The varying climate feedback parameter connected to the atmosphere-ocean coupling and ocean circulation

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

Models indicate a time-varying radiative response of the Earth system to CO_2 forcing (Andrews et al. 2012; Zhou et al. 2016). This variation implies a significant uncertainty in the estimates of climate sensitivity to increasing atmospheric CO_2 concentration (Hawkins and Sutton 2009; Grose et al. 2018). In energy-balance models, the temporal variation is represented as an additional feedback mechanism (Winton et al. 2010; Geoffroy et al. 2013a; Rohrschneider et al. 2019), which also depends on the ocean temperature change. Models and observations also indicate that a spatio-temporal pattern in surface warming controls this additional contribution to the radiative response by modulating the tropospheric instability (Ceppi and Gregory 2017; Zhou et al. 2016). Some authors focus on the atmospheric mechanisms that drive the feedback change (Stevens et al. 2016), reducing the role of the ocean's energy uptake variations. For the first time, I derive, using a linearized conceptual energy-balance model (Winton et al. 2010; Geoffroy et al. 2013a; Rohrschneider et al. 2019), an explicit mathematical expression of the radiative response and its temporal evolution. This expression connects the spatio-temporal warming pattern to a dynamical thermal capacity, stemming from changes in the ocean energy uptake. In comparison with more realistic energy-balance frameworks, and unlike the notion of additional feedback mechanisms, I show that an expanded effective thermal capacity better explains the variation of the radiative response, naturally connects with the spatio-temporal surface warming pattern and provides a non-circular framework to explain the variation of the climate feedback parameter.

Significance statement. Understanding the factors that change the Earth's radiative response to forcing is central to reduce the uncertainty in the climate sensitivity estimates. The current atmospheric-only view on the problem of the time-varying climate feedback parameter unnecessarily hides the ocean's role. This work shows a novel perspective for the problem, enabling the development of a more general theory.

1. Introduction

Climate models show a wide range of temporal variation in their radiative response to CO_2 forcing (Senior and Mitchell 2000; Andrews et al. 2012; Ceppi and Gregory 2017). This variation appears in numerical experiments where the atmospheric CO_2 concentration is raised and maintained constant afterwards. The rise in the atmospheric CO_2 concentration modifies the Earth's emissivity to long-wave radiation, resulting in surface warming. Surface warming modifies the radiative flux at the top of the atmosphere (TOA). The modified flux tends to cancel the energy imbalance introduced by the radiative forcing. Surface warming also changes other variables, such as the atmospheric temperature and humidity, that further modify the radiative flux. These changes are the feedback mechanisms on surface warming. The net rate at which the globally-averaged surface warming reduces the globally-averaged TOA imbalance is known as the climate feedback parameter.

If the feedback mechanisms did not change with time, the climate feedback parameter would be constant, and a diagram of globally-averaged TOA imbalance change N versus surface temperature change T_u ($NT-$diagram) would be linear. However, climate models present $NT-$diagrams with different degrees of curvature, indicating a non-constant climate feedback parameter (presented in Figure 1; Andrews et al. 2012; Ceppi and Gregory 2017). The degree of curvature is also modified by forcing strength (Senior and Mitchell 2000; Meraner et al. 2013; Rohrschneider et al. 2019). Temporal and state dependencies can be the root of climate feedback's variation. Observations indicate that a spatio-temporal pattern of surface warming modifies cloud feedback in decadal timescales by altering atmospheric stability, leading to feedback changes that depend not only on surface warming (Zhou et al. 2016; Mauritsen 2016; Ceppi and Gregory 2017). The spatio-temporal warming pattern should depend also on the state of the system leading to contributions to the climate feedback's temporal and state dependencies.

In a globally-averaged energy-balance framework, N should be equal to the forcing F plus the radiative response of the system R, $N = F + R$. Following the classical picture of the linearized feedback mechanisms depending only on the surface warming (Gregory et al. 2002) T_u, we should have $R \sim \lambda T_\mathrm{u}$, where λ is a constant climate feedback parameter. Thus, if we consider a constant forcing F, the slope of the $NT-$diagram would be constant and equal to λ, in contradiction with observations and complex models as discussed above. Thus, either the non-linear component plays a more significant role, or the feedback mechanisms depend on more than the surface warming.

The problem cannot be solved by introducing more structure in the system. For instance, if we introduce two coupled layers (Winton et al. 2010; Geoffroy et al. 2013b; Rohrschneider et al. 2019). One layer represents the atmosphere, land and the mixed upper ocean layer: the upper layer. The deep layer corresponds to the deep ocean. Both layers have different thermal capacities. The thermal capacities provide two timescales. One can show that these timescales only delay the equilibrium but do not alter the climate feedback parameter λ. For the upper layer, the energy budget is $N_\mathrm{u} = F + \lambda T_\mathrm{u} - H$, where H is the coupling between the upper and the deep layer: the deep-ocean energy uptake. The deep layer budget is, therefore, $N_\mathrm{d} = H$. The sum of both budgets provides the planetary imbalance at the TOA $N = F + \lambda T_\mathrm{u}$. Thus, N has the same form as before.

Geoffroy et al. (2013a) introduced a perturbed deep-ocean energy uptake in the upper-layer budget, $H' := \varepsilon H$, where ε is the efficacy parameter. The deep-layer budget is still equal to H, leading to a different energy imbalance at the TOA: $N = F + \lambda T_\mathrm{u} + (H - H')$. Although the

$H - H'$ term appears to violate the energy conservation principle, it can be interpreted as additional atmospheric feedbacks (e.g. Armour et al. 2013; Stevens et al. 2016) that depend on the deep ocean's temperature. However, this explanation seems to be ad-hoc.

One of the weakness of the classical approach is the linear approximation. The climate feedback parameter λ is defined in terms of the derivative of R evaluated at the initial state. In the same manner, the deep-ocean energy uptake H is defined in terms of a derivative evaluated at the initial state. Therefore, the seemingly ad-hoc term introduced by Geoffroy et al. (2013a) has a more concrete interpretation: the difference between the actual H' and the reference H results in a surplus $(1-\varepsilon)H$. This surplus comes from a dynamically-expanded capacity of the deep ocean to uptake energy: an effective thermal capacity. This effective thermal capacity influences the flux between the upper and deep layers, modifying the surface temperature from below. Once the surface temperature is modified, the usual atmospheric feedbacks change.

Considering the analytical solutions of the modified two-layer model, I derive for the first time an explicit mathematical expression for the slope of the $NT-$diagrams, including the explicit time evolution of this slope. At its core, this expression has the ratio of change of the energy stored by the upper and deep layers and supports the more concrete interpretation that I presented above. The interpretation of these results is that the atmosphere-ocean coupling sets the spatio-temporal warming pattern. Afterwards, the atmosphere adjusts, leading to the changes in the feedback mechanisms. I first show the theory of this novel mathematical expression using the linearized two-layer model. Afterwards, I take a step back away from the approximation to show why the interpretation given above is more relevant.

2. Theory

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

$$\begin{cases} C_{\mathrm{u}}\dot{T}_{\mathrm{u}} = F + \lambda T_{\mathrm{u}} - \varepsilon\gamma(T_{\mathrm{u}} - T_{\mathrm{d}}) \\ C_{\mathrm{d}}\dot{T}_{\mathrm{d}} = \gamma(T_{\mathrm{u}} - T_{\mathrm{d}}) \end{cases}$$

where the first equation corresponds to the upper-layer budget and the second equation to the deep layer. The parameters λ and γ are the climate feedback parameter and the rate of deep-ocean energy uptake in the neighbourhood of the initial state. C_{u} and C_{d} are respectively the thermal capacities (per unit area) of the upper and deep layers. T_{u} and T_{d} are temperature anomalies referred to the initial state and the dotted quantities are time total derivatives. The planetary imbalance is the sum of both equations, resulting in $N = F + \lambda T_{\mathrm{u}} + (1-\varepsilon)\gamma(T_{\mathrm{u}} - T_{\mathrm{d}})$. Nonetheless, it is better to write these equations in the following fashion

$$\begin{cases} \dot{T}_{\mathrm{u}} = F' + \lambda' T_{\mathrm{u}} - \varepsilon\gamma'(T_{\mathrm{u}} - T_{\mathrm{d}}) \\ \dot{T}_{\mathrm{d}} = \gamma'_d(T_{\mathrm{u}} - T_{\mathrm{d}}) \end{cases} \tag{1}$$

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 (1) 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. In the following, I proceed by summarizing the relevant facts, leaving the full mathematical discussion to the appendix A of this article.

The homogeneous ($F' \equiv 0$) version of the system (1) has two distinct eigenvalues $\mu_{\pm} := (\hat{\lambda} \pm \kappa)/2$, where $\hat{\lambda} := \lambda' - \varepsilon\gamma' - \gamma'_d$ and $\kappa^2 := \hat{\lambda}^2 + 4\lambda'\gamma'_d$. These eigenvalues provide two distinct eigenvectors, forming a basis in which the full system (1) is uncoupled and, therefore, has a straight-forward solution. The eigensolutions $T_{\pm}$ are the solutions associated with each eigenvalue. Afterwards, one can return to the original representation, finding that T_{u} and T_{d} are linear combinations of

$T_{\pm}$. These linear combinations are the normal modes: the symmetric mode $T_s := T_+ + T_-$ and the antisymmetric mode $T_a := T_+ - T_-$. The main result of this process is that $T_\mathrm{u} = T_s$ and $T_\mathrm{d} = \alpha T_s + \beta T_a$, where α and β are scalars that depend on the coefficients of the system (1). The normal-mode representation makes explicit the coupling of the deep layer with the upper layer.

3. Results

(i) The explicit slope of the NT−diagram From the solutions to system (1) written in terms of the normal modes, one can obtain an expression for the slope of the NT−diagram, $\dot{N}/\dot{T}_\mathrm{u}$, of a system under constant forcing. In the appendix, I derive the following closed expression for the slope in terms of the derivatives of the normal modes and as a factor of the constant climate feedback parameter λ

$$\frac{\dot{N}}{\dot{T}_\mathrm{u}} = \left\{ \frac{\varepsilon + 1}{2\varepsilon} + \frac{\varepsilon - 1}{2\varepsilon} \frac{C_\mathrm{u}\kappa}{|\lambda|} \left[\left(\frac{\varepsilon}{C_\mathrm{u}} + \frac{1}{C_\mathrm{d}} \right) \frac{\gamma}{\kappa} - \frac{\dot{T}_a}{\dot{T}_s} \right] \right\} \lambda \tag{2}$$

The main characteristic of equation (2) is the square-bracket term of its right-hand side. It contains two parts. The first one sets a basic enhanced slope and contains the sum of the inverse of the thermal capacities as if we had an electrical circuit with capacitors in series. The second part provides the time evolution. It is a ratio of the changes in energy content. This ratio compares the change in energy content of the deep layer with that of the upper layer. To confirm the importance of the square-bracket term, one can take the limit as $\varepsilon \longrightarrow 1$, where the pattern effect is cancelled in equation (2)

$$\lim_{\varepsilon \to 1} \frac{\dot{N}}{\dot{T}_\mathrm{u}} = \lambda$$

The strong coupling between the upper and deep layers disappears. We end up with a constant slope. However, if $\varepsilon \neq 1$, the climate feedback parameter varies with the ratio of the changes in energy content from the deep to the upper layer around a basic value that depends on the thermal capacities of the system, the square bracket term in equation (2).

(ii) Explicit expression for the ratio term Using the full mathematical expressions for the solutions T_s and T_a for constant forcing, I write the ratio term in equation (2) as

$$\frac{\dot{T}_a}{\dot{T}_s} = \tanh\left[\frac{\kappa}{2}(t-t_0) + \operatorname{arctanh}\left(\frac{\hat{\lambda}+2\gamma_d'}{\kappa}\right)\right] \tag{3}$$

The ratio (3) grows in a sigmoidal fashion from -1 to 1. This hyperbolic tangent has a scaling factor $(\kappa/2)$ that sets the rate of change of the hyperbolic tangent between its extreme values. It also has a shift (the arctanh term) that determines when the hyperbolic tangent crosses zero, governing the contribution of the last term in equation (2). Both scaling and shift are in terms of C_u, C_d, ε, γ and λ.

The interpretation of equation (3) is that, after the initial forcing, the deep ocean warms up slower than the upper layer, steepening the slope of the $NT-$diagram. Once the ratio reaches the sign-reversal point, the last term's contribution in equation (2) only flattens the slope of the $NT-$diagram. The scaling factor and the shift of the ratio (3) set the timescale for the flattening. Equation (3) expresses precisely the time evolution of the climate feedback parameter that others have only approximated through numerical experiments with the modified two-layer model (Geoffroy et al. 2013a; Rohrschneider et al. 2019). Additionally, it establishes a third timescale in the Earth system, related to the atmosphere-ocean coupling.

(iii) Explicit expression of the climate feedback parameter Using the explicit expressions, the equation (2) for the climate feedback parameter is

$$\frac{\dot{N}}{\dot{T}_\mathrm{u}} = \frac{\varepsilon+1}{2\varepsilon}\left(1+\frac{\varepsilon-1}{\varepsilon+1}\frac{C_\mathrm{u}\kappa}{|\lambda|}\left[\left(\frac{\varepsilon}{C_\mathrm{u}}+\frac{1}{C_\mathrm{d}}\right)\frac{\gamma}{\kappa} - \tanh\left(\frac{\kappa}{2}(t-t_0)+\operatorname{arctanh}\left(\frac{\hat{\lambda}+2\gamma_d'}{\kappa}\right)\right)\right]\right)\lambda \tag{4}$$

The factor of λ is composed of terms that are positive except for the ratio term coming from equation (3). One can prove that at the start ($t = t_0$) the slope is

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

and from here up to the sign reversal of the ratio term, the slope flattens. The flattening is gentle at first, but towards the sign reversal it accelerates.

At the time of sign reversal we have

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

and from here and on, the ratio term becomes positive, leading to an even flatter slope. The flattening decelerates and becomes gentle again. The asymptotic value of the slope of the NT−diagram is

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

(iv) Numerical estimates of the atmosphere-ocean coupling By substituting in expression (4) the parameter values found by Geoffroy et al. (2013a), I find the timescale for the sign reversal of the $\dot{T}_a/\dot{T}_s$ ratio term. I use the multimodel mean values reported by Geoffroy et al. (2013a). For the multimodel average values ($C_\mathrm{u} = 8.2\,\mathrm{W\,yr\,m^{-2}\,K^{-1}}$, $C_\mathrm{d} = 109\,\mathrm{W\,yr\,m^{-2}\,K^{-1}}$, $\gamma = 0.67\,\mathrm{W\,m^{-2}\,K^{-1}}$, $\lambda = -1.18\,\mathrm{W\,m^{-2}\,K^{-1}}$ and $\varepsilon = 1.28$) the sign reversal of the ratio term takes place after 18.3 years. This timescale lies between the fast (4.2 years) and slow (290 years) timescales established in terms of the thermal capacities alone (Geoffroy et al. 2013a). This timescale is when we are at one half of the change between the initial and final values of the slope.

I calculate the time for sign reversal using the rest of values in the tables of Geoffroy et al. (2013a) and obtain that the multimodel average is 18.8 years. The minimum value is 8.8 years for GISS-E2-R, whereas the maximum is 25.1 years for CNRM-CM5.1. If I compare with their estimates of the fast and the slow timescales, even the extreme values fit well between both. Enlightening is that the timescale of the sign reversal seems to fit with the de-facto 20-year standard to evaluate the change in slope (e.g. Ceppi and Gregory 2017).

I also compare between the multimodel averages for all parameters and with the thermal capacities as calculated by Jiménez-de-la-Cuesta and Mauritsen (2019): $C_\mathrm{u} = 7.2\,\mathrm{W\,yr\,m^{-2}\,K^{-1}}$, $C_\mathrm{d} = 367\,\mathrm{W\,yr\,m^{-2}\,K^{-1}}$. The calculated deep-layer thermal capacity is larger than the CMIP5 multi-model average, whereas the calculated value for the upper layer is smaller than the CMIP5 average. From these differences, we can note changes in the slope evolution (figure 2). Although the difference in final slopes is small, the calculated thermal capacities strongly shift the sign-reversal timescale: a deeper deep ocean lengthens the sign-reversal timescale, whereas a shallower upper layer shortens it.

4. Analysis and Discussion

(i) Consequences of the equation (4) We have two terms in the factor of equation (4): the identity term and the $(\varepsilon - 1)$−term. The second term is only active if $\varepsilon \neq 1$, and has two contributions. The first one is a constant contribution linked to the thermal capacities of the system. The second contribution is time-varying and depends on the ratio $\dot{T}_a/\dot{T}_s$. This ratio measures the proportion of energy that goes into the deep ocean compared to that stored in the upper layer. Together, these terms provide a physical picture in which the slope's variation is determined by a basic thermal capacity, which is expanded dynamically. The expansion stems from the changing energy fluxes between the upper and deep layers that differ from the flux at the starting state represented by γ. This flux difference changes the surface temperature and connects with the evolving spatio-temporal warming pattern in a more concrete fashion, given that the evolution and spatial distribution of the sea surface temperature corresponds to changes in the energy fluxes.

I showed above that the thermal capacities have a strong effect on the timescale at which the slope of the $NT-$diagram changes (figure 2). Thermal capacities in complex models depend strongly on the depth of the ocean mixed-layer and, therefore, on the atmosphere-ocean coupling, providing diverse behaviors (figure 3)

The above interpretation of the two-layer model in the context of the real Earth System is the following: The relative change in the energy fluxes due to the atmosphere-ocean coupling compels the atmospheric feedbacks to adjust. Thus, the magnitude of the changes in the atmospheric radiative response depend on the atmosphere-ocean coupling. One can argue that the non-local free-tropospheric warming from the deep-convective warm regions is enough to explain the change in feedbacks. However, for the pattern effect to act as proposed by Zhou et al. (2016); Mauritsen (2016); Ceppi and Gregory (2017), one needs the surface temperature variation in the subsidence areas. Here the atmosphere-ocean coupling and oceanic circulation enter to change the surface temperature, determining the spatio-temporal warming pattern. Thus, the atmosphere-ocean coupling can play a larger role than thought before (Kiehl 2007).

(ii) State and forcing dependence In this article, I ignored the dependence on the strength of forcing (Senior and Mitchell 2000; Meraner et al. 2013; Rohrschneider et al. 2019). However, such dependence should come from the reference values ε, λ and γ that exist under a particular forcing. Values of λ and γ are first-order derivatives in the neighbourhood of the starting states. Therefore, we need to explore the physics behind the ε parameter to understand how it can change depending on forcing. The physics of ε is the atmosphere-ocean coupling, including circulation. Therefore, we should explore how forcing impacts the atmosphere-ocean coupling, resulting in changes in the spatio-temporal warming pattern.

There are versions of linearized energy balance models in which a simple non-linear term is introduced (Rohrschneider et al. 2019). Although higher-order terms in the Taylor expansion of either the radiative response R or the energy uptake H can provide additional information, the dependencies arising from the atmosphere-ocean coupling, as shown in this article, are far more important in light of the results presented above.

(iii) Non-linear planetary energy balance Above I presented evidence favouring the ocean's energy uptake central role in determining the spatio-temporal warming pattern and its effects on the atmospheric feedback mechanisms. I test this idea in a more general theoretical framework by writing the planetary energy budget in another widely-known representation

$$\frac{d}{dt}(CT_{\mathrm{u}}) = (1-\alpha)S + G - \epsilon\sigma(fT_{\mathrm{u}})^4 \tag{5}$$

where $S := S(t)$ in $\mathrm{W\,m^{-2}}$ is the incoming solar radiative flux at the TOA, α is the planetary albedo, $G := G(t)$ in $\mathrm{W\,m^{-2}}$ represents the remaining inputs (natural and anthropogenic), and the last term is the usual planetary long-wave emission, in $\mathrm{W\,m^{-2}}$, as a grey-body of emissivity ϵ and surface temperature T_{u} with f the lapse-rate scaling factor for the emission temperature. At first inspection, we have the origin of the feedback mechanisms: the planetary albedo α, the emissivity ϵ and the scaling factor f. On the one hand, we have the short-wave strand, the planetary albedo $\alpha := \alpha(T_{\mathrm{u}}, q_{cld,w}, \ldots)$ that is a function of, e.g., the surface temperature (determining ice-sheet and sea-ice area) and the amount of liquid water in the atmosphere forming clouds. On the other hand, we have the long-wave thread, the emissivity and the lapse-rate scaling factor $\epsilon, f := \mathrm{f}(T_{\mathrm{u}}, q_v, q_{cld,w}, \ldots)$, depending on, e.g. the surface temperature, and the amount of water vapor and cloud liquid water in the atmosphere.

The atmospheric feedback mechanisms cannot rely on any temperature we define inside the ocean. The ocean affects α, ϵ and f only through changing T_{u}. In equation (5), we cannot see such dependence. Therefore, here we would be tempted to artificially introduce it by saying that α, ϵ and f depend on another temperature in the ocean, as others have interpreted from the modified two-layer model. In this work, I have shown that there is another more natural place where the ocean enters into play: the energy imbalance at the TOA, N.

We usually picture N as $N = (\mathrm{d}/\mathrm{d}t)(CT_\mathrm{u}) = C\dot{T}_\mathrm{u}$. However, we can have processes that modify the thermal capacity: For example, (a) if the mixed-layer depth varies, it modifies the amount of water in contact with the atmosphere; (b) deep-water upwelling not only reduces the surface temperature but increases the thermal capacity of the mixed layer; (c) a similar effect comes from water from ice sheet melting. Therefore, the sea surface temperature is controlled by the oceanic circulation and the ocean-atmosphere interaction, setting the warming pattern. Thus, a more proper definition for N is $N = (\mathrm{d}/\mathrm{d}t)(CT_\mathrm{u}) - \dot{C}T_\mathrm{u} = C\dot{T}_\mathrm{u}$. The term $\dot{C}T_\mathrm{u}$ considers these processes that modify the sea surface temperature and the ocean energy fluxes from below.

If we rewrite equation (5) using the corrected N:

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

The last term of equation (6) is the representation of the effect of the spatio-temporal warming pattern. The factor $\dot{C}$ needs a new differential equation that describes the temporal variations of the energy uptake due to the ocean circulation and atmosphere-ocean interactions. If we want to use this non-linear framework, we would also need additional differential equations or constitutive relationships for α, ϵ and f.

However, to look at the effect of this additional dynamical thermal capacity term on the atmospheric feedbacks, let us consider the total derivative of N. In the appendix B, I present the details of this derivation. The expression for the total derivative of equation (6) divided by $\dot{T}_{\mathrm{u}}$ is

$$\frac{\dot{N}}{\dot{T}_{\mathrm{u}}} = \left[(1-\alpha)\frac{\dot{S}}{\dot{T}_{\mathrm{u}}} + \frac{\dot{G}}{\dot{T}_{\mathrm{u}}}\right] - \left[1 + \frac{S}{4\epsilon\sigma f(fT_{\mathrm{u}})^3}\frac{\dot{\alpha}}{\dot{T}_{\mathrm{u}}} + \frac{T_{\mathrm{u}}}{4\epsilon}\frac{\dot{\epsilon}}{\dot{T}_{\mathrm{u}}} + \frac{T_{\mathrm{u}}}{f}\frac{\dot{f}}{\dot{T}_{\mathrm{u}}} + \frac{\dot{C}}{4\epsilon\sigma f(fT_{\mathrm{u}})^3} + \frac{T}{4\epsilon\sigma f(fT_{\mathrm{u}})^3}\frac{\ddot{C}}{\dot{T}_{\mathrm{u}}}\right]4\epsilon\sigma f(fT_{\mathrm{u}})^3 \qquad (7)$$

The first term is the contribution from the forcing variation, whereas the second term is the contribution of the radiative response. This last term is the analogue for the expression of the square bracket term in equation (2). This last term has a non-dimensional factor multiplying the product of Planck feedback. In the non-dimensional factor we have a sum of terms. Each term represents the contributions of feedbacks and other processes to the radiative response in comparison to the Planck feedback. The first term in this factor is one, for the Planck feedback. The second term is the contribution of the planetary-albedo feedback, including surface albedo as well as short-wave cloud feedback. The third term is the contribution of the emissivity feedback, including water-vapor feedback. The fourth term is the lapse-rate feedback. The last two terms are the contribution of the atmosphere-ocean interaction and ocean circulation to the radiative response.

Equation (7) is the non-linear analogue of equation (2) if we consider that the forcing is stationary. In a linearization, the first four terms would provide a term analogous to λ. The last two terms, considering that at least $\ddot{C} \neq 0$, would provide the analogue of the remaining terms corresponding to a situation where $\varepsilon \neq 1$. If $\varepsilon = 1$, then $\dot{C} = 0$ and the linearization of equation (7), would provide a constant feedback parameter if we accept that α, ϵ and f covary with T_{u} except for sign. These facts provide more support for interpreting the modified two-layer model as introducing a dynamical thermal capacity, turning the limelight to the role of the ocean circulation and the atmosphere-ocean interactions in setting the sea surface temperature patterns.

5. Conclusions

I presented for the first time an explicit mathematical expression of the slope of the $NT-$diagrams. For that end, I used the linearized framework of the two-layer energy balance model. From the analysis, I uncover another timescale in the Earth System: the timescale of the changes of the climate feedback parameter. The timescale is related to the ratio of change in the energy content between the deep- and the upper-layer (atmosphere, land and mixed upper ocean). In CMIP5 models this time scale is around 18 years, providing theoretical support to the 20-year standard timescale used to study the change in the climate feedback parameter.

The mathematical expression and the analysis of the modified linearized two-layer model suggest that the spatio-temporal warming pattern is a product of the atmosphere-ocean coupling (the H term in equations). Linearization uses the value of the coupling at the starting state. When introducing a modified coupling (H') that represents departures from the starting state, the modified two-layer model can represent the variation of the slope of the $NT-$diagrams. The found mathematical expression suggest that the difference $H - H'$ acts as a dynamically-enlarged thermal capacity that depends on the ratio of change in the energy content between layers, changing the surface temperature. In the real Earth System, therefore, the atmosphere-ocean coupling and the circulation produce the evolving spatio-temporal warming pattern, to which the atmospheric feedbacks respond. I also shortly discussed this interpretation in terms of a non-linear framework, where feedback mechanisms not depending on the surface temperature are complicated to introduce. Instead of introducing exogenous dependencies in the feedback terms, the solution is that the thermal capacity of the system varies. The variation represents the atmosphere-ocean coupling and ocean circulation and relates to the spatio-temporal surface warming pattern. Its effect on the surface temperature then seamlessly translates to changes on the atmospheric feedback mechanisms.

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 this highly theoretical ideas.

Data availability statement. All the data used by this study is described fully in the main text. The numbers for reproducing the figure 2 and 3 are contained in the articles by Geoffroy et al. (2013a) and Jiménez-de-la-Cuesta and Mauritsen (2019). The theoretical considerations are fully described in the appendix to this article. Therefore, this article is fully reproducible and almost self contained.

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.

Matrix form of the equations

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

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

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.

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

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

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

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

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

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

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

$$-\lambda'\gamma'_d+\varepsilon\gamma'\gamma'_d+\mu\gamma'_d-\lambda'\mu+\varepsilon\gamma'\mu+\mu^2-\varepsilon\gamma'\gamma'_d=0$$

$$-\lambda'\gamma'_d+\mu\gamma'_d-\lambda'\mu+\varepsilon\gamma'\mu+\mu^2=0$$

$$-\lambda'\gamma'_d-(\lambda'-\varepsilon\gamma'-\gamma'_d)\mu+\mu^2=0$$

The solutions of equation (A5) are

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

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

$$(\lambda'-\varepsilon\gamma'-\gamma'_d)^2+4\lambda'\gamma'_d\leq 0$$

$$(\lambda'-\varepsilon\gamma')^2-2(\lambda'-\varepsilon\gamma')\gamma'_d+\gamma'^2_d+4\lambda'\gamma'_d\leq 0$$

$$\lambda'^2-2\lambda'\varepsilon\gamma'+(\varepsilon\gamma')^2-2(\lambda'-\varepsilon\gamma')\gamma'_d+\gamma'^2_d+4\lambda'\gamma'_d\leq 0$$

$$\lambda'^2-2\lambda'\varepsilon\gamma'+(\varepsilon\gamma')^2-2\lambda'\gamma'_d+2\varepsilon\gamma'\gamma'_d+\gamma'^2_d+4\lambda'\gamma'_d\leq 0$$

$$(\lambda'/\gamma'_d)^2-2(\lambda'/\gamma'_d)\varepsilon(\gamma'/\gamma'_d)+(\varepsilon(\gamma'/\gamma'_d))^2+2\varepsilon(\gamma'/\gamma'_d)+1+2(\lambda'/\gamma'_d)\leq 0$$

$$(\lambda'/\gamma'_d)^2-2(\lambda'/\gamma'_d)[\varepsilon(\gamma'/\gamma'_d)-1]+(\varepsilon(\gamma'/\gamma'_d))^2+2\varepsilon(\gamma'/\gamma'_d)+1\leq 0$$

$$(\lambda'/\gamma'_d)^2-2(\lambda'/\gamma'_d)[\varepsilon(\gamma'/\gamma'_d)-1]+(\varepsilon(\gamma'/\gamma'_d)+1)^2\leq 0$$

$$(\lambda'/\gamma'_d)^2+(\varepsilon(C_\mathrm{d}/C_\mathrm{u})+1)^2\leq 2(\lambda'/\gamma'_d)[\varepsilon(C_\mathrm{d}/C_\mathrm{u})-1]$$

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 $\varepsilon > C_\mathrm{u}/C_\mathrm{d}$. Given that $\varepsilon \geq 1$ 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 $\varepsilon < 1$ according to Geoffroy et al. (2013a). These also fit here. In the last condition of the above expression we require that $\varepsilon(C_\mathrm{d}/C_\mathrm{u}) - 1 > 0$. If $\varepsilon \geq C_\mathrm{u}/C_\mathrm{d}$ this is fulfilled. $C_\mathrm{u}/C_\mathrm{d}$ is a small quantity and, in the models that have a lesser than one ε, always the ε 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 $\hat{\lambda} := \lambda' - \varepsilon\gamma' - \gamma'_d$ 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 $\hat{\lambda}$

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

and then the matrices for each eigenvalue have the form

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

$$= \begin{pmatrix} \mu_\mp + \gamma'_d & \gamma'_d \\ \varepsilon\gamma' & \mu_\mp - (\lambda' - \varepsilon\gamma') \end{pmatrix}$$

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

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

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

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

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

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

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

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

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

or as a vector in the temperature basis

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

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

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

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

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)

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

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

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

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

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

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

and the temperature basis vectors in the eigenvector representation are

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

Matrix in the eigenvector representation. Solutions

With these results, I can write the matrix **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 $(i, j, k, \ldots)$ have the possible values u,d; and greek indices $(\alpha, \beta, \zeta \ldots)$ 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

$$\mathbf{e}_\alpha = \Theta^i_\alpha \mathbf{e}_i$$

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

$$\mathbf{e}_\alpha = \Theta^i_\alpha \mathbf{e}_i = \Theta^i_\alpha \Lambda^\beta_i \mathbf{e}_\beta$$

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

$$\mathbf{e}_\alpha = \delta^\beta_\alpha \mathbf{e}_\beta = \mathbf{e}_\alpha$$

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

Now, matrix 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: $f(\mathbf{v}) = f(v^j \mathbf{e}_j) = v^j f(\mathbf{e}_j) = v^j \mathsf{A}^k_j \mathbf{e}_k$. Thus, the matrix 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 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

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

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, $\Lambda^{-1}\mathbf{A}$.

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

and multiply the result by Λ

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

$$= \frac{1}{\kappa}\begin{pmatrix} \mu_+^2 + \mu_+\gamma'_d + \varepsilon\gamma'\gamma'_d + \mu_-\gamma'_d + \gamma_d^{'2} & -\mu_+\mu_- - \mu_+\gamma'_d - \varepsilon\gamma'\gamma'_d - \mu_-\gamma'_d - \gamma_d^{'2} \\ \mu_-\mu_+ + \mu_-\gamma'_d + \varepsilon\gamma'\gamma'_d + \mu_+\gamma'_d + \gamma_d^{'2} & -\mu_-^2 - \mu_-\gamma'_d - \varepsilon\gamma'\gamma'_d - \mu_+\gamma'_d - \gamma_d^{'2} \end{pmatrix}$$

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

$$= \frac{1}{\kappa}\begin{pmatrix} \mu_+^2 - \mu_+\mu_- & \lambda'\gamma'_d - \lambda'\gamma'_d \\ -\lambda'\gamma'_d + \lambda'\gamma'_d & -\mu_-^2 + \mu_+\mu_- \end{pmatrix} = \frac{1}{\kappa}\begin{pmatrix} \mu_+\kappa & 0 \\ 0 & \mu_-\kappa \end{pmatrix} = \begin{pmatrix} \mu_+ & 0 \\ 0 & \mu_- \end{pmatrix}$$

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

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

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

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

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

the forcing components are

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

and the solutions in the temperature representation are

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

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

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

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.

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

From now on, I write $T_s := T_+ + T_-$ and $T_a := T_+ - T_-$.

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 $N = N_\mathrm{u} + N_\mathrm{d} = C_\mathrm{u}\dot{T}_\mathrm{u} + C_\mathrm{d}\dot{T}_\mathrm{d}$

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

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

$$
\begin{aligned}
\dot{T}_s &= \mu_+T_+ + \mu_-T_- + (F'_+ + F'_-) = \mu_+T_+ + (\mu_+ - \kappa)T_- + (F'_+ + F'_-) \\
&= \mu_+T_s - \kappa T_- + (F'_+ + F'_-) = \mu_+T_s - \frac{\kappa}{2}(T_s - T_a) + (F'_+ + F'_-) \\
&= \frac{\hat{\lambda}}{2}T_s + \frac{\kappa}{2}T_a + (F'_+ + F'_-) = \frac{\hat{\lambda}}{2}T_s + \frac{\kappa}{2}T_a + F' \\
\dot{T}_a &= \mu_+T_+ - \mu_-T_- + (F'_+ - F'_-) = \mu_+T_+ - (\mu_+ - \kappa)T_- + (F'_+ - F'_-) \\
&= \mu_+T_a + \kappa T_- + (F'_+ - F'_-) = \mu_+T_a + \frac{\kappa}{2}(T_s - T_a) + (F'_+ - F'_-) \\
&= \frac{\kappa}{2}T_s + \frac{\hat{\lambda}}{2}T_a + (F'_+ - F'_-) = \frac{\kappa}{2}T_s + \frac{\hat{\lambda}}{2}T_a + \frac{\hat{\lambda}+2\gamma'_d}{\kappa}F' \;\therefore \\
N &= \frac{1}{2}\left(\hat{\lambda}\mathrm{C}_s + \kappa\mathrm{C}_a\right)T_s + \frac{1}{2}\left(\hat{\lambda}\mathrm{C}_a + \kappa\mathrm{C}_s\right)T_a + \left(\mathrm{C}_s + \mathrm{C}_a\frac{\hat{\lambda}+2\gamma'_d}{\kappa}\right)F'
\end{aligned}
$$

Further expanding the coefficients

$$
\begin{aligned}
\hat{\lambda}\mathrm{C}_s + \kappa\mathrm{C}_a &= \hat{\lambda}C_\mathrm{u} - \frac{C_\mathrm{d}}{2\varepsilon\gamma'}(\hat{\lambda}^2 + 2\gamma_d'\hat{\lambda} - \kappa^2) = \hat{\lambda}C_\mathrm{u} - \frac{C_\mathrm{d}}{2\varepsilon\gamma'}(\hat{\lambda}^2 + 2\gamma_d'\hat{\lambda} - \hat{\lambda}^2 - 4\gamma_d'\lambda') \\
&= 2\frac{C_\mathrm{u}}{\varepsilon}\left(\lambda' + \frac{\varepsilon - 1}{2}\hat{\lambda}\right) \\
\hat{\lambda}\mathrm{C}_a + \kappa\mathrm{C}_s &= \kappa C_\mathrm{u} - \frac{C_\mathrm{d}}{2\varepsilon\gamma'}(\kappa\hat{\lambda} + 2\gamma_d'\kappa - \kappa\hat{\lambda}) = \kappa C_\mathrm{u} - \frac{C_\mathrm{u}}{\varepsilon}\kappa = \kappa\frac{C_\mathrm{u}}{\varepsilon}(\varepsilon - 1) \\
\mathrm{C}_s + \mathrm{C}_a\frac{\hat{\lambda} + 2\gamma_d'}{\kappa} &= C_\mathrm{u} - \frac{C_\mathrm{d}}{2\varepsilon\gamma'}(\hat{\lambda} + 2\gamma_d' - \hat{\lambda} - 2\gamma_d') = C_\mathrm{u}
\end{aligned}
$$

then the imbalance is

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

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 $T_\mathrm{u} = T_s$, then $\dot{T}_\mathrm{u} = \dot{T}_s$. Therefore, the total derivative of the planetary imbalance is

$$
\dot{N} = (\partial_t N) + (\partial_{T_s} N)\dot{T}_s + (\partial_{T_a} N)\dot{T}_a
$$

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

$$
\frac{\dot{N}}{\dot{T}_\mathrm{u}} = (\partial_t N)\frac{1}{\dot{T}_s} + (\partial_{T_s} N) + (\partial_{T_a} N)\frac{\dot{T}_a}{\dot{T}_s}
$$

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.

$$\begin{aligned}
\frac{\dot{N}}{\dot{T}_\mathrm{u}} &= \frac{C_\mathrm{u}}{\varepsilon}\left[\varepsilon\frac{\dot{F}'}{\dot{T}_s} + \left(\lambda' + \frac{\varepsilon-1}{2}\hat{\lambda}\right) + \kappa\frac{\varepsilon-1}{2}\frac{\dot{T}_a}{\dot{T}_s}\right]\frac{\lambda}{\lambda} \\
&= \left[\frac{C_\mathrm{u}}{\lambda}\frac{\dot{F}'}{\dot{T}_s} + \left(\frac{\lambda'}{\varepsilon\lambda'} + \frac{\varepsilon-1}{2\varepsilon}\frac{\hat{\lambda}}{\lambda'}\right) + \frac{\varepsilon-1}{2\varepsilon}\frac{\kappa}{\lambda'}\frac{\dot{T}_a}{\dot{T}_s}\right]\lambda
\end{aligned}$$

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

$$\begin{aligned}
\frac{\dot{N}}{\dot{T}_\mathrm{u}} &= \left\{\frac{C_\mathrm{u}}{\lambda}\frac{\dot{F}'}{\dot{T}_s} + \left[\frac{1}{\varepsilon} + \frac{\varepsilon-1}{2\varepsilon}\left(\frac{\lambda'-\varepsilon\gamma'-\gamma'_d}{\lambda'}\right)\right] + \frac{\varepsilon-1}{2\varepsilon}\frac{\kappa}{\lambda'}\frac{\dot{T}_a}{\dot{T}_s}\right\}\lambda \\
&= \left\{\frac{C_\mathrm{u}}{\lambda}\frac{\dot{F}'}{\dot{T}_s} + \left[\frac{2}{2\varepsilon} + \frac{\varepsilon-1}{2\varepsilon}\left(1-\varepsilon\frac{\gamma}{\lambda} - \frac{C_\mathrm{u}}{C_\mathrm{d}}\frac{\gamma}{\lambda}\right)\right] + \frac{\varepsilon-1}{2\varepsilon}\frac{C_\mathrm{u}\kappa}{\lambda}\frac{\dot{T}_a}{\dot{T}_s}\right\}\lambda \\
&= \left[\frac{C_\mathrm{u}}{\lambda}\frac{\dot{F}'}{\dot{T}_s} + \frac{\varepsilon+1}{2\varepsilon} - \frac{\varepsilon-1}{2\varepsilon}\left(\varepsilon+\frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\frac{\gamma}{\lambda} + \frac{\varepsilon-1}{2\varepsilon}\frac{C_\mathrm{u}\kappa}{\lambda}\frac{\dot{T}_a}{\dot{T}_s}\right]\lambda \\
&= \left[\frac{C_\mathrm{u}}{\lambda}\frac{\dot{F}'}{\dot{T}_s} + \frac{\varepsilon+1}{2\varepsilon} - \frac{\varepsilon-1}{2\varepsilon}\left(\varepsilon+\frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\frac{\gamma}{\lambda} + \frac{\varepsilon-1}{2\varepsilon}\frac{C_\mathrm{u}\kappa}{\lambda}\frac{\dot{T}_a}{\dot{T}_s}\right]\lambda \\
&= \left\{\frac{C_\mathrm{u}}{\lambda}\frac{\dot{F}'}{\dot{T}_s} + \frac{\varepsilon+1}{2\varepsilon} - \frac{\varepsilon-1}{2\varepsilon\lambda}\left[\left(\varepsilon+\frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\gamma - C_\mathrm{u}\kappa\frac{\dot{T}_a}{\dot{T}_s}\right]\right\}\lambda \\
&= \left\{\frac{C_\mathrm{u}}{\lambda}\frac{\dot{F}'}{\dot{T}_s} + \frac{\varepsilon+1}{2\varepsilon} - \frac{\varepsilon-1}{2\varepsilon\lambda}C_\mathrm{u}\kappa\left[\left(\varepsilon+\frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\frac{\gamma}{C_\mathrm{u}\kappa} - \frac{\dot{T}_a}{\dot{T}_s}\right]\right\}\lambda \\
&= \left\{\frac{C_\mathrm{u}}{\lambda}\frac{\dot{F}'}{\dot{T}_s} + \frac{\varepsilon+1}{2\varepsilon} - \frac{\varepsilon-1}{2\varepsilon}\frac{C_\mathrm{u}\kappa}{\lambda}\left[\left(\varepsilon+\frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\frac{\gamma}{C_\mathrm{u}\kappa} - \frac{\dot{T}_a}{\dot{T}_s}\right]\right\}\lambda \\
&= \left\{-\frac{C_\mathrm{u}}{|\lambda|}\frac{\dot{F}'}{\dot{T}_s} + \frac{\varepsilon+1}{2\varepsilon} + \frac{\varepsilon-1}{2\varepsilon}\frac{C_\mathrm{u}\kappa}{|\lambda|}\left[\left(\varepsilon+\frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\frac{\gamma}{C_\mathrm{u}\kappa} - \frac{\dot{T}_a}{\dot{T}_s}\right]\right\}\lambda
\end{aligned}$$

$$\frac{\dot{N}}{\dot{T}_\mathrm{u}} = \left\{-\frac{C_\mathrm{u}}{|\lambda|}\frac{\dot{F}'}{\dot{T}_s} + \frac{\varepsilon+1}{2\varepsilon}\left(1 + \frac{\varepsilon-1}{\varepsilon+1}\frac{C_\mathrm{u}\kappa}{|\lambda|}\left[\left(\varepsilon+\frac{C_\mathrm{u}}{C_\mathrm{d}}\right)\frac{\gamma}{C_\mathrm{u}\kappa} - \frac{\dot{T}_a}{\dot{T}_s}\right]\right)\right\}\lambda \tag{A15}$$

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

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

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 $\varepsilon \longrightarrow 1$, then we recover the classical linear dependence of the imbalance on T_u

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

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

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

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

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

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

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

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

where

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

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

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

where the integrals are easily computed

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

and, upon reduction, the C_i are

$$
\begin{aligned}
C_1 &= \frac{F_c'}{\lambda'} e^{-(\hat{\lambda}/2)(\tau - t_0)} \left\{ -\kappa \cosh\left[\frac{\kappa}{2}(t - t_0)\right] + (2\lambda' - \hat{\lambda}) \sinh\left[\frac{\kappa}{2}(t - t_0)\right] + \kappa e^{(\hat{\lambda}/2)(t - t_0)} \right\} \\
C_2 &= \frac{F_c'}{\lambda'} e^{-(\hat{\lambda}/2)(\tau - t_0)} \left\{ -(2\lambda' - \hat{\lambda}) \cosh\left[\frac{\kappa}{2}(t - t_0)\right] + \kappa \sinh\left[\frac{\kappa}{2}(t - t_0)\right] + (2\lambda' - \hat{\lambda}) e^{(\hat{\lambda}/2)(t - t_0)} \right\}
\end{aligned}
$$

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 $t \geq t_0$)

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

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

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)

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

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

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

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

$$\frac{\dot{T}_a}{\dot{T}_s} = \frac{\frac{\hat{\lambda}+2\gamma_d'}{\kappa}\cosh\left[\frac{\kappa}{2}(t-t_0)\right] + \sinh\left[\frac{\kappa}{2}(t-t_0)\right]}{\cosh\left[\frac{\kappa}{2}(t-t_0)\right] + \frac{\hat{\lambda}+2\gamma_d'}{\kappa}\sinh\left[\frac{\kappa}{2}(t-t_0)\right]}$$

$$= \frac{\frac{\hat{\lambda}+2\gamma_d'}{\kappa} + \tanh\left[\frac{\kappa}{2}(t-t_0)\right]}{1 + \frac{\hat{\lambda}+2\gamma_d'}{\kappa}\tanh\left[\frac{\kappa}{2}(t-t_0)\right]}$$

Formally, above result have the alternative form

$$\frac{\dot{T}_a}{\dot{T}_s} = \tanh\left[\frac{\kappa}{2}(t-t_0) + \operatorname{arctanh}\left(\frac{\hat{\lambda}+2\gamma_d'}{\kappa}\right)\right]$$

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

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

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

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

$$\hat{\lambda} + \gamma_d' \leq \lambda'$$

$$-\varepsilon\gamma' \leq 0$$

the last inequality is always true, since ε, γ' are positive constants. Thus,

$$\frac{\dot{T}_a}{\dot{T}_s} = \tanh\left[\frac{\kappa}{2}(t-t_0) + \operatorname{arctanh}\left(\frac{\hat{\lambda}+2\gamma_d'}{\kappa}\right)\right] \tag{A21}$$

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

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

Variation of the climate feedback parameter

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

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

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

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

and at the time of sign reversal

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

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

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

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

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

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, doi:10.1029/2012GL051607.

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

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, doi:10.1073/pnas.1714308114.

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, doi: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, doi: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, doi:10.1175/1520-0442(2002)015<3117:AOBEOT>2.0.CO;2.

Grose, M. R., J. Gregory, R. Colman, and T. Andrews, 2018: What Climate Sensitivity Index Is Most Useful for Projections? *Geophys. Res. Lett.*, **45 (3)**, 1559–1566, doi:10.1002/2017GL075742.

Hawkins, E., and R. Sutton, 2009: The Potential to Narrow Uncertainty in Regional Climate Predictions. *Bull. Amer. Meteor. Soc.*, **90 (8)**, 1095–1107, doi:10.1175/2009BAMS2607.1.

Jiménez-de-la-Cuesta, D., and T. Mauritsen, 2019: Emergent constraints on Earth's transient and equilibrium response to doubled CO_2 from post-1970s warming. *Nat. Geosci.*, **12 (11)**, 902–905, doi:10.1038/s41561-019-0463-y.

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

Mauritsen, T., 2016: Clouds cooled the Earth. *Nat. Geosci.*, **9 (12)**, 865–867, doi: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, doi:10.1002/2013GL058118.

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, doi: 10.1007/s00382-019-04686-4.

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

Stevens, B., S. C. Sherwood, S. Bony, and M. J. Webb, 2016: Prospects for narrowing bounds on Earth's equilibrium climate sensitivity. *Earths Future*, **4** **(11)**, 512–522, doi: 10.1002/2016EF000376.

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, doi: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, doi:10.1038/ngeo2828.

LIST OF FIGURES

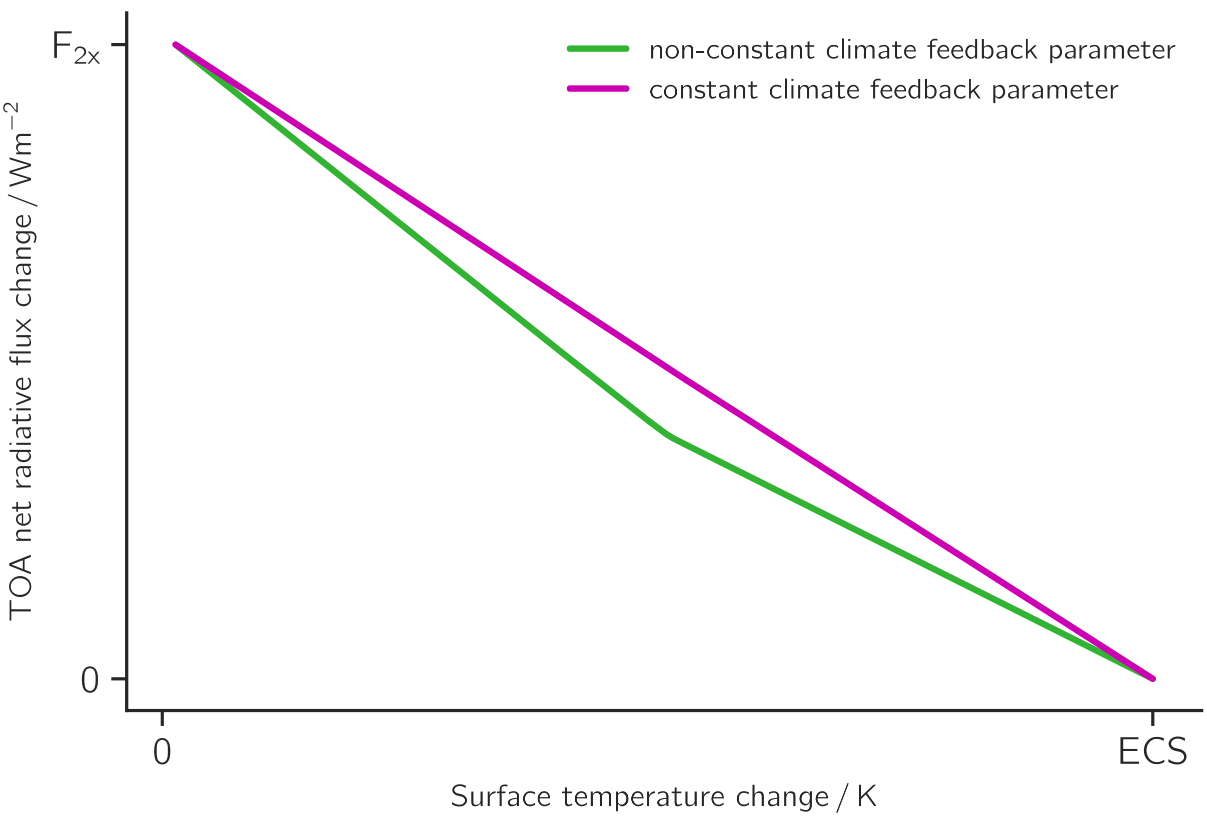

Fig. 1. Schematic representation of an $NT-$diagram for constant forcing due to a doubling of the atmospheric carbon dioxide concentration (F_{2x}). Magenta line represents the relationship between the TOA net radiatiave flux change with the surface temperature change if the feedback mechanisms on surface warming were constant (constant slope). Green line shows the case found in most models, where the slope varies throughout the process. Given that most models are not run until equilibrium, the evolving slope introduces considerable uncertainty in the equilibrium climate sensitivity (ECS) estimates

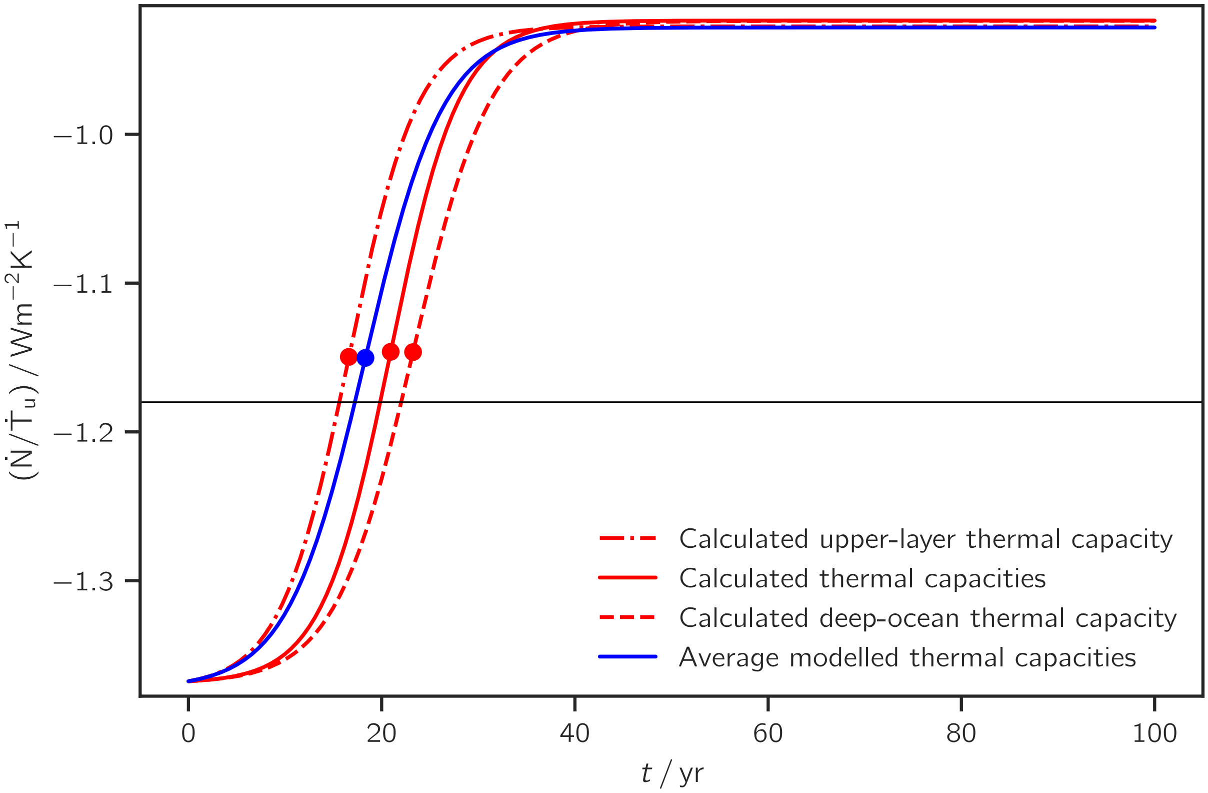

FIG. 2. Evolution of the slope of an NT−diagram. Blue solid line, with the average parameters from CMIP5 models obtained by Geoffroy et al. (2013a). Red solid line, with the thermal capacities as calculated by Jiménez-de-la-Cuesta and Mauritsen (2019). Red dashed line, with C_d as in Jiménez-de-la-Cuesta and Mauritsen (2019). Red dash-dotted line, with C_u as in Jiménez-de-la-Cuesta and Mauritsen (2019). Dots represent the slope values when the ratio term $\dot{T}_a/\dot{T}_s$ has the sign reversal. Thin black line is the constant $\lambda = -1.18\,\mathrm{W\,m^{-2}\,K^{-1}}$.

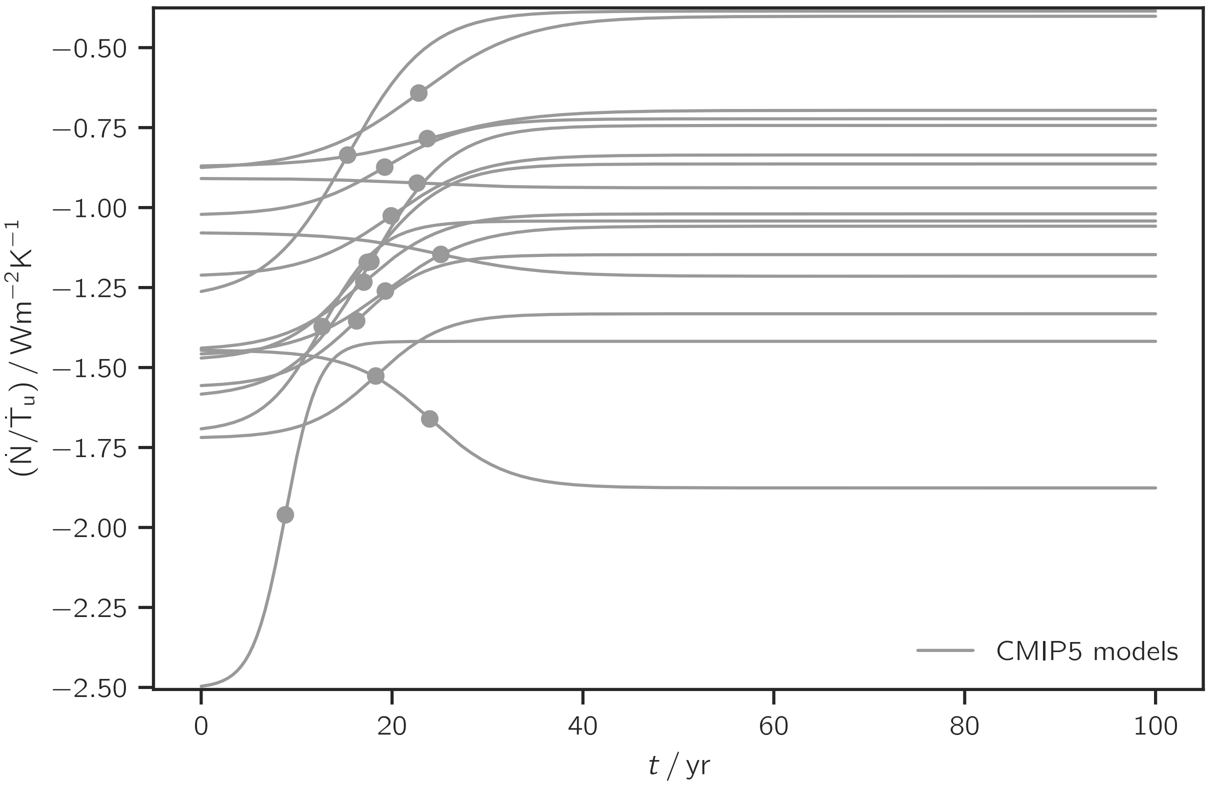

FIG. 3. Evolution of the slope of the NT−diagram. CMIP5 model behaviour using the fitted parameters presented by Geoffroy et al. (2013a). Dots indicate the time of the sign reversal. Note that three models (CNRM-CM5.1, BNU-ESM and INM-CM4) show a steepening slope instead of flattening. For these models, the fitted ε is lesser than one.