weaker inversion, less stratiform cloud cover, and less reflected short-wave radiation. This mechanism suggests that the varying net radiative feedback originates from a process that depends on more than surface warming. Furthermore, several modeling studies found that warming in specific regions leads to a more negative net feedback than when applying warming in other regions (Dong et al. 2019).

Inspired by earlier views on the term $H' - H$ as a perturbed thermal capacity, I show why this term cannot be seen as a peculiar atmospheric radiative feedback mechanism but as a changing thermal capacity. The evolving warming pattern is consistent with a changing oceanic circulation that redistributes the energy, gradually changing the surface temperature and, as a result, the radiative feedback mechanisms. The global effect is as if the thermal capacity of the system changes. First, I show the consistency of the idea by using a non-linear version of equation (1). Second, I put in context this result within the linearized framework of equations (4), finding an equation for the varying planetary thermal capacity. Third, I derive for the first time a mathematical expression for the magnitude of the net radiative feedback in $NT-$diagrams, using the explicit solutions of the linear ordinary differential equations (4) in terms of their normal modes. I find that the variation of the net radiative feedback depends on the ratio of the change in the energy content between the upper and deep layers, in a similar way as the varying planetary thermal capacity. This fact further shows that a changing effective thermal capacity explains better the variation of the net feedback, even in the case of equation (4). As a corollary, I show that the twenty-year timescale for evaluating the pattern effect can be justified by the expression I have derived.

2. Theory

a. Non-linear framework

If E is the Earth's energy content, then its change $N = \dot{E}$ should equal the difference between the TOA incoming and outgoing radiative fluxes. Let us write the incoming TOA radiative flux in terms of the solar incoming radiative flux $S := S(t)$, the planetary albedo α, and the net radiative flux coming from other natural or anthropogenic sources $G := G(t)$. We approximate the outgoing radiative flux as that of a grey-body of emissivity ϵ at the emission temperature $T_{\mathrm{e}} = fT_{\mathrm{u}}$, where f is the lapse-rate scaling factor that relates surface temperature T_{u} to T_{e}. With these elements, the planetary energy budget is

$$N = (1 - \alpha)S + G - \epsilon\sigma(fT_{\mathrm{u}})^4, \tag{6}$$

where N, S and G have units of $\mathrm{W\,m^{-2}}$, T_{u} units are K, and α, ϵ, and f are non-dimensional functions of the variables that describe the system. The planetary albedo depends on the cloud types and cover and the cryosphere extent. Thus, the planetary albedo can depend on the surface temperature and the cloud liquid water content (q_{cw}) or, $\alpha := \alpha(T_{\mathrm{u}}, q_{\mathrm{cw}}, \ldots)$. In the case of the emissivity and lapse-rate scaling factor, the relevant quantity should be the amount of water vapor (q_{v}), additionally to T_{u} and q_{cw}. Therefore, $\epsilon := \epsilon(T_{\mathrm{u}}, q_{\mathrm{v}}, q_{\mathrm{cw}}, \ldots)$ and $f := f(T_{\mathrm{u}}, q_{\mathrm{v}}, q_{\mathrm{cw}}, \ldots)$. The atmospheric radiative feedback mechanisms are contained in α, ϵ, and f. As the state variables evolve, α, ϵ, f change and, consequently, the TOA net radiative flux.

The interpretation of $H' - H$ in equation (5) as an atmospheric radiative feedback is completely ad-hoc in the context of equation (6). If we included $H' - H$ in α, ϵ, or f, another hidden atmospheric state variable would enter the definition of α, ϵ, or f. Directly claiming for regional temperature features in the surface temperature as the hidden variable is not an option since the model is globally averaged. Therefore, one runs out of options to assign a definite physical meaning to $H' - H$ in terms of radiative feedback mechanisms.

The original idea behind $H' - H$ is that the effect of the evolving spatial pattern of warming is connected to a change in the deep ocean's energy uptake. In other words, one temporarily is storing much more energy than expected in the deep ocean, allowing the surface to warm less. As time passes, this larger-than-expected energy uptake is not possible anymore due to changes in the ocean circulation, leading to a different surface warming distribution, which is characteristic of the new ocean state. Therefore, a regional differential warming produced by a new ocean circulation state has a global effect. Consequently, $H' - H$ is an expression of the change in the ocean energy distribution and can be expressed as a change in the planetary thermal capacity of equation (6), mapping a horizontal spatial pattern of warming to a change of the ocean's distribution of energy along the vertical direction. This planetary thermal capacity is the effective thermal capacity associated with the ocean circulation.

Precisely, the planetary thermal capacity is present in the energy content: $E = CT_{\mathrm{u}}$, where C has units of $\mathrm{J\,m^{-2}\,K^{-1}}$. If C is constant, then $\dot{E} = C\dot{T}_{\mathrm{u}} = N$. Defining $N := C\dot{T}_{\mathrm{u}}$ and introducing the varying C results in $\dot{E} = C\dot{T}_{\mathrm{u}} + \dot{C}T_{\mathrm{u}} = N + \dot{C}T_{\mathrm{u}}$. Thus, the planetary energy budget has the new form

$$N = (1 - \alpha)S + G - \epsilon\sigma(fT_{\mathrm{u}})^4 - \dot{C}T_{\mathrm{u}}, \tag{7}$$

consequently, $H' - H$ in equation (5) perfectly fits as a linearization of the last term in equation (7)

$$\dot{C}T_{\mathrm{u}} \sim H' - H$$

Therefore, in this perspective, the perturbed ocean energy uptake is not an exceptional atmospheric radiative feedback, has a definite physical interpretation that does not violate the conservation of energy, and connects the spatial pattern of warming with a changing ocean circulation.

b. The modified linearized two-layer model

I will now use the modified linearized two-layer model to derive an explicit mathematical expression for the net radiative feedback. With this mathematical expression, I find that the traces of the relationship of the pattern effect with ocean circulation are present even in this linearized energy budget.

The following equations define the modified linearized two-layer model (Geoffroy et al. 2013a)

$$\begin{cases} C_{\mathrm{u}}\dfrac{d\,\Delta T_{\mathrm{u}}}{dt} = F+\lambda\Delta T_{\mathrm{u}}-H' \\ C_{\mathrm{d}}\dfrac{d\,\Delta T_{\mathrm{d}}}{dt} = \qquad\qquad H \end{cases} \tag{8}$$

where the first equation corresponds to the upper-layer budget and the second equation to the deep layer. The climate feedback parameter λ has units of $\mathrm{W\,m^{-2}\,K^{-1}}$. H is the ocean energy uptake approximated by $H \approx \gamma(\Delta T_{\mathrm{u}} - \Delta T_{\mathrm{d}})$, where γ is the rate of the deep-ocean energy uptake in $\mathrm{W\,m^{-2}\,K^{-1}}$. H' is the perturbed energy uptake such that $H' = \hat{\varepsilon}H$, where $\hat{\varepsilon}$ is the non-dimensional efficacy of the deep-ocean energy uptake: a measure of the pattern effect. Geoffroy et al. (2013a) consider $\hat{\varepsilon}$ constant. C_{u} and C_{d} are respectively the (fixed) thermal capacities of the upper and deep layers in $\mathrm{J\,m^{-2}\,K^{-1}}$. All these parameters in equations (8) are valid in a neighborhood of the reference climate state $(T_{\mathrm{u}}^*, T_{\mathrm{d}}^*)$, for their values are the ones taken about this state. ΔT_{u} and ΔT_{d} are the temperature anomalies referred to $(T_{\mathrm{u}}^*, T_{\mathrm{d}}^*)$.

For easing the algebraic manipulations, it is better to write equations (8) in the following fashion

$$\begin{cases} \dfrac{d\,\Delta T_{\mathrm{u}}}{dt} = F'+\lambda'\Delta T_{\mathrm{u}}-\hat{\varepsilon}\gamma'(\Delta T_{\mathrm{u}} - \Delta T_{\mathrm{d}}) \\ \dfrac{d\,\Delta T_{\mathrm{d}}}{dt} = \qquad\qquad\qquad \gamma_d'(\Delta T_{\mathrm{u}} - \Delta T_{\mathrm{d}}) \end{cases} \tag{9}$$

where $F' := F/C_{\mathrm{u}}$ with units of $\mathrm{K\,s^{-1}}$ and, $\lambda' := \lambda/C_{\mathrm{u}}$, $\gamma' := \gamma/C_{\mathrm{u}}$ and $\gamma_d' := \gamma/C_{\mathrm{d}}$ with units of $\mathrm{s^{-1}}$. Equations (9) are a system of linear ordinary differential equations (Geoffroy et al. 2013a; Rohrschneider et al. 2019). Although the solutions are standard and widely discussed in other articles (e.g. Geoffroy et al. 2013a; Rohrschneider et al. 2019), here I will use the normal mode approach. The solutions written in terms of the normal modes are more elegant and ease the algebraic transformations. In the following, I summarize the relevant facts, leaving the full mathematical discussion in the appendix A of this article.

The homogeneous ($F' \equiv 0$) version of the system (9) has two distinct eigenvalues $\mu_\pm := (\hat{\lambda} \pm \kappa)/2$, where $\hat{\lambda} := \lambda' - \hat{\varepsilon}\gamma' - \gamma_d'$ and $\kappa^2 := \hat{\lambda}^2 + 4\lambda'\gamma_d'$. These eigenvalues provide two distinct eigenvectors, forming a basis in which the full system (9) is uncoupled and, therefore, has a straight-forward solution. The eigensolutions $\Delta T_\pm$ are the solutions associated with each eigenvalue. Afterwards, one can return to the original representation, finding that ΔT_u and ΔT_d are linear combinations of $\Delta T_\pm$. These linear combinations are the normal modes: the symmetric mode $\Delta T_\mathrm{s} := \Delta T_+ + \Delta T_-$ and the antisymmetric mode $\Delta T_\mathrm{a} := \Delta T_+ - \Delta T_-$. The main result of this process is that

$$\begin{cases} \Delta T_\mathrm{u} = \Delta T_\mathrm{s} \\ \Delta T_\mathrm{d} = -\dfrac{\hat{\lambda} + 2\gamma'_d}{2\hat{\varepsilon}\gamma'}\Delta T_\mathrm{s} + \dfrac{\kappa}{2\hat{\varepsilon}\gamma'}\Delta T_\mathrm{a} \end{cases} \tag{10}$$

c. Planetary thermal capacity in the modified linearized two-layer model

Let us define $\Delta E = E - E^*$, where $E^* = C^* T_\mathrm{u}^*$ and $C = C^* + \Delta C$. C^* is the planetary thermal capacity at the reference climate state $(T_\mathrm{u}^*, T_\mathrm{d}^*)$. Additionally, we postulate that the total change in the planetary energy only comes from F and the original R

$$\frac{d\,\Delta E}{dt} = \dot{E} \approx F + \lambda \Delta T_\mathrm{u}. \tag{11}$$

Summing both equations of system (8), expanding, and using the relationship (10) we obtain

$$\begin{aligned} C_\mathrm{u}\frac{d\,\Delta T_\mathrm{u}}{dt} + C_\mathrm{d}\frac{d\,\Delta T_\mathrm{d}}{dt} &= \dot{E} - (H' - H) \;\therefore \\ \dot{E} &= C_\mathrm{u}\frac{d\,\Delta T_\mathrm{u}}{dt} + H + (H' - H) \\ C^*\frac{d\,\Delta T_\mathrm{u}}{dt} + \Delta C\frac{d\,\Delta T_\mathrm{u}}{dt} + \frac{d\,\Delta C}{dt}T_\mathrm{u} &= C_\mathrm{u}\frac{d\,\Delta T_\mathrm{u}}{dt} + \hat{\varepsilon}H \\ C^* + \Delta C + \frac{d\,\Delta C}{dt}\frac{T_\mathrm{u}}{\frac{d\,\Delta T_\mathrm{u}}{dt}} &= C_\mathrm{u} + \hat{\varepsilon}C_\mathrm{d}\frac{\frac{d\,\Delta T_\mathrm{d}}{dt}}{\frac{d\,\Delta T_\mathrm{u}}{dt}} \\ C^* + \Delta C + \frac{d\,\Delta C}{dt}\frac{T_\mathrm{u}}{\frac{d\,\Delta T_\mathrm{u}}{dt}} &= \left(C_\mathrm{u} - \frac{\hat{\lambda} + 2\gamma'_d}{2\gamma'}C_\mathrm{d}\right) + \frac{\kappa}{2\gamma'}C_\mathrm{d}\frac{\frac{d\,\Delta T_\mathrm{a}}{dt}}{\frac{d\,\Delta T_\mathrm{u}}{dt}}. \end{aligned} \tag{12}$$

From expression (12), as C^* is constant by definition, it should be equal to the quantity inside the parenthesis. Therefore, we can rewrite previous equation in two parts

$$C^* = C_\mathrm{u} - \frac{\hat{\lambda} + 2\gamma'_d}{2\gamma'} C_\mathrm{d}, \tag{13}$$

$$\frac{d\,\Delta C}{dt} + \frac{\frac{d\,\Delta T_\mathrm{u}}{dt}}{T_\mathrm{u}} \Delta C = \frac{\kappa}{2\gamma'} C_\mathrm{d} \frac{\frac{d\,\Delta T_\mathrm{a}}{dt}}{T_\mathrm{u}}. \tag{14}$$

Equation (13) tells us that the basic planetary thermal capacity depends on the initial state of the system. However, Equation (14) provides a more interesting information: the planetary thermal capacity varies regardless of the pattern effect. This fact is reasonable as the initial difference in the thermal capacities of the layers sets the basic distribution of the energy between layers. However, when the pattern effect is active, the relationship between the upper- and deep-layer temperatures changes, per Equations (10). In reality, this change means a different vertical distribution of energy in the ocean coming from a different ocean circulation and, therefore, a different surface warming pattern.

d. Net radiative feedback expression

I now write $\dot{N}$, the total derivative of the imbalance $N_\mathrm{u} + N_\mathrm{d}$, in terms of the normal modes, and divide by the time derivative of ΔT_u to get an explicit mathematical expression for the magnitude of the net radiative feedback and its evolution. I reorder the terms to write the expression as a multiple of the climate feedback parameter λ. In the factor, I separate the radiative forcing ($\mathcal{F}_\mathrm{for}$), radiative response ($\mathcal{F}_\mathrm{res}$), and pattern effect ($\mathcal{F}_\mathrm{pat}$) components of the magnitude

$$\begin{aligned} \lambda_t = \frac{\dot{N}}{\frac{d\,\Delta T_\mathrm{u}}{dt}} &= \left(\mathcal{F}_\mathrm{for} + \mathcal{F}_\mathrm{res} + \mathcal{F}_\mathrm{pat}\right)\lambda \\ &= \left[\mathcal{F}_\mathrm{for} + \mathcal{F}_\mathrm{res} + \frac{\hat{\varepsilon} - 1}{2\hat{\varepsilon}}\left(\mathcal{F}_\mathrm{pat,\,stat} - \mathcal{F}_\mathrm{pat,\,dyn}\right)\right]\lambda. \end{aligned} \tag{15}$$

The $\mathcal{F}_{\text{pat}}$ has two components: the static ($\mathcal{F}_{\text{pat, stat}}$) and dynamical ($\mathcal{F}_{\text{pat, dyn}}$). Each term has the following expression

$$\mathcal{F}_{\text{for}} = -\frac{1}{|\lambda|}\frac{\dot{F}}{\frac{d\,\Delta T_{\text{s}}}{dt}}, \tag{16}$$

$$\mathcal{F}_{\text{res}} = \frac{\hat{\varepsilon}+1}{2\hat{\varepsilon}}, \tag{17}$$

$$\mathcal{F}_{\text{pat, stat}} = C_{\text{u}}\frac{\gamma}{|\lambda|}\left(\frac{\hat{\varepsilon}}{C_{\text{u}}}+\frac{1}{C_{\text{d}}}\right), \tag{18}$$

$$\mathcal{F}_{\text{pat, dyn}} = C_{\text{u}}\frac{\kappa}{|\lambda|}\frac{\frac{d\,\Delta T_{\text{a}}}{dt}}{\frac{d\,\Delta T_{\text{s}}}{dt}}. \tag{19}$$

These expressions (15) – (19) are general for any kind of forcing. One just needs the solutions in terms of normal modes to use them. Let us analyze each term.

- The forcing component (16) simply compares the evolution of F with the evolution of the surface temperature change, given that $\Delta T_{\text{u}} = \Delta T_{\text{s}}$ (first equation of system (10)). This component only contributes if the forcing is time-varying.

- The response component (17) is constant and will only give a correction to the original λ if $\hat{\varepsilon} \neq 1$.

- The pattern effect component is only active if $\hat{\varepsilon} \neq 1$. In case it is active, we have the contributions of the static and dynamical terms.

    1. The static term (18) has three factors. One of them is a sum of the inverse of the thermal capacities of the system. This arrangement is similar to the inverse of the total capacitance of electric capacitors in series. Therefore, it can be interpreted as the effect of the initial state of the ocean energy distribution as discussed for equation (7).

    2. The dynamical term (19) has the ratio of the time derivatives of ΔT_{s} and ΔT_{a}, explicitly relating this term to the expression of the time-dependent planetary thermal capacity in Equation (14).

One should recall that $\hat{\varepsilon} = 1$ means that there is no effect of the energy redistribution due to ocean circulation changes on the surface temperature: no pattern effect. In that case, the only components that contribute to equation (15) are $\mathcal{F}_{\text{for}}$ and $\mathcal{F}_{\text{res}}$. It does not mean that $\mathcal{F}_{\text{pat, stat}}$ and $\mathcal{F}_{\text{pat, dyn}}$ are zero, but that their effects on the net radiative feedback are absent. If this situation had been possible in reality, ocean circulation and ocean energy distribution would have been decoupled from the spatial pattern of warming.

3. Results

a. The explicit slope of the NT−diagram when abruptly changing the atmospheric CO_2

In the `abrupt-4xCO2` experiments, the variation of the net radiative feedback was detected as a curvature in the NT−diagram. I obtain for the first time a concrete expression of the net radiative feedback in those experiments, using equation (15) and the normal-mode solutions for constant radiative forcing. The solutions provide the following form for the components (16) – (19)

$$\mathcal{F}_{\text{for}} = 0, \tag{20}$$

$$\mathcal{F}_{\text{res}} = \frac{\hat{\varepsilon} + 1}{2\hat{\varepsilon}}, \tag{21}$$

$$\mathcal{F}_{\text{pat, stat}} = C_{\text{u}} \frac{\gamma}{|\lambda|} \left(\frac{\hat{\varepsilon}}{C_{\text{u}}} + \frac{1}{C_{\text{d}}} \right), \tag{22}$$

$$\mathcal{F}_{\text{pat, dyn}} = C_{\text{u}} \frac{\kappa}{|\lambda|} \tanh \left[\frac{\kappa}{2}(t - t_0) + \operatorname{arctanh}(Z) \right], \tag{23}$$

$$Z = \frac{\hat{\lambda} + 2\gamma'_d}{\kappa} < 0. \tag{24}$$

One can notice that the time-dependent ratio in equation (19) takes a very elegant and simple form, even though the complexity of the mathematical expressions of the normal-mode solutions (appendix A).

The time-evolving part of equation (23) is an hyperbolic tangent. A plain hyperbolic tangent, $\tanh(t)$, is a monotonically increasing s-shaped or sigmoidal curve, and its possible values are between -1 and 1, crossing zero at $t = 0$. The extreme values -1 and 1 are asymptotes. Leaving out the term $\operatorname{arctanh}(Z)$, our function is similar to $\tanh[(\kappa/2)(t - t_0)]$. This function still has -1 and 1 as asymptotes but crosses zero at $t = t_0$. Depending on the value of $\kappa > 0$, the evolution between asymptotes would be faster. If κ were very large, the function would resemble a step function. The smaller the κ, the gentle the change of the function between asymptotes. Therefore $\kappa/2$ is a scaling factor. We conclude the analysis by adding $\operatorname{arctanh}(Z)$. This term shifts the argument of the hyperbolic tangent. If we evaluate $\tanh[(\kappa/2)(t-t_0)+\operatorname{arctanh}(Z)]$ at t_0, we obtain $\tanh(\operatorname{arctanh}(Z)) = Z < 0$. Therefore, the zero crossing is not anymore at t_0 but at a posterior time and the value of the function at t_0 is negative. I call time of sign reversal (t_{rev}) to the new time where the function becomes zero. This time is

$$t_{\mathrm{rev}} = t_0 + \frac{2}{\kappa} \operatorname{arctanh}|Z| \,. \tag{25}$$

Therefore, $\mathcal{F}_{\mathrm{pat,\,dyn}} < 0$ for $t_0 < t < t_{\mathrm{rev}}$ and non-negative otherwise.

Since in $t_0 < t < t_{\mathrm{rev}}$ $\mathcal{F}_{\mathrm{pat,\,dyn}}$ is negative, the dynamical component strengthens the net radiative feedback, as $\mathcal{F}_{\mathrm{pat}}$ will be larger than without the dynamical component. Nonetheless, the net radiative feedback still becomes less negative as time evolves. In contrast, for $t > t_{\mathrm{rev}}$ $\mathcal{F}_{\mathrm{pat,\,dyn}} \geq 0$, the dynamical component now contributes to weaken even more the feedback. This means that the time of sign reversal is a new timescale in the system. Before t_{rev}, the dynamical component dampens the weakening of the net radiative feedback. However, after t_{rev}, the dynamical component promotes the weakening. This fact leads to the notable curvature of the $NT-$diagrams and is close associated with the varying planetary thermal capacity.

b. Numerical estimates of the time of sign reversal in models

Following the method shown by Geoffroy et al. (2013a), I calculate the thermal, circulation and radiative parameters of the modified linearized two-layer model for a selection of 52 models of the phases 5 and 6 of the climate model inter-comparison project (CMIP). The ensemble means are in table 1. Using equation (15) and the estimated parameters, the theoretical change in the net radiative feedback $\Delta\lambda_t = \lambda_t(150\,\mathrm{yr}) - \lambda_t(1\,\mathrm{yr})$ is calculated. It is compared with the difference in the slopes obtained from the regressions of N on T from the first twenty years, and from the years 21 to 150. Figure 2 shows that the theoretical expression simulates correctly the change in the net radiative feedback ($r = 0.93$).

Ensemble	F / W m^{-2}	C / W yr m^{-2} K^{-1}		λ / W m^{-2} K^{-1}	γ / W m^{-2} K^{-1}	$\hat{\varepsilon}$ / 1	t_{rev} / yr
		C_{u}	C_{d}				
CMIP5	7.52	8.53	105.17	−1.21	0.68	1.26	18.53
CMIP6	7.48	8.06	95.88	−1.02	0.66	1.30	18.31

TABLE 1. CMIP5 and CMIP6 ensemble averages of the thermal and radiative parameters of the modified linearized two-layer model and estimates of the sign reversal timescale t_{rev}.

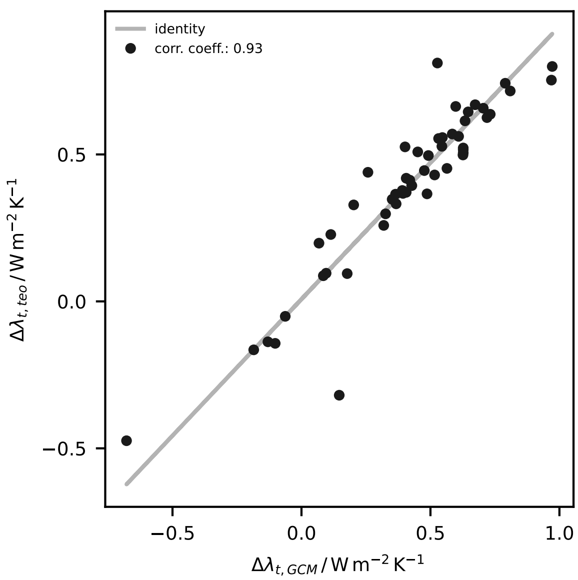

FIG. 2. Comparison between the theoretical change in the net radiative feedback and the corresponding from GCMs. Grey line, the 1-1 line. Black dots, theoretical estimate based in the estimated parameters of the modified linearized two-layer model versus the change estimated using regression from the NT-diagrams.

Given that t_{rev} provides a new timescale, it probably serves as a justification for how we calculate the change in the net radiative feedback: the twenty-year timescale used in this study or, e.g., Ceppi and Gregory (2017). The ensemble means for t_{rev} are consistent: around 18 years for the sign reversal in either ensemble (Table 1): after 18 years, the $\mathcal{F}_{\mathrm{pat,\,dyn}}$ term contributes to further the weakening of the net radiative feedback. In Figure 3, we can see the distribution of t_{rev} in the CMIP ensembles. The median is around 18 years and the total range is between 9 and 27 years (from 12 to 25 is the 5-95 percentile range). Thus, the twenty-year timescale for studying the net radiative feedback variation has a theoretical support.

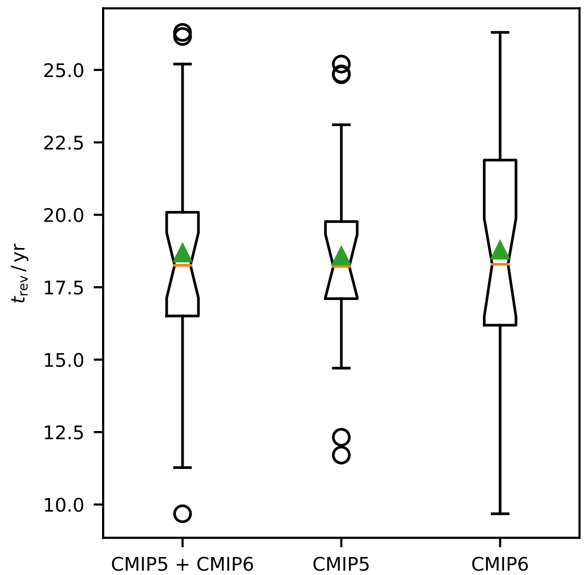

FIG. 3. Time of sign reversal in the CMIP ensembles. Each box represents the inter-quartile range of the data. The orange line is the median and the green triangle shows the mean. The notches on the boxes show the 95 percent confidence interval of the median. The whiskers are at a distance of 1.5 times the inter-quartile range from the first and third quartile.

We can have a look at the diversity of behaviors in the CMIP ensembles. In Figure 4, I show all the models' theoretical evolution of the net radiative feedback. The highlighted models are the ones shown in Figure 1, which shows a model with a strong pattern effect (red), one with a mild pattern effect (grey), and one with a reversed pattern effect (blue). The CMIP5 ensemble has less spread in the starting radiative feedback as well as in the late feedback. The CMIP6 case is more diverse and the late feedbacks are in general more weaker than in the CMIP5 case. Since the amplitude, time of sign reversal and scaling of the hyperbolic tangent of equation (23) depend on the estimates of C_{u}, C_{d}, λ, γ near the starting state, this can explain this diversity in the CMIP ensembles. Additionally, one can look here graphically that the time of sign reversal is more or less constrained in both ensembles, as the mid-point between the early and late feedbacks is attained near to year 20.

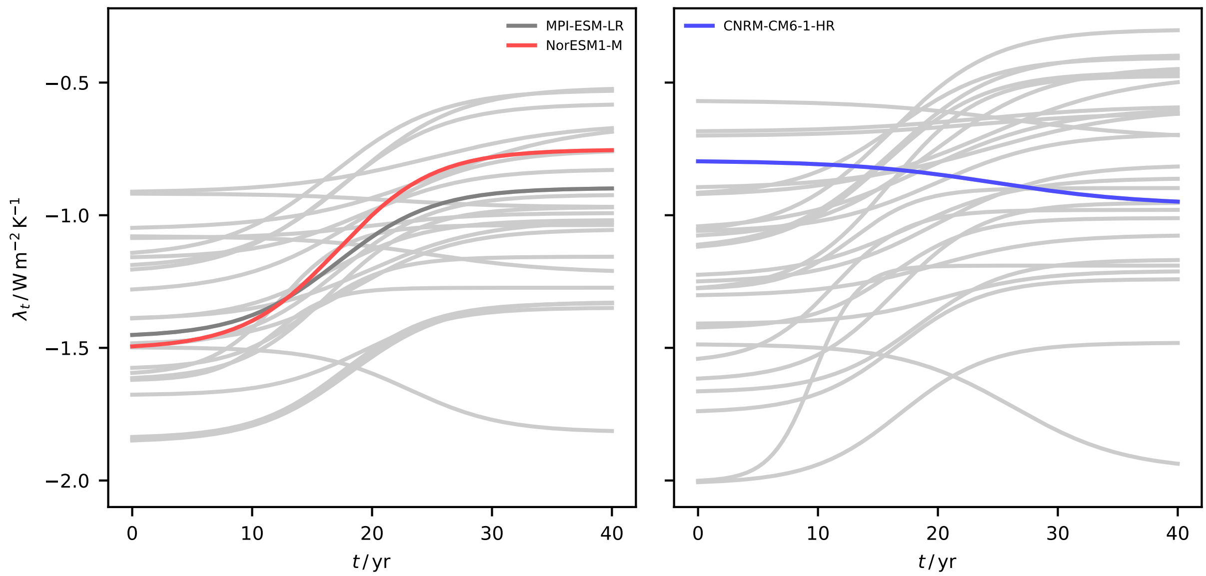

FIG. 4. Theoretical evolution of the radiative feedback. Left: CMIP5 ensemble. Right: CMIP6 ensemble

4. Analysis and Discussion

Winton et al. (2010) have already proposed that an efficacy in the deep-ocean energy uptake would be equivalent to changing the thermal capacity of the deep-ocean layer, as Geoffroy et al. (2013a) also noted. The initial discussion of the non-linear planetary energy budget (Equation 7) and the expression for the dynamical planetary thermal capacity in the linearized framework (Equation 14) show how natural is the concept of a varying planetary thermal capacity, even without pattern effect. When there is a pattern effect, then the relationship between surface and deep-ocean temperatures changes, and the planetary thermal capacity evolves in a different manner.

When looking at the expression of λ_t (Equation 15), the time-varying term $\mathcal{F}_{\text{pat, dyn}}$ (Equation 19) has the same time varying term as in the dynamical planetary thermal capacity (Equation 14). This fact directly connects the varying net radiative feedback to the dynamical planetary thermal capacity. The influence of the $\mathcal{F}_{\text{pat, dyn}}$ term only appears when the pattern effect is active. In contrast, the classical interpretation $H' - H$ as a peculiar radiative feedback mechanism led to inconsistencies: the more serious was about energy conservation. It also left unexplained the origin of the warming pattern and how a spatial pattern could explain a global effect. The dynamical thermal capacity interpretation closes the energy inconsistencies and connects naturally spatial warming patterns to energy distributions in layers, marking a possible course towards understanding the warming pattern and why it is different between contemporary GCMs and observations (e.g., Wills et al. 2022).

One obstacle to understand the thermal-capacity interpretation is the picture of Earth's thermal capacity as that of all the matter in the Earth system's components. However, the planetary thermal capacity is a global representation of how the ocean circulation distributes the energy in the system, as one can interpret from Equations (13) and (14). After forcing, the ocean circulation changes, altering the ocean stratification and which parts of the ocean are active at storing energy. This fact impacts the energy distribution and the efficacy of storing more energy. Consequently, this evolving energy distribution sets the evolving surface warming pattern. In recent studies, the role of circulation changes in the ocean energy uptake and its effect on the regional warming pattern has been uncovered in complex models. The southern ocean temperatures are connected with the tropics (Newsom et al. 2020; Lin et al. 2021; Hu et al. 2021). In the southern ocean the complex interactions between deep-water formation and upwelling shape the long-term ocean overturning circulation and influence the Pacific basin shallower layers (Talley 2015), closely relating the SST in both regions. Thus, in some way, the role of the ocean was always there, even in the conceptual models.

As I showed above, apart from the linearization, the two-layer model (8) preserves the traces of the energy redistribution process. The energy is redistributed between the upper and deep layers. One can then ask to formulate the problem in terms of a two-region model for mimicking the spatial warming pattern directly. Rohrschneider et al. (2019) demonstrated that two-region models are mathematically equivalent to two-layer models, further supporting the discussion on how $H' - H$ represents a physical reality and is not only a mathematical artifice to provide further usability to a broken framework. However, the two-region model assumes different net radiative feedback for the regions, again leaving the origin of this difference unexplained. However, this new interpretation of the modified two-layer model provides the missing link.

In the results, I show that the estimates of the thermal, circulation, and radiative parameters can have a substantial effect on the evolution of the net radiative feedback. In light of the discussion, particularly the thermal capacities and the rate of deep-ocean energy uptake γ represent an initial energy distribution about the reference state (T_u^*, T_d^*). This energy distribution evolves differently, depending on the magnitude of the deep-ocean energy uptake efficacy $\hat{\varepsilon}$. This parameter represents the magnitude of the coupling of the energy distribution and the surface temperature. Thus, it should be related directly to physical quantities, e.g., the ocean stratification in the regions of upwelling of deep-water formation. In consequence, GCMs will show diverse behaviors for the variation of the net radiative feedback as their initial energy uptake and the rate at which it changes with ocean circulation widely varies (Kiehl 2007). Perhaps, this diversity in GCMs is part of the reason why GCMs cannot fully reproduce the observed warming pattern (e.g., Wills et al. 2022). This fact is worrying, given that our climate change projections can be biased low.

Although the framework of the two-layer model (8) and the equation (15) can provide estimates for the variation of the net radiative feedback and theoretically justify the timescale used to study this variation, one should remember that this simple model has limitations. The three main limitations are

1. The assumed radiative response R neglects the dependency on atmospheric state variables other than the surface temperature,
2. The linearization neglects the existence of complex emergent behaviors such as tipping points,
3. The unknown relationship between the surface temperature spatial pattern and the distribution of the energy content in the ocean, limiting our capability to provide good estimates for $\hat{\varepsilon}$ and estimate the error of considering $\hat{\varepsilon}$ constant.

Therefore, some details in the theoretical evolution of the radiative feedback (Figure 4) can be different between the complex models and nature. Nonetheless, these limitations should be the starting point to find what are the actual relationships between the evolving spatial warming pattern and the energy distribution in the ocean. For that end, one should use observations, Earth system model output, new experiments tailored to isolate mechanisms, and other simplified models for specific mechanisms. This process will help to put in context equation (14) and possibly reveal that $\hat{\varepsilon}$ is not constant, relaxing the constraint imposed and providing further information on its physics. Such uses of the conceptual frameworks have been useful in related problems, and there are recent advances (e.g. Datseris et al. 2022). Thus, checking when the assumptions of the conceptual models break and understanding the reasons advance us towards a better conceptual understanding of the climate system.

In my analysis of the two-layer model, the dependence of the variation of the net radiative feedback with the strength of forcing (Senior and Mitchell 2000; Meraner et al. 2013; Rohrschneider et al. 2019) is missing. However, such dependence should come from the values of $\hat{\varepsilon}$, λ, γ, and the thermal capacities under a particular forcing and, probably, non-linearities. We should always remember that the thermal capacities, λ, and γ are only approximations of the actual quantities in the neighborhood of the starting states. Therefore, we need a consistent theory on how the different types and magnitudes of forcing modify (a) the coupling between ocean energy distribution and surface temperature, (b) the atmospheric radiative feedback mechanisms, and (c) the rate of energy uptake. Such a theory should describe the Earth system not in the tiny details or as an aggregate of separate disciplines but as an integrated system. The idea can be better expressed as the difference between describing a tree as an aggregate of cells of different types with different functions; and describing the whole tree in terms of certain characteristic variables. In the best case, the needed theory for the climate is incomplete. However, having such a basic conceptual theory of climate will help us better interpret complex model results, find more hidden relationships between important variables and, possibly, reduce the uncertainty in observational estimates of climate sensitivity.

5. Conclusions

In the context of the modified linearized two-layer model (8), I show that variation of net radiative feedback due to the evolving spatial pattern of warming cannot be directly explained by a hidden variable in the atmospheric radiative feedback mechanisms. To show this fact, I discuss how this view is utterly artificial in the context of a global non-linear version of the energy budget (7) and provide an alternative interpretation. This alternative perspective proposes that the planetary thermal capacity used in equations (7) and (8) change, because the ocean circulation changes the distribution of energy in the ocean, the efficacy of the energy uptake and the sea surface temperature. This new perspective is consistent with recent studies (Newsom et al. 2020; Hu et al. 2021; Lin et al. 2021). I also present for the first time an explicit mathematical expression of the net radiative feedback in the two-layer model (8) and particularize it for a case of constant forcing. From the analysis, I

1. confirm that the the time-varying term (Equation 23) mimics the redistribution of energy by comparing the energy in the upper and deep layers, varying the net radiative feedback,

2. connect this time-varying term with the dynamical planetary thermal capacity (Equation 14),
3. uncover another timescale t_{rev}: the timescale for the change in the net radiative feedback in the GCM-based `abrupt-4xCO2` experiments.

Using the parameters estimated in the same way as Geoffroy et al. (2013a) did, I find that t_{rev} is around 18 years in CMIP models, providing theoretical support to the 20-year standard timescale used to study the variations in the net radiative feedback in `abrupt-4xCO2` experiments. These results should motivate us to continue developing a conceptual characterization of the Earth system. This conceptual theory is necessary to interpret our complex models better, find hidden relationships between variables, or reduce the uncertainty in observationally-informed estimates of future climate change.

Acknowledgments. I thank Hauke Schmidt, Jiawei Bao, Christopher Hedemann and Moritz Günther for reading and correcting the manuscript and for the lively discussions on how to present these highly theoretical ideas. I also thank the three anonymous reviewers that kindly revised the manuscript and enriched the presentation of the materials with their suggestions.

Data availability statement. The theoretical considerations are fully described in the appendices of this article. The software to reproduce the numerical results can be found in Jiménez-de-la-Cuesta (2022a). All the post-processed CMIP data is deposited in Jiménez-de-la-Cuesta (2022b).

APPENDIX A

Mathematical analysis of the modified two-layer model

In Classical Mechanics, a very coarse thinking would be reducing the field to the task of solving the equation $\dot{\mathbf{p}} = \mathbf{F}$ for any force term, either analytically or numerically. Going further leads to conservation principles and formulations of Classical Mechanics that provide more information without actually obtaining solutions, if that is possible at all. In this appendix, reduced to the scale of a simplified framework, I show that by delving deep into the mathematics of a system of linear ordinary differential equations, the structure of the solutions and its physical interpretation, one can obtain a new view on an old problem.

The appendix is written in an exhaustive way and I leave few things without development. The cases in which I do not show some algebraic step is because the necessary step has been already done or is very simple. For simplicity ΔT_u and ΔT_d are always rewritten as T_u and T_d for the two-layer model.

Matrix form of the equations

The equations of two-layer model Geoffroy et al. (2013a) are

$$\begin{aligned} N_\mathrm{u} &= C_\mathrm{u}\dot{T}_\mathrm{u} = F+\lambda T_\mathrm{u}-\hat{\varepsilon}\gamma(T_\mathrm{u}-T_\mathrm{d}) \\ N_\mathrm{d} &= C_\mathrm{d}\dot{T}_\mathrm{d} = \qquad\qquad \gamma(T_\mathrm{u}-T_\mathrm{d}) \end{aligned} \tag{A1}$$

and the planetary imbalance is $N = N_\mathrm{u} + N_\mathrm{d}$. I present another form of the equations, where I divide by the thermal capacities.

$$\begin{aligned} \dot{T}_\mathrm{u} &= \tfrac{F}{C_\mathrm{u}}+\tfrac{\lambda}{C_\mathrm{u}}T_\mathrm{u}-\hat{\varepsilon}\tfrac{\gamma}{C_\mathrm{u}}(T_\mathrm{u}-T_\mathrm{d}) \\ \dot{T}_\mathrm{d} &= \qquad\qquad \tfrac{\gamma}{C_\mathrm{d}}(T_\mathrm{u}-T_\mathrm{d}) \end{aligned}$$

If I define $F' := F/C_\mathrm{u}$, $\lambda' := \lambda/C_\mathrm{u}$, $\gamma' := \gamma/C_\mathrm{u}$, $\gamma'_d := \gamma/C_\mathrm{d}$, one can write the equations in a lean way

$$\begin{aligned} \dot{T}_\mathrm{u} &= F'+\lambda' T_\mathrm{u}-\hat{\varepsilon}\gamma'(T_\mathrm{u}-T_\mathrm{d}) \\ \dot{T}_\mathrm{d} &= \qquad\qquad \gamma'_d(T_\mathrm{u}-T_\mathrm{d}) \end{aligned} \tag{A2}$$

I will put the system in matrix form. I define $\mathbf{T} := (T_\mathrm{u}, T_\mathrm{d})$, $\mathbf{F}' := (F', 0)$ and

$$\mathsf{A} := \begin{pmatrix} \lambda' - \hat{\varepsilon}\gamma' & \gamma'_d \\ \hat{\varepsilon}\gamma' & -\gamma'_d \end{pmatrix} \tag{A3}$$

and the system can be written

$$\dot{\mathbf{T}} = \mathbf{F}'+\mathbf{T}\mathsf{A} \tag{A4}$$

which is the representation of the system in the temperature basis.

Eigenvalues and eigenvectors

I want to analyse the normal modes of the system. For that end, I need the eigenvalues of the homogeneous system obtained as the solutions of the characteristic equation

$$(\lambda' - \hat{\varepsilon}\gamma' - \mu)(-\gamma'_d - \mu) - \hat{\varepsilon}\gamma'\gamma'_d = 0 \tag{A5}$$

$$\begin{aligned}
-\lambda'\gamma'_d + \hat{\varepsilon}\gamma'\gamma'_d + \mu\gamma'_d - \lambda'\mu + \hat{\varepsilon}\gamma'\mu + \mu^2 - \hat{\varepsilon}\gamma'\gamma'_d &= 0 \\
-\lambda'\gamma'_d + \mu\gamma'_d - \lambda'\mu + \hat{\varepsilon}\gamma'\mu + \mu^2 &= 0 \\
-\lambda'\gamma'_d - (\lambda' - \hat{\varepsilon}\gamma' - \gamma'_d)\mu + \mu^2 &= 0
\end{aligned}$$

The solutions of equation (A5) are

$$\mu = \frac{(\lambda' - \hat{\varepsilon}\gamma' - \gamma'_d) \pm \left[(\lambda' - \hat{\varepsilon}\gamma' - \gamma'_d)^2 + 4\lambda'\gamma'_d\right]^{1/2}}{2} \tag{A6}$$

and, given that in the Earth $C_\mathrm{u} < C_\mathrm{d}$, one can prove that there are two real and different eigenvalues. One needs to check that the square root term is not complex or zero. This only happens if the sum within the square root is negative or zero

$$\begin{aligned}
(\lambda' - \hat{\varepsilon}\gamma' - \gamma'_d)^2 + 4\lambda'\gamma'_d &\le 0 \\
(\lambda' - \hat{\varepsilon}\gamma')^2 - 2(\lambda' - \hat{\varepsilon}\gamma')\gamma'_d + \gamma_d'^2 + 4\lambda'\gamma'_d &\le 0 \\
\lambda'^2 - 2\lambda'\hat{\varepsilon}\gamma' + (\hat{\varepsilon}\gamma')^2 - 2(\lambda' - \hat{\varepsilon}\gamma')\gamma'_d + \gamma_d'^2 + 4\lambda'\gamma'_d &\le 0 \\
\lambda'^2 - 2\lambda'\hat{\varepsilon}\gamma' + (\hat{\varepsilon}\gamma')^2 - 2\lambda'\gamma'_d + 2\hat{\varepsilon}\gamma'\gamma'_d + \gamma_d'^2 + 4\lambda'\gamma'_d &\le 0 \\
(\lambda'/\gamma'_d)^2 - 2(\lambda'/\gamma'_d)\hat{\varepsilon}(\gamma'/\gamma'_d) + (\hat{\varepsilon}(\gamma'/\gamma'_d))^2 + 2\hat{\varepsilon}(\gamma'/\gamma'_d) + 1 + 2(\lambda'/\gamma'_d) &\le 0 \\
(\lambda'/\gamma'_d)^2 - 2(\lambda'/\gamma'_d)[\hat{\varepsilon}(\gamma'/\gamma'_d) - 1] + (\hat{\varepsilon}(\gamma'/\gamma'_d))^2 + 2\hat{\varepsilon}(\gamma'/\gamma'_d) + 1 &\le 0 \\
(\lambda'/\gamma'_d)^2 - 2(\lambda'/\gamma'_d)[\hat{\varepsilon}(\gamma'/\gamma'_d) - 1] + (\hat{\varepsilon}(\gamma'/\gamma'_d) + 1)^2 &\le 0 \\
(\lambda'/\gamma'_d)^2 + (\hat{\varepsilon}(C_\mathrm{d}/C_\mathrm{u}) + 1)^2 &\le 2(\lambda'/\gamma'_d)[\hat{\varepsilon}(C_\mathrm{d}/C_\mathrm{u}) - 1]
\end{aligned}$$

In the last inequality, the left-hand side is always positive. The right-hand side depends on the sign of the factors. The middle factor is negative since λ' is negative and γ'_d is positive. The third factor is positive provided that $\hat{\varepsilon} > C_\mathrm{u}/C_\mathrm{d}$. Given that $\hat{\varepsilon} \geq 1$ and $C_\mathrm{u} < C_\mathrm{d}$, then the third factor is positive in our case. Then the right-hand side is negative. Thus, we obtained a contradiction by supposing that the square root term was negative or zero. Therefore, the conclusion is that the eigenvalues are two real and distinct numbers. Some CMIP5 models show $\hat{\varepsilon} < 1$ according to Geoffroy et al. (2013a). These also fit here. In the last condition of the above expression we require that $\hat{\varepsilon}(C_\mathrm{d}/C_\mathrm{u}) - 1 > 0$. If $\hat{\varepsilon} \geq C_\mathrm{u}/C_\mathrm{d}$ this is fulfilled. $C_\mathrm{u}/C_\mathrm{d}$ is a small quantity and, in the models that have a lesser than one $\hat{\varepsilon}$, always the $\hat{\varepsilon}$ is larger than this small quantity by an order of magnitude. Thus, what I had said until now and will be said afterwards applies to all cases.

I call the solutions μ_+ and μ_-, depending on the sign of the square root term. Let us rewrite their expression in more lean fashion. I define $\hat{\lambda} := \lambda' - \hat{\varepsilon}\gamma' - \gamma'_d$ and we call κ the square root term. Then, I rewrite the solutions (A6) as

$$\mu_\pm = \frac{\hat{\lambda} \pm \kappa}{2} \tag{A7}$$

Now that I know the eigenvalues, one should get the eigenvectors of the system and solve it easily. The eigenvectors are the generators of the kernel of the operators $\mathbf{A} - \mu_\pm\,\mathrm{id}$. Let us write the diagonal of the matrix $\mathbf{A}$ with the definition of $\hat{\lambda}$

$$\mathbf{A} = \begin{pmatrix} \hat{\lambda} + \gamma'_d & \gamma'_d \\ \hat{\varepsilon}\gamma' & \hat{\lambda} - (\lambda' - \hat{\varepsilon}\gamma') \end{pmatrix}$$

and then the matrices for each eigenvalue have the form

$$\begin{aligned} \mathbf{A} - \mu_\pm\,\mathrm{id} &= \begin{pmatrix} \hat{\lambda} + \gamma'_d - \mu_\pm & \gamma'_d \\ \hat{\varepsilon}\gamma' & \hat{\lambda} - (\lambda' - \hat{\varepsilon}\gamma') - \mu_\pm \end{pmatrix} \\ &= \begin{pmatrix} \mu_\mp + \gamma'_d & \gamma'_d \\ \hat{\varepsilon}\gamma' & \mu_\mp - (\lambda' - \hat{\varepsilon}\gamma') \end{pmatrix} \end{aligned}$$

Since eigenvalues are real and distinct, there should be two linearly-independent eigenvectors, one for each eigenvalue. These vectors should fulfill that $\mathbf{e}_\pm(\mathbf{A} - \mu_\pm\,\mathrm{id}) = 0$. Solving that linear system, I find the eigenvectors in temperature representation

$$\mathbf{e}_{\pm} = \mathbf{e}_{u} - \frac{\mu_{\mp} + \gamma'_d}{\hat{\varepsilon}\gamma'}\mathbf{e}_d \tag{A8}$$

The procedure to get the result is to solve the system of homogeneous linear equations $\mathbf{e}_{\pm}(\mathbf{A} - \mu_{\pm}\,\mathrm{id}) = 0$

$$\begin{cases} (\mu_{\mp} + \gamma'_d)e_{\pm,u} & +\hat{\varepsilon}\gamma' e_{\pm,d} = 0 \\ \gamma'_d e_{\pm,u} & +[\mu_{\mp} - (\lambda' - \hat{\varepsilon}\gamma')]e_{\pm,d} = 0 \end{cases}$$

I solve the first equation for the component $e_{\pm,d}$, and substitute this result on the second equation

$$\begin{aligned} e_{\pm,d} &= -\frac{\mu_{\mp} + \gamma'_d}{\hat{\varepsilon}\gamma'}e_{\pm,u} \longrightarrow \\ \left(\gamma'_d - \frac{[\mu_{\mp} - (\lambda' - \hat{\varepsilon}\gamma')](\mu_{\mp} + \gamma'_d)}{\hat{\varepsilon}\gamma'}\right) e_{\pm,u} &= 0 \\ \frac{\hat{\varepsilon}\gamma'\gamma'_d - [\mu_{\mp} - (\lambda' - \hat{\varepsilon}\gamma')](\mu_{\mp} + \gamma'_d)}{\hat{\varepsilon}\gamma'} e_{\pm,u} &= 0,\ (\hat{\varepsilon}, \gamma' \neq 0)\ \therefore \end{aligned}$$

$$\begin{aligned} \left\{\hat{\varepsilon}\gamma'\gamma'_d - [\mu_{\mp} - (\lambda' - \hat{\varepsilon}\gamma')](\mu_{\mp} + \gamma'_d)\right\} e_{\pm,u} &= 0 \\ \left\{\hat{\varepsilon}\gamma'\gamma'_d + [(\lambda' - \hat{\varepsilon}\gamma') - \mu_{\mp}](\gamma'_d + \mu_{\mp})\right\} e_{\pm,u} &= 0 \\ -\left\{-\hat{\varepsilon}\gamma'\gamma'_d + [(\lambda' - \hat{\varepsilon}\gamma') - \mu_{\mp}](-\gamma'_d - \mu_{\mp})\right\} e_{\pm,u} &= 0 \end{aligned}$$

and in the last expression we have two options: either $e_{\pm,u}$ is zero or the term within curly braces is zero. However, the expression in curly braces is the characteristic equation (A5) and then always vanishes identically. This means that $e_{\pm,u} = \alpha \in \mathbb{R}$ can be chosen arbitrarily. I plug in this result in the expression for $e_{\pm,d}$ and get that

$$\begin{aligned} e_{\pm,u} &= \alpha \\ e_{\pm,d} &= -\frac{\mu_{\mp} + \gamma'_d}{\hat{\varepsilon}\gamma'}\alpha \end{aligned}$$

or as a vector in the temperature basis

$$\mathbf{e}_{\pm} = e_{\pm,u}\mathbf{e}_u + e_{\pm,d}\mathbf{e}_d$$

$$\mathbf{e}_{\pm} = \alpha\mathbf{e}_u - \frac{\mu_{\mp} + \gamma'_d}{\hat{\varepsilon}\gamma'}\alpha\mathbf{e}_d$$

and since α is arbitrary this means we are in front of a subspace of vectors. I choose a basis by selecting $\alpha = 1$.

$$\mathbf{e}_{\pm} = \mathbf{e}_u - \frac{\mu_{\mp} + \gamma'_d}{\hat{\varepsilon}\gamma'}\mathbf{e}_d$$

which is the same as the equation (A8).

Now, I can derive the expressions of the temperature basis vectors in terms of the two eigenvectors. If one solves for e_u in equation (A8)

$$\mathbf{e}_{\pm} + \frac{\mu_{\mp} + \gamma'_d}{\hat{\varepsilon}\gamma'}\mathbf{e}_d = \mathbf{e}_u$$

but we have here two expressions in a condensed way. Therefore,

$$\mathbf{e}_{-} + \frac{\mu_{+} + \gamma'_d}{\hat{\varepsilon}\gamma'}\mathbf{e}_d = \mathbf{e}_{+} + \frac{\mu_{-} + \gamma'_d}{\hat{\varepsilon}\gamma'}\mathbf{e}_d$$

$$\left(\frac{\mu_{+} + \gamma'_d}{\hat{\varepsilon}\gamma'} - \frac{\mu_{-} + \gamma'_d}{\hat{\varepsilon}\gamma'}\right)\mathbf{e}_d = \mathbf{e}_{+} - \mathbf{e}_{-}$$

$$\frac{(\mu_{+} + \gamma'_d) - (\mu_{-} + \gamma'_d)}{\hat{\varepsilon}\gamma'}\mathbf{e}_d = \mathbf{e}_{+} - \mathbf{e}_{-}$$

$$\frac{\mu_{+} - \mu_{-}}{\hat{\varepsilon}\gamma'}\mathbf{e}_d = \mathbf{e}_{+} - \mathbf{e}_{-}$$

$$\mathbf{e}_d = \frac{\hat{\varepsilon}\gamma'}{\mu_{+} - \mu_{-}}(\mathbf{e}_{+} - \mathbf{e}_{-})$$

Thus, I have expressed $\mathbf{e}_d$ in terms of the eigenvectors.

Now, I substitute the last result on one of the expressions for $\mathbf{e}_u$.

$$\mathbf{e}_+ + \frac{\mu_- + \gamma'_d}{\hat{\varepsilon}\gamma'}\mathbf{e}_d = \mathbf{e}_u$$
$$\mathbf{e}_+ + \frac{\mu_- + \gamma'_d}{\hat{\varepsilon}\gamma'}\frac{\hat{\varepsilon}\gamma'}{\mu_+ - \mu_-}(\mathbf{e}_+ - \mathbf{e}_-) = \mathbf{e}_u$$
$$\mathbf{e}_+ + \frac{\mu_- + \gamma'_d}{\mu_+ - \mu_-}(\mathbf{e}_+ - \mathbf{e}_-) = \mathbf{e}_u$$
$$\left(1 + \frac{\mu_- + \gamma'_d}{\mu_+ - \mu_-}\right)\mathbf{e}_+ - \frac{\mu_- + \gamma'_d}{\mu_+ - \mu_-}\mathbf{e}_- = \mathbf{e}_u$$
$$\frac{\mu_+ - \mu_- + \mu_- + \gamma'_d}{\mu_+ - \mu_-}\mathbf{e}_+ - \frac{\mu_- + \gamma'_d}{\mu_+ - \mu_-}\mathbf{e}_- = \mathbf{e}_u$$
$$\frac{\mu_+ + \gamma'_d}{\mu_+ - \mu_-}\mathbf{e}_+ - \frac{\mu_- + \gamma'_d}{\mu_+ - \mu_-}\mathbf{e}_- = \mathbf{e}_u$$

and the temperature basis vectors in the eigenvector representation are

$$\begin{aligned}\mathbf{e}_u &= \frac{\mu_+ + \gamma'_d}{\mu_+ - \mu_-}\mathbf{e}_+ - \frac{\mu_- + \gamma'_d}{\mu_+ - \mu_-}\mathbf{e}_- \\ \mathbf{e}_d &= \frac{\hat{\varepsilon}\gamma'}{\mu_+ - \mu_-}(\mathbf{e}_+ - \mathbf{e}_-)\end{aligned} \tag{A9}$$

Matrix in the eigenvector representation. Solutions

With these results, I can write the matrix $\mathbf{A}$ (A3) in the eigenvector basis and it should be the following diagonal matrix

$$\mathbf{B} = \begin{pmatrix} \mu_+ & 0 \\ 0 & \mu_- \end{pmatrix} \tag{A10}$$

I show how to get to this result. Let subscripts represent rows and superscripts represent columns. I define that latin indices $(i, j, k, \dots)$ have the possible values u, d; and greek indices $(\alpha, \beta, \zeta \dots)$ have possible values $+, -$. Also, repeated indices in expressions mean summation over the set of possible values. With these considerations, equation (A9) is

$$\mathbf{e}_i = \Lambda_i^{\alpha}\mathbf{e}_{\alpha}$$

where the rows of matrix Λ contain the coordinates of each of the vectors of the temperature basis in the eigenvector representation. Analogously, equation (A8) is

$$\mathbf{e}_\alpha = \Theta_\alpha^i \mathbf{e}_i$$

where matrix Θ has in its rows the coordinates the eigenvector basis in the temperature representation. This means that

$$\mathbf{e}_\alpha = \Theta_\alpha^i \mathbf{e}_i = \Theta_\alpha^i \Lambda_i^\beta \mathbf{e}_\beta$$

which is only possible if the matrices Λ and Θ are inverse of each other

$$\mathbf{e}_\alpha = \delta_\alpha^\beta \mathbf{e}_\beta = \mathbf{e}_\alpha$$

Thus, we write $\Theta = \Lambda^{-1}$.

Now, matrix $\mathbf{A}$ is the temperature representation of a linear operator f. If $\mathbf{v} = v^j \mathbf{e}_j$ is a vector in the temperature representation, then the action of the linear operator f should be $f(\mathbf{v}) = f(v^j \mathbf{e}_j) = v^j f(\mathbf{e}_j)$. Then the action of f on a vector expressed in a given basis only depends on the action of the operator on the basis: $f(\mathbf{v}) = f(v^j \mathbf{e}_j) = v^j f(\mathbf{e}_j) = v^j \mathbf{A}_j^k \mathbf{e}_k$. Thus, the matrix $\mathbf{A}$ has in its rows the coordinates in the temperature representation of the action of f over each basis vector. Once one understands what is happening under the hood, what we want is the matrix $\mathbf{B}$, which is the representation of f in the eigenvector basis. Therefore, I begin with the basic relationship in the temperature representation and introduce the change of representation using the alternative representation of equations (A8) and (A9)

$$
\begin{aligned}
f(\mathbf{e}_i) &= \mathbf{A}_i^j \Lambda_j^\zeta \mathbf{e}_\zeta \\
f(\Lambda_i^\alpha \mathbf{e}_\alpha) &= \mathbf{A}_i^j \Lambda_j^\zeta \mathbf{e}_\zeta \\
\Lambda_i^\alpha f(\mathbf{e}_\alpha) &= \mathbf{A}_i^j \Lambda_j^\zeta \mathbf{e}_\zeta \\
(\Lambda^{-1})_\beta^i \Lambda_i^\alpha f(\mathbf{e}_\alpha) &= (\Lambda^{-1})_\beta^i \mathbf{A}_i^j \Lambda_j^\zeta \mathbf{e}_\zeta \\
f(\mathbf{e}_\beta) &= (\Lambda^{-1})_\beta^i \mathbf{A}_i^j \Lambda_j^\zeta \mathbf{e}_\zeta,\ f(\mathbf{e}_\beta) := \mathbf{B}_\beta^\zeta \mathbf{e}_\zeta \\
\mathbf{B}_\beta^\zeta &= (\Lambda^{-1})_\beta^i \mathbf{A}_i^j \Lambda_j^\zeta
\end{aligned}
$$

or in matrix notation $\mathbf{B} = \Lambda^{-1}\mathbf{A}\Lambda$. Then, I multiply the matrices

$$
\Lambda^{-1} = \begin{pmatrix} 1 & -\frac{\mu_- + \gamma_d'}{\hat{\varepsilon}\gamma'} \\ 1 & -\frac{\mu_+ + \gamma_d'}{\hat{\varepsilon}\gamma'} \end{pmatrix},\ \mathbf{A} = \begin{pmatrix} \hat{\lambda} + \gamma_d' & \gamma_d' \\ \hat{\varepsilon}\gamma' & -\gamma_d' \end{pmatrix},\ \Lambda = \begin{pmatrix} \frac{\mu_+ + \gamma_d'}{\mu_+ - \mu_-} & -\frac{\mu_- + \gamma_d'}{\mu_+ - \mu_-} \\ \frac{\hat{\varepsilon}\gamma'}{\mu_+ - \mu_-} & -\frac{\hat{\varepsilon}\gamma'}{\mu_+ - \mu_-} \end{pmatrix}
$$

First, note that $\mu_+ - \mu_- = \kappa$. One also looks at the following quantities that will help in the process: $\mu_+ + \mu_- = \hat{\lambda}$ and $\mu_+\mu_- = \frac{1}{4}(\hat{\lambda}^2 - \kappa^2) = \frac{1}{4}(\hat{\lambda}^2 - \hat{\lambda}^2 - 4\lambda'\gamma_d') = -\lambda'\gamma_d'$. I proceed with the first product, $\Lambda^{-1}\mathbf{A}$.

$$
\begin{aligned}
\Lambda^{-1}\mathbf{A} &= \begin{pmatrix} 1 & -\frac{\mu_- + \gamma_d'}{\hat{\varepsilon}\gamma'} \\ 1 & -\frac{\mu_+ + \gamma_d'}{\hat{\varepsilon}\gamma'} \end{pmatrix} \begin{pmatrix} \hat{\lambda} + \gamma_d' & \gamma_d' \\ \hat{\varepsilon}\gamma' & -\gamma_d' \end{pmatrix} \\
&= \begin{pmatrix} \hat{\lambda} + \gamma_d' - \mu_- - \gamma_d' & \left(1 + \frac{\mu_- + \gamma_d'}{\hat{\varepsilon}\gamma'}\right)\gamma_d' \\ \hat{\lambda} + \gamma_d' - \mu_+ - \gamma_d' & \left(1 + \frac{\mu_+ + \gamma_d'}{\hat{\varepsilon}\gamma'}\right)\gamma_d' \end{pmatrix} \\
&= \begin{pmatrix} \hat{\lambda} - \mu_- & \frac{\hat{\varepsilon}\gamma' + \mu_- + \gamma_d'}{\hat{\varepsilon}\gamma'}\gamma_d' \\ \hat{\lambda} - \mu_+ & \frac{\hat{\varepsilon}\gamma' + \mu_+ + \gamma_d'}{\hat{\varepsilon}\gamma'}\gamma_d' \end{pmatrix} \\
&= \begin{pmatrix} \mu_+ & \frac{\hat{\varepsilon}\gamma' + \mu_- + \gamma_d'}{\hat{\varepsilon}\gamma'}\gamma_d' \\ \mu_- & \frac{\hat{\varepsilon}\gamma' + \mu_+ + \gamma_d'}{\hat{\varepsilon}\gamma'}\gamma_d' \end{pmatrix}
\end{aligned}
$$

and multiply the result by Λ

$$
\begin{aligned}
\Lambda^{-1}\mathbf{A}\Lambda &= \begin{pmatrix} \mu_+ & \frac{\hat{\varepsilon}\gamma'+\mu_-+\gamma'_d}{\hat{\varepsilon}\gamma'}\gamma'_d \\ \mu_- & \frac{\hat{\varepsilon}\gamma'+\mu_++\gamma'_d}{\hat{\varepsilon}\gamma'}\gamma'_d \end{pmatrix}\begin{pmatrix} \frac{\mu_++\gamma'_d}{\mu_+-\mu_-} & -\frac{\mu_-+\gamma'_d}{\mu_+-\mu_-} \\ \frac{\hat{\varepsilon}\gamma'}{\mu_+-\mu_-} & -\frac{\hat{\varepsilon}\gamma'}{\mu_+-\mu_-} \end{pmatrix} \\
&= \frac{1}{\kappa}\begin{pmatrix} \mu_+^2+\mu_+\gamma'_d+\hat{\varepsilon}\gamma'\gamma'_d+\mu_-\gamma'_d+\gamma_d'^2 & -\mu_+\mu_--\mu_+\gamma'_d-\hat{\varepsilon}\gamma'\gamma'_d-\mu_-\gamma'_d-\gamma_d'^2 \\ \mu_-\mu_++\mu_-\gamma'_d+\hat{\varepsilon}\gamma'\gamma'_d+\mu_+\gamma'_d+\gamma_d'^2 & -\mu_-^2-\mu_-\gamma'_d-\hat{\varepsilon}\gamma'\gamma'_d-\mu_+\gamma'_d-\gamma_d'^2 \end{pmatrix} \\
&= \frac{1}{\kappa}\begin{pmatrix} \mu_+^2+(\hat{\lambda}+\hat{\varepsilon}\gamma'+\gamma'_d)\gamma'_d & -\mu_+\mu_--(\hat{\lambda}+\hat{\varepsilon}\gamma'+\gamma'_d)\gamma'_d \\ \mu_-\mu_++(\hat{\lambda}+\hat{\varepsilon}\gamma'+\gamma'_d)\gamma'_d & -\mu_-^2-(\hat{\lambda}+\hat{\varepsilon}\gamma'+\gamma'_d)\gamma'_d \end{pmatrix} \\
&= \frac{1}{\kappa}\begin{pmatrix} \mu_+^2-\mu_+\mu_- & \lambda'\gamma'_d-\lambda'\gamma'_d \\ -\lambda'\gamma'_d+\lambda'\gamma'_d & -\mu_-^2+\mu_+\mu_- \end{pmatrix} = \frac{1}{\kappa}\begin{pmatrix} \mu_+\kappa & 0 \\ 0 & \mu_-\kappa \end{pmatrix} = \begin{pmatrix} \mu_+ & 0 \\ 0 & \mu_- \end{pmatrix}
\end{aligned}
$$

the last line is the result that we wanted to check.

In the eigenvector representation the system (A4) has the following form

$$\dot{\mathbf{T}} = \mathbf{F}'+\mathbf{TB} \tag{A11}$$

and, therefore, is decoupled. Therefore, I can solve each equation separately. I only need to transform the forcing vector to the eigenvector representation.

The equations are

$$\dot{T}_\pm = F'_\pm + \mu_\pm T_\pm$$

and the solutions of a generic initial value problem are

$$T_\pm = \left(T_{\pm,0} + \int_{t_0}^{t} F'_\pm e^{-\mu_\pm(\tau-t_0)}\mathrm{d}\tau\right) e^{\mu_\pm(t-t_0)} \tag{A12}$$

where the initial values in the eigenvector representation in terms of the initial values in the temperature representation are

$$T_{\pm,0} = \pm\frac{1}{\mu_+-\mu_-}\left[(\mu_\pm+\gamma'_d)T_{\mathrm{u},0}+\hat{\varepsilon}\gamma' T_{\mathrm{d},0}\right]$$

the forcing components are

$$F'_{\pm} = \pm\frac{\mu_{\pm} + \gamma'_d}{\mu_+ - \mu_-}F'$$

and the solutions in the temperature representation are

$$\begin{aligned}
T_\mathrm{u} &= T_+ + T_- \\
T_\mathrm{d} &= -\frac{\mu_- + \gamma'_d}{\hat{\varepsilon}\gamma'}T_+ - \frac{\mu_+ + \gamma'_d}{\hat{\varepsilon}\gamma'}T_-
\end{aligned}$$

If I further expand the T_d solution, the form of the solutions is more elegant

$$\begin{aligned}
T_\mathrm{u} &= T_+ + T_- \\
T_\mathrm{d} &= -\frac{\hat{\lambda} + 2\gamma'_d}{2\hat{\varepsilon}\gamma'}(T_+ + T_-) + \frac{\kappa}{2\hat{\varepsilon}\gamma'}(T_+ - T_-)
\end{aligned} \tag{A13}$$

since it shows that the solutions in the temperature space are in a sort of symmetric and antisymmetric combinations of the solutions in the eigenvector representation. These are the normal modes. One thing to note is that the upper temperature is the symmetric mode and the deep temperature is a mixture of symmetric and antisymmetric modes.

I show how I got the solutions (A13). Just expand the T_d equation.

$$\begin{aligned}
T_\mathrm{d} &= -\frac{\mu_- + \gamma'_d}{\hat{\varepsilon}\gamma'}T_+ - \frac{\mu_+ + \gamma'_d}{\hat{\varepsilon}\gamma'}T_- \\
&= -\frac{1}{\hat{\varepsilon}\gamma'}\left[\left(\frac{\hat{\lambda} - \kappa}{2} + \gamma'_d\right)T_+ + \left(\frac{\hat{\lambda} + \kappa}{2} + \gamma'_d\right)T_-\right] \\
&= -\frac{1}{\hat{\varepsilon}\gamma'}\left[\left(\frac{\hat{\lambda} + 2\gamma'_d}{2} - \frac{\kappa}{2}\right)T_+ + \left(\frac{\hat{\lambda} + 2\gamma'_d}{2} + \frac{\kappa}{2}\right)T_-\right] \\
&= -\frac{1}{2\hat{\varepsilon}\gamma'}\left[(\hat{\lambda} + 2\gamma'_d)(T_+ + T_-) - \kappa(T_+ - T_-)\right]
\end{aligned}$$

From now on, I write $T_\mathrm{s} := T_+ + T_-$ and $T_\mathrm{a} := T_+ - T_-$.

Planetary imbalance

Now, I will find an expression for the planetary imbalance in terms of the equations (A13). The mathematical expression that I should expand is $N = N_\mathrm{u} + N_\mathrm{d} = C_\mathrm{u}\dot{T}_\mathrm{u} + C_\mathrm{d}\dot{T}_\mathrm{d}$

$$\begin{aligned}
C_\mathrm{u}\dot{T}_\mathrm{u} &= C_\mathrm{u}\dot{T}_\mathrm{s} \\
C_\mathrm{d}\dot{T}_\mathrm{d} &= -C_\mathrm{d}\frac{\hat{\lambda} + 2\gamma'_d}{2\hat{\varepsilon}\gamma'}\dot{T}_\mathrm{s} + C_\mathrm{d}\frac{\kappa}{2\hat{\varepsilon}\gamma'}\dot{T}_\mathrm{a} \therefore \\
N &= C_\mathrm{u}\dot{T}_\mathrm{s} - C_\mathrm{d}\frac{\hat{\lambda} + 2\gamma'_d}{2\hat{\varepsilon}\gamma'}\dot{T}_\mathrm{s} + C_\mathrm{d}\frac{\kappa}{2\hat{\varepsilon}\gamma'}\dot{T}_\mathrm{a} \\
&= \left(C_\mathrm{u} - C_\mathrm{d}\frac{\hat{\lambda} + 2\gamma'_d}{2\hat{\varepsilon}\gamma'}\right)\dot{T}_\mathrm{s} + C_\mathrm{d}\frac{\kappa}{2\hat{\varepsilon}\gamma'}\dot{T}_\mathrm{a} \\
&= C_\mathrm{s}\dot{T}_\mathrm{s} + C_\mathrm{a}\dot{T}_\mathrm{a}
\end{aligned}$$

Now, $\dot{T}_\pm = F'_\pm + \mu_\pm T_\pm$, then

$$\begin{aligned}
\dot{T}_\mathrm{s} &= \mu_+T_+ + \mu_-T_- + (F'_+ + F'_-) = \mu_+T_+ + (\mu_+ - \kappa)T_- + (F'_+ + F'_-) \\
&= \mu_+T_\mathrm{s} - \kappa T_- + (F'_+ + F'_-) = \mu_+T_\mathrm{s} - \frac{\kappa}{2}(T_\mathrm{s} - T_\mathrm{a}) + (F'_+ + F'_-) \\
&= \frac{\hat{\lambda}}{2}T_\mathrm{s} + \frac{\kappa}{2}T_\mathrm{a} + (F'_+ + F'_-) = \frac{\hat{\lambda}}{2}T_\mathrm{s} + \frac{\kappa}{2}T_\mathrm{a} + F' \\
\dot{T}_\mathrm{a} &= \mu_+T_+ - \mu_-T_- + (F'_+ - F'_-) = \mu_+T_+ - (\mu_+ - \kappa)T_- + (F'_+ - F'_-) \\
&= \mu_+T_\mathrm{a} + \kappa T_- + (F'_+ - F'_-) = \mu_+T_\mathrm{a} + \frac{\kappa}{2}(T_\mathrm{s} - T_\mathrm{a}) + (F'_+ - F'_-) \\
&= \frac{\kappa}{2}T_\mathrm{s} + \frac{\hat{\lambda}}{2}T_\mathrm{a} + (F'_+ - F'_-) = \frac{\kappa}{2}T_\mathrm{s} + \frac{\hat{\lambda}}{2}T_\mathrm{a} + \frac{\hat{\lambda} + 2\gamma'_d}{\kappa}F' \therefore \\
N &= \frac{1}{2}\left(\hat{\lambda}C_\mathrm{s} + \kappa C_\mathrm{a}\right)T_\mathrm{s} + \frac{1}{2}\left(\hat{\lambda}C_\mathrm{a} + \kappa C_\mathrm{s}\right)T_\mathrm{a} + \left(C_\mathrm{s} + C_\mathrm{a}\frac{\hat{\lambda} + 2\gamma'_d}{\kappa}\right)F'
\end{aligned}$$

Further expanding the coefficients

$$\hat{\lambda}C_\mathrm{s} + \kappa C_\mathrm{a} = \hat{\lambda}C_\mathrm{u} - \frac{C_\mathrm{d}}{2\hat{\varepsilon}\gamma'}(\hat{\lambda}^2 + 2\gamma'_d\hat{\lambda} - \kappa^2) = \hat{\lambda}C_\mathrm{u} - \frac{C_\mathrm{d}}{2\hat{\varepsilon}\gamma'}(\hat{\lambda}^2 + 2\gamma'_d\hat{\lambda} - \hat{\lambda}^2 - 4\gamma'_d\lambda')$$
$$= 2\frac{C_\mathrm{u}}{\hat{\varepsilon}}\left(\lambda' + \frac{\hat{\varepsilon}-1}{2}\hat{\lambda}\right)$$
$$\hat{\lambda}C_\mathrm{a} + \kappa C_\mathrm{s} = \kappa C_\mathrm{u} - \frac{C_\mathrm{d}}{2\hat{\varepsilon}\gamma'}(\kappa\hat{\lambda} + 2\gamma'_d\kappa - \kappa\hat{\lambda}) = \kappa C_\mathrm{u} - \frac{C_\mathrm{u}}{\hat{\varepsilon}}\kappa = \kappa\frac{C_\mathrm{u}}{\hat{\varepsilon}}(\hat{\varepsilon} - 1)$$
$$C_\mathrm{s} + C_\mathrm{a}\frac{\hat{\lambda} + 2\gamma'_d}{\kappa} = C_\mathrm{u} - \frac{C_\mathrm{d}}{2\hat{\varepsilon}\gamma'}(\hat{\lambda} + 2\gamma'_d - \hat{\lambda} - 2\gamma'_d) = C_\mathrm{u}$$

then the imbalance is

$$N = \frac{C_\mathrm{u}}{\hat{\varepsilon}}\left[\hat{\varepsilon}F' + \left(\lambda' + \frac{\hat{\varepsilon}-1}{2}\hat{\lambda}\right)T_\mathrm{s} + \kappa\frac{\hat{\varepsilon}-1}{2}T_\mathrm{a}\right] \tag{A14}$$

From here, I derive the slope of a $NT-$diagram. In such a diagram, N is plotted versus T_u. If we naïvely take the partial derivative of equation (A14) with respect to T_u, we will arrive to a constant slope. This is contrary to the evidence that it will change with time. An $NT-$diagram is one projection of the phase space of the system. Then, the $NT-$diagram slope does not only depend on how N varies with T_u. It is a comparison of how the changes of T_u are expressed in changes of N. Then, the slope is the total derivative $\mathrm{d}N/\mathrm{d}T_\mathrm{u}$. By virtue of the chain rule, $\mathrm{d}N/\mathrm{d}T_\mathrm{u} = \dot{N}(\mathrm{d}t/\mathrm{d}T_\mathrm{u})$. In a neighborhood where $T_\mathrm{u}(t)$ is injective, $\mathrm{d}t/\mathrm{d}T_\mathrm{u} = 1/\dot{T}_\mathrm{u}$. Therefore, the slope $\mathrm{d}N/\mathrm{d}T_\mathrm{u}$ is the ratio of two total derivatives: $\dot{N}$ and $\dot{T}_\mathrm{u}$.

We know that $T_\mathrm{u} = T_\mathrm{s}$, then $\dot{T}_\mathrm{u} = \dot{T}_\mathrm{s}$. Therefore, the total derivative of the planetary imbalance is

$$\dot{N} = (\partial_t N) + (\partial_{T_\mathrm{s}}N)\dot{T}_\mathrm{s} + (\partial_{T_\mathrm{a}}N)\dot{T}_\mathrm{a}$$

that is a change depending only on time, a second change depending only on changes of T_s and a third depending on changes of T_a. Therefore, the ratio of total derivative of planetary imbalance and total derivative of T_u is

$$\frac{\dot{N}}{\dot{T}_\mathrm{u}} = (\partial_t N)\frac{1}{\dot{T}_\mathrm{s}} + (\partial_{T_\mathrm{s}}N) + (\partial_{T_\mathrm{a}}N)\frac{\dot{T}_\mathrm{a}}{\dot{T}_\mathrm{s}}$$

As one can see in the above expression, the ratio includes the derivative of the imbalance with respect to T_u but is not the only contribution. One contribution comes from the explicit dependence on time of N and how it compares with the dependency of T_u. The other contribution comes from the antisymmetric mode and how it changes in relation to the symmetric one. From equation (A14), I can write the precise expression of the slope as a factor of λ.

I multiply equation (A14) by λ/λ and reorganise.

$$\begin{aligned}
\frac{\dot N}{\dot T_\mathrm{u}} &= \frac{C_\mathrm{u}}{\hat\varepsilon}\left[\hat\varepsilon\frac{\dot F'}{\dot T_\mathrm{s}} + \left(\lambda' + \frac{\hat\varepsilon-1}{2}\hat\lambda\right) + \kappa\frac{\hat\varepsilon-1}{2}\frac{\dot T_\mathrm{a}}{\dot T_\mathrm{s}}\right]\frac{\lambda}{\lambda}\\
&= \left[\frac{C_\mathrm{u}}{\lambda}\frac{\dot F'}{\dot T_\mathrm{s}} + \left(\frac{\lambda'}{\hat\varepsilon\lambda'} + \frac{\hat\varepsilon-1}{2\hat\varepsilon}\frac{\hat\lambda}{\lambda'}\right) + \frac{\hat\varepsilon-1}{2\hat\varepsilon}\frac{\kappa}{\lambda'}\frac{\dot T_\mathrm{a}}{\dot T_\mathrm{s}}\right]\lambda
\end{aligned}$$

then we will expand the terms to separate the terms that vanish when $\hat\varepsilon = 1$

$$\begin{aligned}
\frac{\dot N}{\dot T_\mathrm{u}} &= \left\{\frac{C_\mathrm{u}}{\lambda}\frac{\dot F'}{\dot T_\mathrm{s}} + \left[\frac{1}{\hat\varepsilon} + \frac{\hat\varepsilon-1}{2\hat\varepsilon}\left(\frac{\lambda' - \hat\varepsilon\gamma' - \gamma'_d}{\lambda'}\right)\right] + \frac{\hat\varepsilon-1}{2\hat\varepsilon}\frac{\kappa}{\lambda'}\frac{\dot T_\mathrm{a}}{\dot T_\mathrm{s}}\right\}\lambda\\
&= \left\{\frac{C_\mathrm{u}}{\lambda}\frac{\dot F'}{\dot T_\mathrm{s}} + \left[\frac{2}{2\hat\varepsilon} + \frac{\hat\varepsilon-1}{2\hat\varepsilon}\left(1 - \hat\varepsilon\frac{\gamma}{\lambda} - \frac{C_\mathrm{u}}{C_\mathrm{d}}\frac{\gamma}{\lambda}\right)\right] + \frac{\hat\varepsilon-1}{2\hat\varepsilon}\frac{C_\mathrm{u}\kappa}{\lambda}\frac{\dot T_\mathrm{a}}{\dot T_\mathrm{s}}\right\}\lambda\\
&= \left[\frac{C_\mathrm{u}}{\lambda}\frac{\dot F'}{\dot T_\mathrm{s}} + \frac{\hat\varepsilon+1}{2\hat\varepsilon} - \frac{\hat\varepsilon-1}{2\hat\varepsilon}\left(\hat\varepsilon + \frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\frac{\gamma}{\lambda} + \frac{\hat\varepsilon-1}{2\hat\varepsilon}\frac{C_\mathrm{u}\kappa}{\lambda}\frac{\dot T_\mathrm{a}}{\dot T_\mathrm{s}}\right]\lambda\\
&= \left[\frac{C_\mathrm{u}}{\lambda}\frac{\dot F'}{\dot T_\mathrm{s}} + \frac{\hat\varepsilon+1}{2\hat\varepsilon} - \frac{\hat\varepsilon-1}{2\hat\varepsilon}\left(\hat\varepsilon + \frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\frac{\gamma}{\lambda} + \frac{\hat\varepsilon-1}{2\hat\varepsilon}\frac{C_\mathrm{u}\kappa}{\lambda}\frac{\dot T_\mathrm{a}}{\dot T_\mathrm{s}}\right]\lambda\\
&= \left\{\frac{C_\mathrm{u}}{\lambda}\frac{\dot F'}{\dot T_\mathrm{s}} + \frac{\hat\varepsilon+1}{2\hat\varepsilon} - \frac{\hat\varepsilon-1}{2\hat\varepsilon\lambda}\left[\left(\hat\varepsilon + \frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\gamma - C_\mathrm{u}\kappa\frac{\dot T_\mathrm{a}}{\dot T_\mathrm{s}}\right]\right\}\lambda\\
&= \left\{\frac{C_\mathrm{u}}{\lambda}\frac{\dot F'}{\dot T_\mathrm{s}} + \frac{\hat\varepsilon+1}{2\hat\varepsilon} - \frac{\hat\varepsilon-1}{2\hat\varepsilon\lambda}C_\mathrm{u}\kappa\left[\left(\hat\varepsilon + \frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\frac{\gamma}{C_\mathrm{u}\kappa} - \frac{\dot T_\mathrm{a}}{\dot T_\mathrm{s}}\right]\right\}\lambda\\
&= \left\{\frac{C_\mathrm{u}}{\lambda}\frac{\dot F'}{\dot T_\mathrm{s}} + \frac{\hat\varepsilon+1}{2\hat\varepsilon} - \frac{\hat\varepsilon-1}{2\hat\varepsilon}\frac{C_\mathrm{u}\kappa}{\lambda}\left[\left(\hat\varepsilon + \frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\frac{\gamma}{C_\mathrm{u}\kappa} - \frac{\dot T_\mathrm{a}}{\dot T_\mathrm{s}}\right]\right\}\lambda\\
&= \left\{-\frac{C_\mathrm{u}}{|\lambda|}\frac{\dot F'}{\dot T_\mathrm{s}} + \frac{\hat\varepsilon+1}{2\hat\varepsilon} + \frac{\hat\varepsilon-1}{2\hat\varepsilon}\frac{C_\mathrm{u}\kappa}{|\lambda|}\left[\left(\hat\varepsilon + \frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\frac{\gamma}{C_\mathrm{u}\kappa} - \frac{\dot T_\mathrm{a}}{\dot T_\mathrm{s}}\right]\right\}\lambda
\end{aligned}$$

$$\frac{\dot N}{\dot T_\mathrm{u}} = \left\{-\frac{C_\mathrm{u}}{|\lambda|}\frac{\dot F'}{\dot T_\mathrm{s}} + \frac{\hat\varepsilon+1}{2\hat\varepsilon}\left(1 + \frac{\hat\varepsilon-1}{\hat\varepsilon+1}\frac{C_\mathrm{u}\kappa}{|\lambda|}\left[\left(\hat\varepsilon + \frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\frac{\gamma}{C_\mathrm{u}\kappa} - \frac{\dot T_\mathrm{a}}{\dot T_\mathrm{s}}\right]\right)\right\}\lambda \tag{A15}$$

The term in square brackets in equation (A15) is the key term that provides a $NT-$diagram with evolving slope when the forcing is constant. The second part of this term provides the temporal evolution, whereas the first part is a constant term that sets the base enhancement of the slope. Interestingly, this first part contains in particular the thermal capacities of the system.

If I rewrite this first part of the square-brackets term, the terms are shown clearly

$$\frac{\dot{N}}{\dot{T}_\mathrm{u}} = \left\{ -\frac{C_\mathrm{u}}{|\lambda|}\frac{\dot{F}'}{\dot{T}_\mathrm{s}} + \frac{\hat{\varepsilon}+1}{2\hat{\varepsilon}} + \frac{\hat{\varepsilon}-1}{2\hat{\varepsilon}}\frac{C_\mathrm{u}\kappa}{|\lambda|}\left[\left(\frac{\hat{\varepsilon}}{C_\mathrm{u}} + \frac{1}{C_\mathrm{d}} \right)\frac{\gamma}{\kappa} - \frac{\dot{T}_\mathrm{a}}{\dot{T}_\mathrm{s}} \right] \right\} \lambda \tag{A16}$$

Now in the first part it is the sum of the inverse of the thermal capacities as if we have an electrical circuit with capacitors in series. Having such a term in the equation for the slope favors the physical interpretation in terms of thermal capacities, instead of variable feedback mechanisms. The time-evolving ratio term in the second part, that represents the dynamics of the atmosphere-ocean coupling, only strengthens this interpretation.

As a corollary, if the forcing is constant and $\hat{\varepsilon} \longrightarrow 1$, then we recover the classical linear dependence of the imbalance on T_u

$$\lim_{\hat{\varepsilon}\to 1} \frac{\dot{N}}{\dot{T}_\mathrm{u}} = \lambda, \; F = \mathrm{const}$$

Symmetric and antisymmetric modes

From equations (A13), we see that the symmetric and antisymmetric modes are the basis for the description of the solutions. Thus, let us give some explicit expression for the symmetric and antisymmetric modes.

From equation (A12) and the equations for the initial values and the forcing, I can write more explicitly the solution

$$
\begin{aligned}
T_\pm &= \left(T_{\pm,0} + \int_{t_0}^{t} F'_\pm e^{-\mu_\pm(\tau-t_0)}\mathrm{d}\tau\right) e^{\mu_\pm(t-t_0)} \\
&= \left(\pm\frac{1}{\mu_+-\mu_-}[(\mu_\pm+\gamma'_d)T_{\mathrm{u},0}+\hat{\varepsilon}\gamma' T_{\mathrm{d},0}] \pm \frac{\mu_\pm+\gamma'_d}{\mu_+-\mu_-}\int_{t_0}^{t} F' e^{-\mu_\pm(\tau-t_0)}\mathrm{d}\tau\right) e^{\mu_\pm(t-t_0)} \\
&= \pm\frac{e^{(\hat{\lambda}/2)(t-t_0)}}{\mu_+-\mu_-}\left[(\mu_\pm+\gamma'_d)T_{\mathrm{u},0}+\hat{\varepsilon}\gamma' T_{\mathrm{d},0}+(\mu_\pm+\gamma'_d)\int_{t_0}^{t} F' e^{-\mu_\pm(\tau-t_0)}\mathrm{d}\tau\right] e^{\pm(\kappa/2)(t-t_0)} \\
&= \pm\frac{e^{(\hat{\lambda}/2)(t-t_0)}}{\mu_+-\mu_-}\left[\frac{\hat{\lambda}\pm\kappa+2\gamma'_d}{2}T_{\mathrm{u},0}+\frac{2\hat{\varepsilon}\gamma'}{2}T_{\mathrm{d},0}+\frac{\hat{\lambda}\pm\kappa+2\gamma'_d}{2}\int_{t_0}^{t} F' e^{-\mu_\pm(\tau-t_0)}\mathrm{d}\tau\right] e^{\pm(\kappa/2)(t-t_0)} \\
&= \pm\frac{e^{(\hat{\lambda}/2)(t-t_0)}}{2(\mu_+-\mu_-)}\left[(\hat{\lambda}+2\gamma'_d)T_{\mathrm{u},0}+2\hat{\varepsilon}\gamma' T_{\mathrm{d},0}\pm\kappa T_{\mathrm{u},0}+(\hat{\lambda}+2\gamma'_d\pm\kappa)\int_{t_0}^{t} F' e^{-\mu_\pm(\tau-t_0)}\mathrm{d}\tau\right] e^{\pm(\kappa/2)(t-t_0)}
\end{aligned}
$$

Now that I have a more explicit expression, I write the modes

$$
\begin{aligned}
T_+ \pm T_- =& \\
&\frac{e^{(\hat{\lambda}/2)(t-t_0)}}{2(\mu_+-\mu_-)}\left[(\hat{\lambda}+2\gamma'_d)T_{\mathrm{u},0}+2\hat{\varepsilon}\gamma' T_{\mathrm{d},0}+\kappa T_{\mathrm{u},0}+(\hat{\lambda}+2\gamma'_d+\kappa)\int_{t_0}^{t} F' e^{-\mu_+(\tau-t_0)}\mathrm{d}\tau\right] e^{(\kappa/2)(t-t_0)} \\
&\mp\frac{e^{(\hat{\lambda}/2)(t-t_0)}}{2(\mu_+-\mu_-)}\left[(\hat{\lambda}+2\gamma'_d)T_{\mathrm{u},0}+2\hat{\varepsilon}\gamma' T_{\mathrm{d},0}-\kappa T_{\mathrm{u},0}+(\hat{\lambda}+2\gamma'_d-\kappa)\int_{t_0}^{t} F' e^{-\mu_-(\tau-t_0)}\mathrm{d}\tau\right] e^{-(\kappa/2)(t-t_0)} \\
=& \frac{e^{(\hat{\lambda}/2)(t-t_0)}}{\mu_+-\mu_-}\left\{\left[(\hat{\lambda}+2\gamma'_d)T_{\mathrm{u},0}+2\hat{\varepsilon}\gamma' T_{\mathrm{d},0}\right]\frac{e^{(\kappa/2)(t-t_0)}\mp e^{-(\kappa/2)(t-t_0)}}{2}\right. \\
&+\kappa T_{\mathrm{u},0}\frac{e^{(\kappa/2)(t-t_0)}\pm e^{-(\kappa/2)(t-t_0)}}{2} \\
&+\frac{\hat{\lambda}+2\gamma'_d}{2}\left[e^{(\kappa/2)(t-t_0)}\int_{t_0}^{t} F' e^{-\mu_+(\tau-t_0)}\mathrm{d}\tau \mp e^{-(\kappa/2)(t-t_0)}\int_{t_0}^{t} F' e^{-\mu_-(\tau-t_0)}\mathrm{d}\tau\right] \\
&\left.+\frac{\kappa}{2}\left[e^{(\kappa/2)(t-t_0)}\int_{t_0}^{t} F' e^{-\mu_+(\tau-t_0)}\mathrm{d}\tau \pm e^{-(\kappa/2)(t-t_0)}\int_{t_0}^{t} F' e^{-\mu_-(\tau-t_0)}\mathrm{d}\tau\right]\right\}
\end{aligned}
$$

The last two terms inside the curly brackets have a similar form as the combinations of exponential functions in the first two terms. These combinations of exponential functions are hyperbolic functions which can simplify the expressions of the solutions. I would want such a representation but a problem is there: the integrals are not the same, therefore I cannot factorise them together. Notwithstanding, from the definition of hyperbolic sine and cosine functions, I can write $e^{\pm x} =$

$\cosh x \pm \sinh x$. The factors within square brackets in the last two terms can be thought as $e^x I_+ \pm e^{-x} I_-$, where $I_\pm$ are the corresponding integrals. Using the expression of the exponential function in terms of the hyperbolic functions, I expand $e^x I_+ \pm e^{-x} I_- = (\cosh x + \sinh x) I_+ \pm (\cosh x - \sinh x) I_- = (I_+ \pm I_-) \cosh x + (I_+ \mp I_-) \sinh x$. Then, I overcome the limitation and now the two terms are written with hyperbolic functions. The coefficients of the hyperbolic functions are simple combinations of the integrals which can be also expanded easily. I do that now

$$
\begin{aligned}
I_+ + I_- &= \int_{t_0}^{t} F' e^{-\mu_+(\tau - t_0)} \mathrm{d}\tau + \int_{t_0}^{t} F' e^{-\mu_-(\tau - t_0)} \mathrm{d}\tau = \int_{t_0}^{t} F' [e^{-\mu_+(\tau - t_0)} + e^{-\mu_-(\tau - t_0)}] \mathrm{d}\tau \\
&= \int_{t_0}^{t} F' e^{-(\hat{\lambda}/2)(\tau - t_0)} [e^{-(\kappa/2)(\tau - t_0)} + e^{(\kappa/2)(\tau - t_0)}] \mathrm{d}\tau \\
&= 2 \int_{t_0}^{t} F' e^{-(\hat{\lambda}/2)(\tau - t_0)} \cosh\left[\frac{\kappa}{2}(\tau - t_0)\right] \mathrm{d}\tau \\
I_+ - I_- &= \int_{t_0}^{t} F' e^{-\mu_+(\tau - t_0)} \mathrm{d}\tau - \int_{t_0}^{t} F' e^{-\mu_-(\tau - t_0)} \mathrm{d}\tau = \int_{t_0}^{t} F' [e^{-\mu_+(\tau - t_0)} - e^{-\mu_-(\tau - t_0)}] \mathrm{d}\tau \\
&= \int_{t_0}^{t} F' e^{-(\hat{\lambda}/2)(\tau - t_0)} [e^{-(\kappa/2)(\tau - t_0)} - e^{(\kappa/2)(\tau - t_0)}] \mathrm{d}\tau \\
&= -2 \int_{t_0}^{t} F' e^{-(\hat{\lambda}/2)(\tau - t_0)} \sinh\left[\frac{\kappa}{2}(\tau - t_0)\right] \mathrm{d}\tau
\end{aligned}
$$

If one collects terms corresponding to each hyperbolic function in the former expressions for the normal modes, obtains the following

$$T_\mathrm{s} = \frac{e^{(\hat{\lambda}/2)(t - t_0)}}{\kappa} \left\{ C_1 \cosh\left[\frac{\kappa}{2}(t - t_0)\right] + C_2 \sinh\left[\frac{\kappa}{2}(t - t_0)\right] \right\} \tag{A17}$$

$$T_\mathrm{a} = \frac{e^{(\hat{\lambda}/2)(t - t_0)}}{\kappa} \left\{ C_2 \cosh\left[\frac{\kappa}{2}(t - t_0)\right] + C_1 \sinh\left[\frac{\kappa}{2}(t - t_0)\right] \right\} \tag{A18}$$

where

$$
\begin{aligned}
C_1 &= \kappa T_{\mathrm{u},0} \\
&\quad - (\hat{\lambda} + 2\gamma_d') \int_{t_0}^{t} F' e^{-(\hat{\lambda}/2)(\tau - t_0)} \sinh\left[\frac{\kappa}{2}(\tau - t_0)\right] \mathrm{d}\tau + \kappa \int_{t_0}^{t} F' e^{-(\hat{\lambda}/2)(\tau - t_0)} \cosh\left[\frac{\kappa}{2}(\tau - t_0)\right] \mathrm{d}\tau \\
C_2 &= (\hat{\lambda} + 2\gamma_d') T_{\mathrm{u},0} + 2\hat{\varepsilon}\gamma_d' T_{\mathrm{d},0} \\
&\quad + (\hat{\lambda} + 2\gamma_d') \int_{t_0}^{t} F' e^{-(\hat{\lambda}/2)(\tau - t_0)} \cosh\left[\frac{\kappa}{2}(\tau - t_0)\right] \mathrm{d}\tau - \kappa \int_{t_0}^{t} F' e^{-(\hat{\lambda}/2)(\tau - t_0)} \sinh\left[\frac{\kappa}{2}(\tau - t_0)\right] \mathrm{d}\tau
\end{aligned}
$$

These expressions for the normal modes are quite elegant, and the coefficients C_i summarize all the information from the initial conditions and the forcing. The initial condition terms in the C_i correspond to the non-forced response of the system, while the part that is forcing-dependent corresponds to the forced response of the system.

Forced response to constant forcing

If $F' = F'_c \neq 0$ for $t > t_0$ with F'_c constant and $T_{\mathrm{u},0}, T_{\mathrm{d},0} = 0$ for $t = t_0$, then

$$C_1 = F'_c \left\{ -(\hat{\lambda} + 2\gamma'_d) \int_{t_0}^{t} e^{-(\hat{\lambda}/2)(\tau - t_0)} \sinh\left[\frac{\kappa}{2}(\tau - t_0)\right] \mathrm{d}\tau + \kappa \int_{t_0}^{t} e^{-(\hat{\lambda}/2)(\tau - t_0)} \cosh\left[\frac{\kappa}{2}(\tau - t_0)\right] \mathrm{d}\tau \right\}$$

$$C_2 = F'_c \left\{ (\hat{\lambda} + 2\gamma'_d) \int_{t_0}^{t} e^{-(\hat{\lambda}/2)(\tau - t_0)} \cosh\left[\frac{\kappa}{2}(\tau - t_0)\right] \mathrm{d}\tau - \kappa \int_{t_0}^{t} e^{-(\hat{\lambda}/2)(\tau - t_0)} \sinh\left[\frac{\kappa}{2}(\tau - t_0)\right] \mathrm{d}\tau \right\}$$

where the integrals are easily computed

$$\int_{t_0}^{t} e^{-(\hat{\lambda}/2)(\tau - t_0)} \sinh\left[\frac{\kappa}{2}(\tau - t_0)\right] \mathrm{d}\tau = \frac{e^{-(\hat{\lambda}/2)(t - t_0)}}{\lambda' \gamma'_d} \left\{ \frac{\kappa}{2} \cosh\left[\frac{\kappa}{2}(t - t_0)\right] + \frac{\hat{\lambda}}{2} \sinh\left[\frac{\kappa}{2}(t - t_0)\right] \right\} - \frac{\kappa}{2\lambda' \gamma'_d}$$

$$\int_{t_0}^{t} e^{-(\hat{\lambda}/2)(\tau - t_0)} \cosh\left[\frac{\kappa}{2}(\tau - t_0)\right] \mathrm{d}\tau = \frac{e^{-(\hat{\lambda}/2)(t - t_0)}}{\lambda' \gamma'_d} \left\{ \frac{\hat{\lambda}}{2} \cosh\left[\frac{\kappa}{2}(t - t_0)\right] + \frac{\kappa}{2} \sinh\left[\frac{\kappa}{2}(t - t_0)\right] \right\} - \frac{\hat{\lambda}}{2\lambda' \gamma'_d}$$

and, upon reduction, the C_i are

$$C_1 = \frac{F'_c}{\lambda'} e^{-(\hat{\lambda}/2)(\tau - t_0)} \left\{ -\kappa \cosh\left[\frac{\kappa}{2}(t - t_0)\right] + (2\lambda' - \hat{\lambda}) \sinh\left[\frac{\kappa}{2}(t - t_0)\right] + \kappa e^{(\hat{\lambda}/2)(t - t_0)} \right\}$$

$$C_2 = \frac{F'_c}{\lambda'} e^{-(\hat{\lambda}/2)(\tau - t_0)} \left\{ -(2\lambda' - \hat{\lambda}) \cosh\left[\frac{\kappa}{2}(t - t_0)\right] + \kappa \sinh\left[\frac{\kappa}{2}(t - t_0)\right] + (2\lambda' - \hat{\lambda}) e^{(\hat{\lambda}/2)(t - t_0)} \right\}$$

with these expressions is easy to evaluate the terms inside the curly brackets in equations (A17) and (A18) and the symmetric and antisymmetric modes are (for $t \geq t_0$)

$$T_{\mathrm{s}} = \frac{F_c}{\lambda} \left\{ e^{(\hat{\lambda}/2)(t - t_0)} \left(\cosh\left[\frac{\kappa}{2}(t - t_0)\right] + \frac{2\lambda' - \hat{\lambda}}{\kappa} \sinh\left[\frac{\kappa}{2}(t - t_0)\right] \right) - 1 \right\} \tag{A19}$$

$$T_{\mathrm{a}} = \frac{F_c}{\lambda} \left\{ e^{(\hat{\lambda}/2)(t - t_0)} \left(\frac{2\lambda' - \hat{\lambda}}{\kappa} \cosh\left[\frac{\kappa}{2}(t - t_0)\right] + \sinh\left[\frac{\kappa}{2}(t - t_0)\right] \right) - \frac{2\lambda' - \hat{\lambda}}{\kappa} \right\} \tag{A20}$$

where $F_c' := F_c/C_\mathrm{u}$. I can also obtain the explicit time derivatives of both modes. We take the time derivative both equations (A19) and (A20)

$$
\begin{aligned}
\dot{T}_\mathrm{s} &= \frac{F_c}{\lambda} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \frac{\hat{\lambda}}{2} \left(\cosh\left[\frac{\kappa}{2}(t-t_0)\right] + \frac{2\lambda' - \hat{\lambda}}{\kappa} \sinh\left[\frac{\kappa}{2}(t-t_0)\right] \right) \right. \\
&\qquad \left. + \frac{\kappa}{2} \left(\frac{2\lambda' - \hat{\lambda}}{\kappa} \cosh\left[\frac{\kappa}{2}(t-t_0)\right] + \sinh\left[\frac{\kappa}{2}(t-t_0)\right] \right) \right\} \\
&= \frac{F_c}{\lambda} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \lambda' \cosh\left[\frac{\kappa}{2}(t-t_0)\right] + \frac{\lambda'\hat{\lambda} + 2\gamma_d'\lambda'}{\kappa} \sinh\left[\frac{\kappa}{2}(t-t_0)\right] \right\} \\
&= \frac{F_c}{C_\mathrm{u}} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \cosh\left[\frac{\kappa}{2}(t-t_0)\right] + \frac{\hat{\lambda} + 2\gamma_d'}{\kappa} \sinh\left[\frac{\kappa}{2}(t-t_0)\right] \right\} \\
\dot{T}_\mathrm{a} &= \frac{F_c}{\lambda} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \frac{\hat{\lambda}}{2} \left(\frac{2\lambda' - \hat{\lambda}}{\kappa} \cosh\left[\frac{\kappa}{2}(t-t_0)\right] + \sinh\left[\frac{\kappa}{2}(t-t_0)\right] \right) \right. \\
&\qquad \left. + \frac{\kappa}{2} \left(\cosh\left[\frac{\kappa}{2}(t-t_0)\right] + \frac{2\lambda' - \hat{\lambda}}{\kappa} \sinh\left[\frac{\kappa}{2}(t-t_0)\right] \right) \right\} \\
&= \frac{F_c}{\lambda} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \frac{\lambda'\hat{\lambda} + 2\gamma_d'\lambda'}{\kappa} \cosh\left[\frac{\kappa}{2}(t-t_0)\right] + \lambda' \sinh\left[\frac{\kappa}{2}(t-t_0)\right] \right\} \\
&= \frac{F_c}{C_\mathrm{u}} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \frac{\hat{\lambda} + 2\gamma_d'}{\kappa} \cosh\left[\frac{\kappa}{2}(t-t_0)\right] + \sinh\left[\frac{\kappa}{2}(t-t_0)\right] \right\}
\end{aligned}
$$

I present both results jointly to show the simplicity of the derivatives

$$
\begin{aligned}
\dot{T}_\mathrm{s} &= \frac{F_c}{C_\mathrm{u}} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \cosh\left[\frac{\kappa}{2}(t-t_0)\right] + \frac{\hat{\lambda} + 2\gamma_d'}{\kappa} \sinh\left[\frac{\kappa}{2}(t-t_0)\right] \right\} \\
\dot{T}_\mathrm{a} &= \frac{F_c}{C_\mathrm{u}} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \frac{\hat{\lambda} + 2\gamma_d'}{\kappa} \cosh\left[\frac{\kappa}{2}(t-t_0)\right] + \sinh\left[\frac{\kappa}{2}(t-t_0)\right] \right\}
\end{aligned}
$$

With these derivatives, I can calculate the ratio of the antisymmetric mode derivative to the symmetric one that appears in equation (A15)

$$
\begin{aligned}
\frac{\dot{T}_\mathrm{a}}{\dot{T}_\mathrm{s}} &= \frac{\frac{\hat{\lambda}+2\gamma_d'}{\kappa} \cosh\left[\frac{\kappa}{2}(t-t_0)\right] + \sinh\left[\frac{\kappa}{2}(t-t_0)\right]}{\cosh\left[\frac{\kappa}{2}(t-t_0)\right] + \frac{\hat{\lambda}+2\gamma_d'}{\kappa} \sinh\left[\frac{\kappa}{2}(t-t_0)\right]} \\
&= \frac{\frac{\hat{\lambda}+2\gamma_d'}{\kappa} + \tanh\left[\frac{\kappa}{2}(t-t_0)\right]}{1 + \frac{\hat{\lambda}+2\gamma_d'}{\kappa} \tanh\left[\frac{\kappa}{2}(t-t_0)\right]}
\end{aligned}
$$

Formally, above result have the alternative form

$$\frac{\dot{T}_\mathrm{a}}{\dot{T}_\mathrm{s}} = \tanh\left[\frac{\kappa}{2}(t - t_0) + \operatorname{arctanh}\left(\frac{\hat{\lambda} + 2\gamma'_d}{\kappa}\right)\right]$$

This is possible only if $\left|(\hat{\lambda} + 2\gamma'_d)/\kappa\right| \leq 1$. Let us prove that in our case this follows

$$\left|\frac{\hat{\lambda} + 2\gamma'_d}{\kappa}\right| \leq 1$$

$$\frac{\hat{\lambda}^2 + 4\gamma'_d\hat{\lambda} + 4\gamma'^2_d}{\hat{\lambda}^2 + 4\gamma'_d\lambda'} \leq 1$$

$$\hat{\lambda}^2 + 4\gamma'_d\hat{\lambda} + 4\gamma'^2_d \leq \hat{\lambda}^2 + 4\gamma'_d\lambda'$$

$$\hat{\lambda} + \gamma'_d \leq \lambda'$$

$$-\hat{\varepsilon}\gamma' \leq 0$$

the last inequality is always true, since $\hat{\varepsilon}, \gamma'$ are positive constants. Thus,

$$\frac{\dot{T}_\mathrm{a}}{\dot{T}_\mathrm{s}} = \tanh\left[\frac{\kappa}{2}(t - t_0) + \operatorname{arctanh}\left(\frac{\hat{\lambda} + 2\gamma'_d}{\kappa}\right)\right] \tag{A21}$$

Equation (A21) is an hyperbolic tangent that grows from -1 to 1 in a sigmoidal fashion. It has a scaling factor that determines how fast it goes from -1 to 1. It also has a shift that sets where the hyperbolic tangent will cross zero. Both the scaling and shift depend on the thermal and radiative parameters of the system. Since the shift is negative, after the initial forcing the deep ocean (that depends on the antisymmetric mode) warms up slower than the upper ocean. At a latter time, the ratio becomes positive and the contrary happens. The time at which the sign reverses is

$$t_1 = t_0 + \frac{2}{\kappa}\operatorname{arctanh}\left|\frac{\hat{\lambda} + 2\gamma'_d}{\kappa}\right|$$

Variation of the climate feedback parameter

With the solution shown before, the $NT-$diagram has a slope

$$\frac{\dot{N}}{\dot{T}_\mathrm{u}} = \frac{\hat{\varepsilon}+1}{2\hat{\varepsilon}}\left(1+\frac{\hat{\varepsilon}-1}{\hat{\varepsilon}+1}\frac{C_\mathrm{u}\kappa}{|\lambda|}\left[\left(\hat{\varepsilon}+\frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\frac{\gamma}{C_\mathrm{u}\kappa} - \tanh\left(\frac{\kappa}{2}(t-t_0)+\operatorname{arctanh}\left(\frac{\hat{\lambda}+2\gamma'_d}{\kappa}\right)\right)\right]\right)\lambda \quad \text{(A22)}$$

The factor is composed of terms that are positive except for the ratio term coming from equation (A21). The negative ratio for $t \in [t_0, t_1)$ clearly generates a more negative slope, whereas for $t \in (t_1, \infty)$ makes it less negative. At the start one can get the slope

$$\frac{\dot{N}}{\dot{T}_\mathrm{u}} = \left(1+(\hat{\varepsilon}-1)\frac{\gamma}{|\lambda|}\right)\lambda,\ t = t_0$$

and at the time of sign reversal

$$\frac{\dot{N}}{\dot{T}_\mathrm{u}} = \frac{\hat{\varepsilon}+1}{2\hat{\varepsilon}}\left(1+\frac{\hat{\varepsilon}-1}{\hat{\varepsilon}+1}\left(\hat{\varepsilon}+\frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\frac{\gamma}{|\lambda|}\right)\lambda,\ t = t_1$$

After the sign reversal the factor of λ will only decrease up to

$$\lim_{t\to\infty}\frac{\dot{N}}{\dot{T}_\mathrm{u}} = \frac{\hat{\varepsilon}+1}{2\hat{\varepsilon}}\left(1+\frac{\hat{\varepsilon}-1}{\hat{\varepsilon}+1}\frac{C_\mathrm{u}\kappa}{|\lambda|}\left[\left(\hat{\varepsilon}+\frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\frac{\gamma}{C_\mathrm{u}\kappa}-1\right]\right)\lambda$$

Equation (A22) shows the importance of the ratio of the symmetric and antisymmetric modes. Its physical meaning, the relationship between the upper- and deep-ocean warming, sets the strength of the variation of the climate feedback, whereas the constant term sets a base enhancement around which the feedback evolves. The thermal capacities of the system determine this constant term.

APPENDIX B

Feedbacks and pattern effect in a non-linear planetary budget

I start with a planetary imbalance considering a variation of the planetary thermal capacity

$$N = (1-\alpha)S + G - \epsilon\sigma(fT_\mathrm{u})^4 - \dot{C}T_\mathrm{u} \quad \text{(B1)}$$

where S is the incoming solar short-wave flux at the TOA, α is the planetary albedo, G are the remaining natural and anthropogenic energy fluxes, and the last two terms are the planetary long-wave response and the contribution to the radiative response of a varying thermal capacity. As said in the main text, the ocean circulation and the atmosphere-ocean coupling provide the dynamical component of the thermal capacity.

If I compute the total derivative of N then

$$\begin{aligned}
\dot{N} &= \left[(1-\alpha)\dot{S} + \dot{G}\right] - S\dot{\alpha} - \sigma(fT_\mathrm{u})^4\dot{\epsilon} - 4\epsilon\sigma(fT_\mathrm{u})^3(\dot{f}T_\mathrm{u} + f\dot{T}_\mathrm{u}) - \dot{C}\dot{T}_\mathrm{u} - T_\mathrm{u}\ddot{C} \\
&= \left[(1-\alpha)\dot{S} + \dot{G}\right] - \mathcal{R}
\end{aligned}$$

Here we can see the first term is the change from a time-evolving forcing. The rest of the terms, $\mathcal{R}$, are atmospheric feedbacks or the effects of ocean circulation and ocean-atmosphere interaction. The fourth term contains the Planck feedback. Let us compare all the terms of $\mathcal{R}$ in comparison to the Planck feedback term $4\epsilon f\sigma(fT_\mathrm{u})^3\dot{T}_\mathrm{u}$

$$\begin{aligned}
\mathcal{R} &= S\dot{\alpha} + \sigma(fT_\mathrm{u})^4\dot{\epsilon} + 4\epsilon\sigma(fT_\mathrm{u})^3(\dot{f}T_\mathrm{u} + f\dot{T}_\mathrm{u}) + \dot{C}\dot{T}_\mathrm{u} + T_\mathrm{u}\ddot{C} \\
&= 4\epsilon f\sigma(fT_\mathrm{u})^3\dot{T}_\mathrm{u}\left[\frac{S}{4\epsilon f\sigma(fT_\mathrm{u})^3}\frac{\dot{\alpha}}{\dot{T}_\mathrm{u}} + \frac{T_\mathrm{u}}{4\epsilon}\frac{\dot{\epsilon}}{\dot{T}_\mathrm{u}} + \frac{T_\mathrm{u}}{f}\frac{\dot{f}}{\dot{T}_\mathrm{u}} + 1 + \frac{\dot{C}}{4\epsilon f\sigma(fT_\mathrm{u})^3} + \frac{T_\mathrm{u}}{4\epsilon f\sigma(fT_\mathrm{u})^3}\frac{\ddot{C}}{\dot{T}_\mathrm{u}}\right]
\end{aligned}$$

By inserting former expression of $\mathcal{R}$ in the total derivative of the planetary imbalance, reordering and dividing by $\dot{T}_\mathrm{u}$, we get the analogous expression for the slope of the $NT-$diagrams

$$\begin{aligned}
\frac{\dot{N}}{\dot{T}_\mathrm{u}} &= \left[(1-\alpha)\frac{\dot{S}}{\dot{T}_\mathrm{u}} + \frac{\dot{G}}{\dot{T}_\mathrm{u}}\right] \\
&\quad - \left[1 + \frac{S}{4\epsilon f\sigma(fT_\mathrm{u})^3}\frac{\dot{\alpha}}{\dot{T}_\mathrm{u}} + \frac{T_\mathrm{u}}{4\epsilon}\frac{\dot{\epsilon}}{\dot{T}_\mathrm{u}} + \frac{T_\mathrm{u}}{f}\frac{\dot{f}}{\dot{T}_\mathrm{u}} + \frac{\dot{C}}{4\epsilon f\sigma(fT_\mathrm{u})^3} + \frac{T_\mathrm{u}}{4\epsilon f\sigma(fT_\mathrm{u})^3}\frac{\ddot{C}}{\dot{T}_\mathrm{u}}\right]4\epsilon f\sigma(fT_\mathrm{u})^3
\end{aligned}$$

The first contribution in the $\mathcal{R}/\dot{T}_\mathrm{u}$ term is 1, representing the Planck feedback. The second contribution is the planetary albedo feedback. It includes the surface albedo feedback as well as the short-wave cloud feedback. The third contribution is the emissivity feedback, to which mainly contributes the traditional water-vapor feedback. The fourth contribution is a representation of the lapse-rate feedback. The fifth and sixth contributions are not atmospheric feedbacks but the effect of the evolving planetary thermal capacity provided by the atmosphere-ocean interaction and the ocean circulation.

Both the fifth and sixth contributions measure the effect of a changing planetary thermal capacity. The fifth term should be positive but reduces its contribution towards the equilibrium in view of the modified two-layer model results. In the same context, the sixth contribution should change sign, in analogy to the linearized model results.

References

Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor, 2012: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophys. Res. Lett.*, **39 (9)**, L09 712, https://doi.org/10.1029/2012GL051607.

Armour, K. C., 2017: Energy budget constraints on climate sensitivity in light of inconstant climate feedbacks. *Nat. Clim. Change*, **7 (5)**, 331–335, https://doi.org/10.1038/nclimate3278.

Armour, K. C., C. M. Bitz, and G. H. Roe, 2013: Time-Varying Climate Sensitivity from Regional Feedbacks. *J. Climate*, **26 (13)**, 4518–4534, https://doi.org/10.1175/JCLI-D-12-00544.1.

Arrhenius, S., 1896: On the Influence of Carbonic Acid in the Air upon the Temperature of the Ground. *Philosophical Magazine and Journal of Science*, **41**, 237–276.

Budyko, M., 1969: The effect of solar radiation variations on the climate of the Earth. *Tellus*, **21 (5)**, 611–619, https://doi.org/10.1111/j.2153-3490.1969.tb00466.x.

Callendar, G. S., 1938: The artificial production of carbon dioxide and its influence on temperature. *Quarterly Journal of the Royal Meteorological Society*, **64 (275)**, 223–240, https://doi.org/10.1002/qj.49706427503.

Ceppi, P., and J. M. Gregory, 2017: Relationship of tropospheric stability to climate sensitivity and Earth's observed radiation budget. *Proc. Natl. Acad. Sci. (USA)*, **114 (50)**, 13 126–13 131, https://doi.org/10.1073/pnas.1714308114.

Datseris, G., J. Blanco, O. Hadas, S. Bony, R. Caballero, Y. Kaspi, and B. Stevens, 2022: Minimal recipes for global cloudiness. *Earth and Space Science Open Archive*, 14, https://doi.org/10.1002/essoar.10510797.2.

Dong, Y., C. Proistosescu, K. C. Armour, and D. S. Battisti, 2019: Attributing Historical and Future Evolution of Radiative Feedbacks to Regional Warming Patterns using a Green's Function Approach: The Preeminence of the Western Pacific. *J. Climate*, **32 (17)**, 5471–5491, https://doi.org/10.1175/JCLI-D-18-0843.1.

Fourier, J.-B. J., 1827: Mémoire sur les Températures du Globe Terrestre et des Espaces Planétaires. *Mémoires d l'Académie Royale des Sciences de l'Institute de France*, **7**, 570–604.

Geoffroy, O., D. Saint-Martin, G. Bellon, A. Voldoire, D. J. L. Olivié, and S. Tytéca, 2013a: Transient Climate Response in a Two-Layer Energy-Balance Model. Part II: Representation of the Efficacy of Deep-Ocean Heat Uptake and Validation for CMIP5 AOGCMs. *J. Climate*, **26 (6)**, 1859–1876, https://doi.org/10.1175/JCLI-D-12-00196.1.

Geoffroy, O., D. Saint-Martin, D. J. L. Olivié, A. Voldoire, G. Bellon, and S. Tytéca, 2013b: Transient Climate Response in a Two-Layer Energy-Balance Model. Part I: Analytical Solution and Parameter Calibration Using CMIP5 AOGCM Experiments. *J. Climate*, **26 (6)**, 1841–1857, https://doi.org/10.1175/JCLI-D-12-00195.1.

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

Gregory, J. M., and Coauthors, 2004: A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.*, **31 (3)**, L03 205, https://doi.org/10.1029/2003GL018747.

Hansen, J. E., R. A. Ruedy, M. Sato, and K.-W. K. Lo, 2010: Global surface temperature change. *Reviews of Geophysics*, **48 (4)**, RG4004, https://doi.org/10.1029/2010RG000345, URL https://doi.org/10.1029/2010RG000345.

Hansen, J. E., G. Russell, A. Lacis, I. Fung, D. Rind, and P. Stone, 1985: Climate Response Times: Dependence on Climate Sensitivity and Ocean Mixing. *Science*, **229 (4716)**, 857–859, https://doi.org/10.1002/2017MS001208.

Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis, 2010: Probing the Fast and Slow Components of Global Warming by Returning Abruptly to Preindustrial Forcing. *J. Climate*, **23 (9)**, 2418–2427, https://doi.org/10.1175/2009JCLI3466.1.

Hu, S., S.-P. Xie, , and S. M. Kang, 2021: Global Warming Pattern Formation: The Role of Ocean Heat Uptake. *J. Climate*, **35 (6)**, 1885–1899, https://doi.org/10.1175/JCLI-D-21-0317.1.

Jiménez-de-la-Cuesta, D., 2022a: diegojco/pattern-effect: Preprint release. *Software in Github*, https://doi.org/10.5281/zenodo.6530577.

Jiménez-de-la-Cuesta, D., 2022b: Global and tropical band averages for a selection of CMIP5 and CMIP6 models: piControl and abrupt-4xCO2 experiments. *Dataset in Zenodo*, https://doi.org/10.5281/zenodo.6531208.

Kiehl, J., 2007: Twentieth century climate model response and climate sensitivity. *Geophys. Res. Lett.*, **34 (22)**, L22 710, https://doi.org/10.1029/2007GL031383.

Lin, Y.-J., Y.-T. Hwang, J. Lu, F. Liu, and B. E. J. Rose, 2021: The Dominant Contribution of Southern Ocean Heat Uptake to Time-Evolving Radiative Feedback in CESM. *Geophys. Res. Lett.*, **48 (9)**, e2021GL093 302, https://doi.org/10.1029/2021GL093302.

Mauritsen, T., 2016: Clouds cooled the Earth. *Nat. Geosci.*, **9 (12)**, 865–867, https://doi.org/10.1038/ngeo2838.

Meraner, K., T. Mauritsen, and A. Voigt, 2013: Robust increase in equilibrium climate sensitivity under global warming. *Geophys. Res. Lett.*, **40 (2)**, 5944–5948, https://doi.org/10.1002/2013GL058118.

Newsom, E., L. Zanna, S. Khatiwala, and J. M. Gregory, 2020: The Influence of Warming Patterns on Passive Ocean Heat Uptake. *Geophys. Res. Lett.*, **47 (18)**, e2020GL088 429, https://doi.org/10.1029/2020GL088429.

Otto, A., and Coauthors, 2013: Energy budget constraints on climate response. *Nat. Geosci.*, **6 (6)**, 415–416, https://doi.org/10.1038/ngeo1836.

Rohrschneider, T., B. Stevens, and T. Mauritsen, 2019: On simple representations of the climate response to external radiative forcing. *Climate Dyn.*, **3 (5-6)**, 3131–3145, https://doi.org/10.1007/s00382-019-04686-4.

Rugenstein, M. A. A., and K. C. Armour, 2021: Three Flavors of Radiative Feedbacks and Their Implications for Estimating Equilibrium Climate Sensitivity. *Geophys. Res. Lett.*, **48 (15)**, e2021GL092 983, https://doi.org/10.1029/2021GL092983.

Senior, C. A., and J. F. B. Mitchell, 2000: The time-dependence of climate sensitivity. *Geophys. Res. Lett.*, **27 (17)**, 2685–2688, https://doi.org/10.1029/2000GL011373.

Talley, L. D., 2015: Closure of the Global Overturning Circulation Through the Indian, Pacific, and Southern Oceans: Schematics and Transports. *Oceanography*, **26 (1)**, 80–97, https://doi.org/10.5670/oceanog.2013.07.

Wills, R. C. J., Y. Dong, C. Proistosescu, K. C. Armour, and D. S. Battisti, 2022: Systematic Climate Model Biases in the Large-Scale Patterns of Recent Sea-Surface Temperature and Sea-Level Pressure Change. *Geophys. Res. Lett.*, **49 (17)**, e2022GL100 011, https://doi.org/10.1029/2022GL100011.

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

Zhou, C., M. D. Zelinka, and S. A. Klein, 2016: Impact of decadal cloud variations on the Earth's energy budget. *Nat. Geosci.*, **9 (12)**, 871–874, https://doi.org/10.1038/ngeo2828.