

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

3 Diego Jiménez-de-la-Cuesta

4 *Max-Planck-Institut für Meteorologie, Hamburg, Deutschland*

5 *Corresponding author:* Diego Jiménez-de-la-Cuesta, diego.jimenez@mpimet.mpg.de This work
6 has been submitted to the *Journal of Climate*. Copyright in this work may be transferred without
7 further notice.

8 ABSTRACT: When quadrupling the atmospheric CO₂ concentration in relation to pre-industrial
9 levels, most global climate models show an initially strong net radiative feedback that significantly
10 reduces the energy imbalance during the first two decades after the quadrupling. Afterwards,
11 the net radiative feedback weakens, needing more surface warming than before to reduce the
12 remaining energy imbalance. Such weakening radiative feedback has its origin in the tropical
13 oceanic stratiform cloud cover, linked to an evolving spatial warming pattern. In the classical
14 linearized energy balance framework, such variation is represented by an additional term in the
15 planetary budget equation. This additional term is usually interpreted as an ad-hoc emulation
16 of the cloud feedback change, leaving unexplained the relationship between this term and the
17 spatial warming pattern. I use a simple non-linearized energy balance framework to justify that
18 there is a physical interpretation of this term: the evolution of the spatial pattern of warming is
19 explained by changes in the ocean's circulation and energy uptake. Therefore, the global effective
20 thermal capacity of the system also changes, leading to the additional term. In reality, the clouds
21 respond to what occurs in the ocean, changing their radiative effect. In the equation, the term
22 is now a concrete representation of the ocean's role. Additionally, I derive for the first time an
23 explicit mathematical expression of the net radiative feedback and its temporal evolution in the
24 linearized energy balance framework. This mathematical expression supports the new proposed
25 interpretation. As a corollary, it justifies the twenty-year time scale used to study the variation of
26 the net radiative feedback.

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

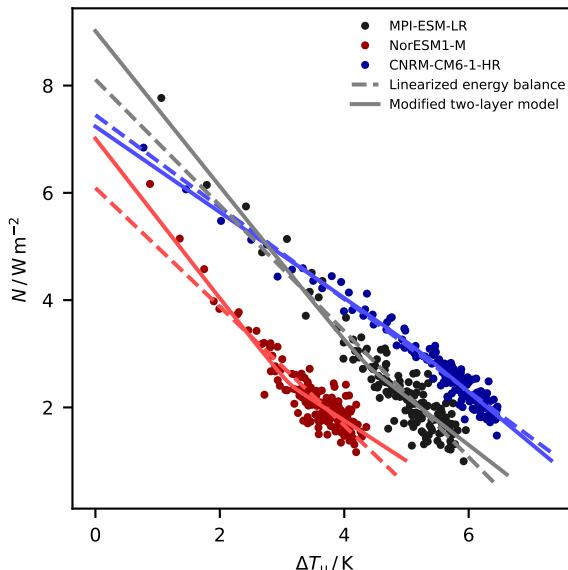
35 1. Introduction

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

$$47 \dot{E} = F + R, \tag{1}$$

48

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



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

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

65 Several studies have used equation (1) together with NT -diagrams for successfully studying the
 66 radiative response and the equilibrium climate sensitivity to CO_2 forcing (ECS) as shown in global
 67 climate models (GCMs) and the historical record (e.g., Senior and Mitchell 2000; Gregory et al.
 68 2002, 2004; Andrews et al. 2012; Otto et al. 2013; Armour et al. 2013; Armour 2017). Given the
 69 quasi-linearity found in the NT -diagrams of GCMs forced with a quadrupling of the pre-industrial
 70 atmospheric CO_2 and assuming that R is only a function of T_u , most of these studies implicitly
 71 used a Taylor series of R truncated at its first-order term (popularized by Gregory et al. 2004).
 72 Consequently, they assumed that (1) the climate state used as the basis for the Taylor series is in
 73 balance, and (2) the changes in T_u due to the CO_2 forcing are small enough to neglect higher-order

74 terms of the series. The result is

$$75 \quad \dot{E} = N \approx F + \left. \frac{dR}{dT_u} \right|_{T_u=T_u^*} \Delta T_u, \quad (2)$$

76

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

$$81 \quad N \approx F + \lambda \Delta T_u. \quad (3)$$

82

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

91 Some authors extended the framework of equation (3) to accommodate this effect (Held et al.
 92 2010; Winton et al. 2010; Geoffroy et al. 2013b,a). First they introduced two layers: a) the upper
 93 layer that includes the atmosphere and the ocean's mixed layer, and b) the deep ocean's layer.
 94 Therefore, the state variables are now the surface (T_u) and the deep-ocean (T_d) temperatures. These
 95 two layers greatly differ on thermal capacities, introducing two timescales: the fast upper layer and
 96 the slow deep layer. They connected both layers with the deep ocean's energy uptake (H), which
 97 should depend on $T_u - T_d$. However, they also introduced a perturbed energy uptake in the upper
 98 layer H' to account for the change in the radiative response

$$99 \quad \begin{cases} N_u \approx F + \lambda \Delta T_u - H' \\ N_d \approx H \end{cases} \quad (4)$$

$$100 \quad N = N_u + N_d \approx F + \lambda \Delta T_u - (H' - H), \quad (5)$$

101

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

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

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

132 shows that a changing effective thermal capacity explains better the variation of the net feedback,
133 even in the case of equation (4). As a corollary, I show that the twenty-year timescale for evaluating
134 the pattern effect can be justified by the expression I have derived.

135 2. Theory

136 a. Non-linear framework

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

$$N = (1 - \alpha)S + G - \epsilon\sigma(fT_u)^4, \quad (6)$$

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

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

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

169 Precisely, the planetary thermal capacity is present in the energy content: $E = CT_u$, where C has
 170 units of $\text{J m}^{-2} \text{K}^{-1}$. If C is constant, then $\dot{E} = C\dot{T}_u = N$. Defining $N := C\dot{T}_u$ and introducing the
 171 varying C results in $\dot{E} = C\dot{T}_u + \dot{C}T_u = N + \dot{C}T_u$. Thus, the planetary energy budget has the new
 172 form

$$N = (1 - \alpha)S + G - \epsilon\sigma(fT_u)^4 - \dot{C}T_u, \quad (7)$$

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

$$\dot{C}T_u \sim H' - H$$

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

177 *b. The modified linearized two-layer model*

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

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

$$\begin{cases} C_u \frac{d \Delta T_u}{dt} = F + \lambda \Delta T_u - H' \\ C_d \frac{d \Delta T_d}{dt} = H \end{cases} \quad (8)$$

183 where the first equation corresponds to the upper-layer budget and the second equation to the deep
 184 layer. The climate feedback parameter λ has units of $\text{W m}^{-2} \text{K}^{-1}$. H is the ocean energy uptake
 185 approximated by $H \approx \gamma(\Delta T_u - \Delta T_d)$, where γ is the rate of the deep-ocean energy uptake in
 186 $\text{W m}^{-2} \text{K}^{-1}$. H' is the perturbed energy uptake such that $H' = \hat{\varepsilon}H$, where $\hat{\varepsilon}$ is the non-dimensional
 187 efficacy of the deep-ocean energy uptake: a measure of the pattern effect. Geoffroy et al. (2013a)
 188 consider $\hat{\varepsilon}$ constant. C_u and C_d are respectively the (fixed) thermal capacities of the upper and
 189 deep layers in $\text{J m}^{-2} \text{K}^{-1}$. All these parameters in equations (8) are valid in a neighborhood of the
 190 reference climate state (T_u^*, T_d^*) , for their values are the ones taken about this state. ΔT_u and ΔT_d
 191 are the temperature anomalies referred to (T_u^*, T_d^*) .

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

$$\begin{cases} \frac{d \Delta T_u}{dt} = F' + \lambda' \Delta T_u - \hat{\varepsilon} \gamma' (\Delta T_u - \Delta T_d) \\ \frac{d \Delta T_d}{dt} = \gamma'_d (\Delta T_u - \Delta T_d) \end{cases} \quad (9)$$

193 where $F' := F/C_u$ with units of K s^{-1} and, $\lambda' := \lambda/C_u$, $\gamma' := \gamma/C_u$ and $\gamma'_d := \gamma/C_d$ with
 194 units of s^{-1} . Equations (9) are a system of linear ordinary differential equations (Geoffroy et al.
 195 2013a; Rohrschneider et al. 2019). Although the solutions are standard and widely discussed in
 196 other articles (e.g. Geoffroy et al. 2013a; Rohrschneider et al. 2019), here I will use the normal
 197 mode approach. The solutions written in terms of the normal modes are more elegant and ease
 198 the algebraic transformations. In the following, I summarize the relevant facts, leaving the full
 199 mathematical discussion in the appendix A of this article.

200 The homogeneous ($F' \equiv 0$) version of the system (9) has two distinct eigenvalues $\mu_{\pm} := (\hat{\lambda} \pm \kappa)/2$,
 201 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,
 202 forming a basis in which the full system (9) is uncoupled and, therefore, has a straight-forward
 203 solution. The eigensolutions ΔT_{\pm} are the solutions associated with each eigenvalue. Afterwards,

204 one can return to the original representation, finding that ΔT_u and ΔT_d are linear combinations of
 205 ΔT_{\pm} . These linear combinations are the normal modes: the symmetric mode $\Delta T_s := \Delta T_+ + \Delta T_-$
 206 and the antisymmetric mode $\Delta T_a := \Delta T_+ - \Delta T_-$. The main result of this process is that

$$\begin{cases} \Delta T_u = & \Delta T_s \\ \Delta T_d = -\frac{\hat{\lambda} + 2\gamma'_d}{2\hat{\epsilon}\gamma'}\Delta T_s + \frac{\kappa}{2\hat{\epsilon}\gamma'}\Delta T_a \end{cases} \quad (10)$$

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

208 Let us define $\Delta E = E - E^*$, where $E^* = C^*T_u^*$ and $C = C^* + \Delta C$. C^* is the planetary thermal
 209 capacity at the reference climate state (T_u^*, T_d^*) . Additionally, we postulate that the total change in
 210 the planetary energy only comes from F and the original R

$$\frac{d \Delta E}{dt} = \dot{E} \approx F + \lambda \Delta T_u. \quad (11)$$

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

$$\begin{aligned} C_u \frac{d \Delta T_u}{dt} + C_d \frac{d \Delta T_d}{dt} &= \dot{E} - (H' - H) \therefore \\ \dot{E} &= C_u \frac{d \Delta T_u}{dt} + H + (H' - H) \\ C^* \frac{d \Delta T_u}{dt} + \Delta C \frac{d \Delta T_u}{dt} + \frac{d \Delta C}{dt} T_u &= C_u \frac{d \Delta T_u}{dt} + \hat{\epsilon} H \\ C^* + \Delta C + \frac{d \Delta C}{dt} \frac{T_u}{\frac{d \Delta T_u}{dt}} &= C_u + \hat{\epsilon} C_d \frac{\frac{d \Delta T_d}{dt}}{\frac{d \Delta T_u}{dt}} \\ C^* + \Delta C + \frac{d \Delta C}{dt} \frac{T_u}{\frac{d \Delta T_u}{dt}} &= \left(C_u - \frac{\hat{\lambda} + 2\gamma'_d}{2\gamma'} C_d \right) + \frac{\kappa}{2\gamma'} C_d \frac{\frac{d \Delta T_a}{dt}}{\frac{d \Delta T_u}{dt}}. \end{aligned} \quad (12)$$

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

$$C^* = C_u - \frac{\hat{\lambda} + 2\gamma'_d}{2\gamma'} C_d, \quad (13)$$

$$\frac{d \Delta C}{dt} + \frac{d \Delta T_u}{T_u} \Delta C = \frac{\kappa}{2\gamma'} C_d \frac{d \Delta T_a}{T_u}. \quad (14)$$

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

222 *d. Net radiative feedback expression*

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

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

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

$$\mathcal{F}_{\text{for}} = -\frac{1}{|\lambda|} \frac{\dot{F}}{d\Delta T_s/dt}, \quad (16)$$

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

$$\mathcal{F}_{\text{pat, stat}} = C_u \frac{\gamma}{|\lambda|} \left(\frac{\hat{\varepsilon}}{C_u} + \frac{1}{C_d} \right), \quad (18)$$

$$\mathcal{F}_{\text{pat, dyn}} = C_u \frac{\kappa}{|\lambda|} \frac{d\Delta T_a/dt}{d\Delta T_s/dt}. \quad (19)$$

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

- 232 • The forcing component (16) simply compares the evolution of F with the evolution of the
 233 surface temperature change, given that $\Delta T_u = \Delta T_s$ (first equation of system (10)). This
 234 component only contributes if the forcing is time-varying.
- 235 • The response component (17) is constant and will only give a correction to the original λ if
 236 $\hat{\varepsilon} \neq 1$.
- 237 • The pattern effect component is only active if $\hat{\varepsilon} \neq 1$. In case it is active, we have the
 238 contributions of the static and dynamical terms.
 - 239 1. The static term (18) has three factors. One of them is a sum of the inverse of the
 240 thermal capacities of the system. This arrangement is similar to the inverse of the total
 241 capacitance of electric capacitors in series. Therefore, it can be interpreted as the effect
 242 of the initial state of the ocean energy distribution as discussed for equation (7).
 - 243 2. The dynamical term (19) has the ratio of the time derivatives of ΔT_s and ΔT_a , explicitly
 244 relating this term to the expression of the time-dependent planetary thermal capacity in
 245 Equation (14).

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

252 3. Results

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

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

$$\mathcal{F}_{\text{for}} = 0, \quad (20)$$

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

$$\mathcal{F}_{\text{pat, stat}} = C_u \frac{\gamma}{|\lambda|} \left(\frac{\hat{\varepsilon}}{C_u} + \frac{1}{C_d} \right), \quad (22)$$

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

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

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

261 The time-evolving part of equation (23) is an hyperbolic tangent. A plain hyperbolic tangent,
 262 $\tanh(t)$, is a monotonically increasing s-shaped or sigmoidal curve, and its possible values are
 263 between -1 and 1 , crossing zero at $t = 0$. The extreme values -1 and 1 are asymptotes. Leaving

264 out the term $\operatorname{arctanh}(Z)$, our function is similar to $\tanh[(\kappa/2)(t - t_0)]$. This function still has -1
 265 and 1 as asymptotes but crosses zero at $t = t_0$. Depending on the value of $\kappa > 0$, the evolution
 266 between asymptotes would be faster. If κ were very large, the function would resemble a step
 267 function. The smaller the κ , the gentle the change of the function between asymptotes. Therefore
 268 $\kappa/2$ is a scaling factor. We conclude the analysis by adding $\operatorname{arctanh}(Z)$. This term shifts the
 269 argument of the hyperbolic tangent. If we evaluate $\tanh[(\kappa/2)(t - t_0) + \operatorname{arctanh}(Z)]$ at t_0 , we obtain
 270 $\tanh(\operatorname{arctanh}(Z)) = Z < 0$. Therefore, the zero crossing is not anymore at t_0 but at a posterior time
 271 and the value of the function at t_0 is negative. I call time of sign reversal (t_{rev}) to the new time
 272 where the function becomes zero. This time is

$$t_{\text{rev}} = t_0 + \frac{2}{\kappa} \operatorname{arctanh} |Z|. \quad (25)$$

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

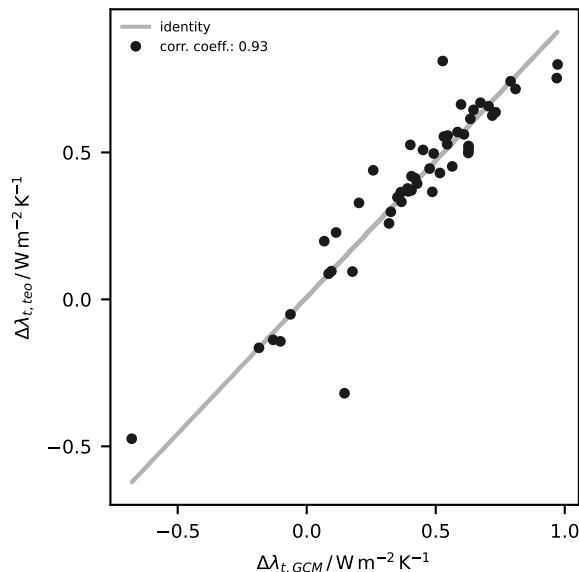
274 Since in $t_0 < t < t_{\text{rev}}$ $\mathcal{F}_{\text{pat, dyn}}$ is negative, the dynamical component strengthens the net radiative
 275 feedback, as \mathcal{F}_{pat} will be larger than without the dynamical component. Nonetheless, the net
 276 radiative feedback still becomes less negative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

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

288 Following the method shown by Geoffroy et al. (2013a), I calculate the thermal, circulation and
 289 radiative parameters of the modified linearized two-layer model for a selection of 52 models of
 290 the phases 5 and 6 of the climate model inter-comparison project (CMIP). The ensemble means
 291 are in table 1. Using equation (15) and the estimated parameters, the theoretical change in the net
 292 radiative feedback $\Delta\lambda_t = \lambda_t(150 \text{ yr}) - \lambda_t(1 \text{ yr})$ is calculated. It is compared with the difference in
 293 the slopes obtained from the regressions of N on T from the first twenty years, and from the years
 294 21 to 150. Figure 2 shows that the theoretical expression simulates correctly the change in the net
 295 radiative feedback ($r = 0.93$).

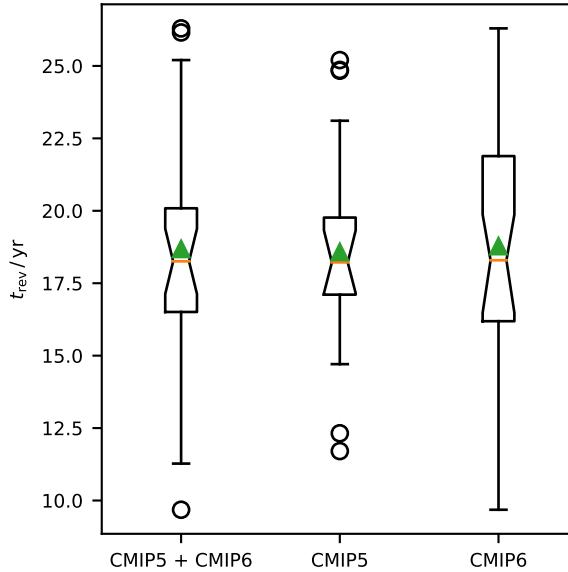
Ensemble	$F / \text{W m}^{-2}$	$C / \text{W yr m}^{-2} \text{K}^{-1}$		$\lambda / \text{W m}^{-2} \text{K}^{-1}$	$\gamma / \text{W m}^{-2} \text{K}^{-1}$	$\hat{\varepsilon} / 1$	$t_{\text{rev}} / \text{yr}$
		C_u	C_d				
CMIP5	7.52	8.53	105.17	-1.21	0.68	1.26	18.53
CMIP6	7.48	8.06	95.88	-1.02	0.66	1.30	18.31

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



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

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



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

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

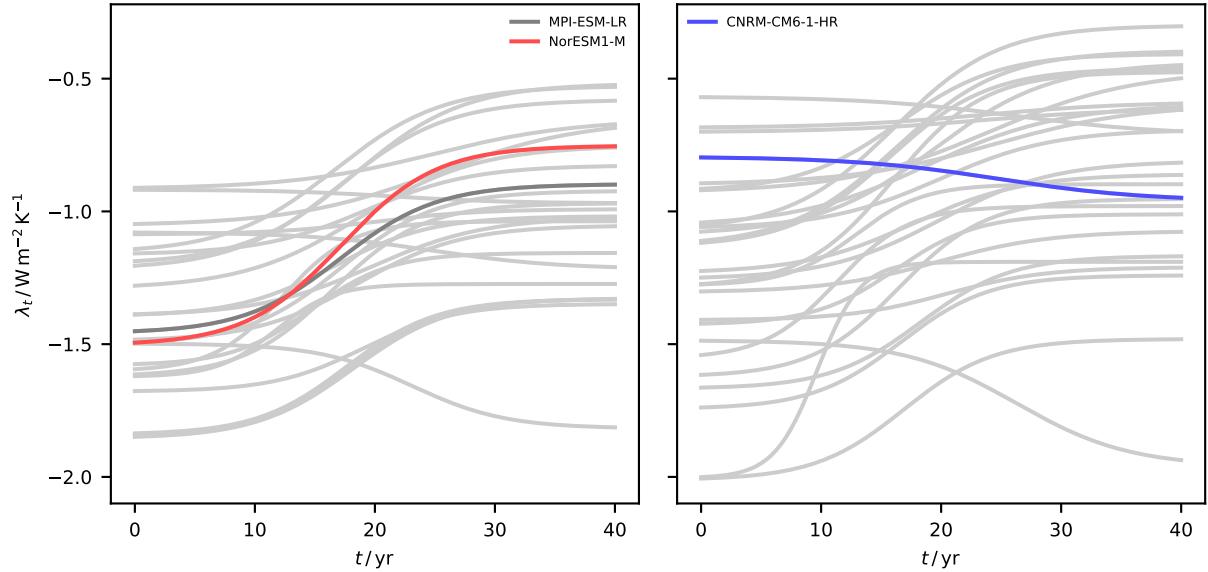


FIG. 4. Theoretical evolution of the radiative feedback. Left: CMIP5 ensemble. Right: CMIP6 ensemble

319 4. Analysis and Discussion

320 Winton et al. (2010) have already proposed that an efficacy in the deep-ocean energy uptake
 321 would be equivalent to changing the thermal capacity of the deep-ocean layer, as Geoffroy et al.
 322 (2013a) also noted. The initial discussion of the non-linear planetary energy budget (Equation
 323 7) and the expression for the dynamical planetary thermal capacity in the linearized framework
 324 (Equation 14) show how natural is the concept of a varying planetary thermal capacity, even without
 325 pattern effect. When there is a pattern effect, then the relationship between surface and deep-ocean
 326 temperatures changes, and the planetary thermal capacity evolves in a different manner.

327 When looking at the expression of λ_t (Equation 15), the time-varying term $\mathcal{F}_{\text{pat, dyn}}$ (Equation 19)
 328 has the same time varying term as in the dynamical planetary thermal capacity (Equation 14). This
 329 fact directly connects the varying net radiative feedback to the dynamical planetary thermal capacity.
 330 The influence of the $\mathcal{F}_{\text{pat, dyn}}$ term only appears when the pattern effect is active. In contrast, the
 331 classical interpretation $H' - H$ as a peculiar radiative feedback mechanism led to inconsistencies:
 332 the more serious was about energy conservation. It also left unexplained the origin of the warming
 333 pattern and how a spatial pattern could explain a global effect. The dynamical thermal capacity
 334 interpretation closes the energy inconsistencies and connects naturally spatial warming patterns

335 to energy distributions in layers, marking a possible course towards understanding the warming
336 pattern and why it is different between contemporary GCMs and observations (e.g., Wills et al.
337 2022).

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

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

361 In the results, I show that the estimates of the thermal, circulation, and radiative parameters can
362 have a substantial effect on the evolution of the net radiative feedback. In light of the discussion,
363 particularly the thermal capacities and the rate of deep-ocean energy uptake γ represent an initial
364 energy distribution about the reference state (T_u^*, T_d^*) . This energy distribution evolves differently,

365 depending on the magnitude of the deep-ocean energy uptake efficacy $\hat{\epsilon}$. This parameter represents
366 the magnitude of the coupling of the energy distribution and the surface temperature. Thus, it should
367 be related directly to physical quantities, e.g., the ocean stratification in the regions of upwelling
368 of deep-water formation. In consequence, GCMs will show diverse behaviors for the variation of
369 the net radiative feedback as their initial energy uptake and the rate at which it changes with ocean
370 circulation widely varies (Kiehl 2007). Perhaps, this diversity in GCMs is part of the reason why
371 GCMs cannot fully reproduce the observed warming pattern (e.g., Wills et al. 2022). This fact is
372 worrying, given that our climate change projections can be biased low.

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

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

383 Therefore, some details in the theoretical evolution of the radiative feedback (Figure 4) can be
384 different between the complex models and nature. Nonetheless, these limitations should be the
385 starting point to find what are the actual relationships between the evolving spatial warming pattern
386 and the energy distribution in the ocean. For that end, one should use observations, Earth system
387 model output, new experiments tailored to isolate mechanisms, and other simplified models for
388 specific mechanisms. This process will help to put in context equation (14) and possibly reveal that
389 $\hat{\epsilon}$ is not constant, relaxing the constraint imposed and providing further information on its physics.
390 Such uses of the conceptual frameworks have been useful in related problems, and there are recent
391 advances (e.g. Datseris et al. 2022). Thus, checking when the assumptions of the conceptual
392 models break and understanding the reasons advance us towards a better conceptual understanding
393 of the climate system.

394 In my analysis of the two-layer model, the dependence of the variation of the net radiative feedback
395 with the strength of forcing (Senior and Mitchell 2000; Meraner et al. 2013; Rohrschneider et al.
396 2019) is missing. However, such dependence should come from the values of $\hat{\epsilon}$, λ , γ , and the
397 thermal capacities under a particular forcing and, probably, non-linearities. We should always
398 remember that the thermal capacities, λ , and γ are only approximations of the actual quantities in
399 the neighborhood of the starting states. Therefore, we need a consistent theory on how the different
400 types and magnitudes of forcing modify (a) the coupling between ocean energy distribution and
401 surface temperature, (b) the atmospheric radiative feedback mechanisms, and (c) the rate of energy
402 uptake. Such a theory should describe the Earth system not in the tiny details or as an aggregate of
403 separate disciplines but as an integrated system. The idea can be better expressed as the difference
404 between describing a tree as an aggregate of cells of different types with different functions; and
405 describing the whole tree in terms of certain characteristic variables. In the best case, the needed
406 theory for the climate is incomplete. However, having such a basic conceptual theory of climate will
407 help us better interpret complex model results, find more hidden relationships between important
408 variables and, possibly, reduce the uncertainty in observational estimates of climate sensitivity.

409 **5. Conclusions**

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

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

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

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

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

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

440 APPENDIX A

441 **Mathematical analysis of the modified two-layer model**

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

449 The appendix is written in an exhaustive way and I leave few things without development. The
450 cases in which I do not show some algebraic step is because the necessary step has been already

451 done or is very simple. For simplicity ΔT_u and ΔT_d are always rewritten as T_u and T_d for the
 452 two-layer model.

453 **Matrix form of the equations**

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

$$\begin{aligned}
 455 \quad N_u &= C_u \dot{T}_u = F + \lambda T_u - \hat{\varepsilon} \gamma (T_u - T_d) \\
 456 \quad N_d &= C_d \dot{T}_d = \gamma (T_u - T_d)
 \end{aligned} \tag{A1}$$

457 and the planetary imbalance is $N = N_u + N_d$. I present another form of the equations, where I
 458 divide by the thermal capacities.

$$\begin{aligned}
 459 \quad \dot{T}_u &= \frac{F}{C_u} + \frac{\lambda}{C_u} T_u - \hat{\varepsilon} \frac{\gamma}{C_u} (T_u - T_d) \\
 460 \quad \dot{T}_d &= \frac{\gamma}{C_d} (T_u - T_d)
 \end{aligned}$$

461 If I define $F' := F/C_u$, $\lambda' := \lambda/C_u$, $\gamma' := \gamma/C_u$, $\gamma'_d := \gamma/C_d$, one can write the equations in a
 462 lean way

$$\begin{aligned}
 463 \quad \dot{T}_u &= F' + \lambda' T_u - \hat{\varepsilon} \gamma' (T_u - T_d) \\
 464 \quad \dot{T}_d &= \gamma'_d (T_u - T_d)
 \end{aligned} \tag{A2}$$

465 I will put the system in matrix form. I define $\mathbf{T} := (T_u, T_d)$, $\mathbf{F}' := (F', 0)$ and

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

468 and the system can be written

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

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

472 **Eigenvalues and eigenvectors**

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

$$475 \quad (\lambda' - \hat{\varepsilon}\gamma' - \mu)(-\gamma'_d - \mu) - \hat{\varepsilon}\gamma'\gamma'_d = 0 \quad (A5)$$

477

$$478 \quad -\lambda'\gamma'_d + \hat{\varepsilon}\gamma'\gamma'_d + \mu\gamma'_d - \lambda'\mu + \hat{\varepsilon}\gamma'\mu + \mu^2 - \hat{\varepsilon}\gamma'\gamma'_d = 0$$

478

$$479 \quad -\lambda'\gamma'_d + \mu\gamma'_d - \lambda'\mu + \hat{\varepsilon}\gamma'\mu + \mu^2 = 0$$

479

$$480 \quad -\lambda'\gamma'_d - (\lambda' - \hat{\varepsilon}\gamma' - \gamma'_d)\mu + \mu^2 = 0$$

480

481

482 The solutions of equation (A5) are

$$483 \quad \mu = \frac{(\lambda' - \hat{\varepsilon}\gamma' - \gamma'_d) \pm [(\lambda' - \hat{\varepsilon}\gamma' - \gamma'_d)^2 + 4\lambda'\gamma'_d]^{1/2}}{2} \quad (A6)$$

483

484

485 and, given that in the Earth $C_u < C_d$, one can prove that there are two real and different eigenval-
 486 ues. One needs to check that the square root term is not complex or zero. This only happens if the
 487 sum within the square root is negative or zero

488

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

489

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

490

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

491

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

492

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

493

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

494

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

496

$$(\lambda'/\gamma'_d)^2 + (\hat{\varepsilon}(C_d/C_u) + 1)^2 \leq 2(\lambda'/\gamma'_d)[\hat{\varepsilon}(C_d/C_u) - 1]$$

497 In the last inequality, the left-hand side is always positive. The right-hand side depends on the
 498 sign of the factors. The middle factor is negative since λ' is negative and γ'_d is positive. The third

499 factor is positive provided that $\hat{\varepsilon} > C_u/C_d$. Given that $\hat{\varepsilon} \geq 1$ and $C_u < C_d$, then the third factor
500 is positive in our case. Then the right-hand side is negative. Thus, we obtained a contradiction
501 by supposing that the square root term was negative or zero. Therefore, the conclusion is that
502 the eigenvalues are two real and distinct numbers. Some CMIP5 models show $\hat{\varepsilon} < 1$ according
503 to Geoffroy et al. (2013a). These also fit here. In the last condition of the above expression we
504 require that $\hat{\varepsilon}(C_d/C_u) - 1 > 0$. If $\hat{\varepsilon} \geq C_u/C_d$ this is fulfilled. C_u/C_d is a small quantity and, in the
505 models that have a lesser than one $\hat{\varepsilon}$, always the $\hat{\varepsilon}$ is larger than this small quantity by an order of
506 magnitude. Thus, what I had said until now and will be said afterwards applies to all cases.

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

$$\mu_{\pm} = \frac{\hat{\lambda} \pm \kappa}{2} \quad (\text{A7})$$

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

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

517 and then the matrices for each eigenvalue have the form

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

521 Since eigenvalues are real and distinct, there should be two linearly-independent eigenvectors,
522 one for each eigenvalue. These vectors should fulfill that $\mathbf{e}_{\pm}(\mathbf{A} - \mu_{\pm} \text{id}) = 0$. Solving that linear

523 system, I find the eigenvectors in temperature representation

$$524 \quad \mathbf{e}_{\pm} = \mathbf{e}_u - \frac{\mu_{\mp} + \gamma'_d}{\hat{\varepsilon}\gamma'} \mathbf{e}_d \quad (A8)$$

525

526 The procedure to get the result is to solve the system of homogeneous linear equations $\mathbf{e}_{\pm}(\mathbf{A} -$
527 $\mu_{\pm} \text{id}) = 0$

$$528 \quad \begin{cases} (\mu_{\mp} + \gamma'_d)e_{\pm,u} & + \hat{\varepsilon}\gamma' e_{\pm,d} = 0 \\ \gamma'_d e_{\pm,u} + [\mu_{\mp} - (\lambda' - \hat{\varepsilon}\gamma')]e_{\pm,d} = 0 \end{cases}$$

529

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

$$531 \quad e_{\pm,d} = -\frac{\mu_{\mp} + \gamma'_d}{\hat{\varepsilon}\gamma'} e_{\pm,u} \longrightarrow$$

$$532 \quad \left(\gamma'_d - \frac{[\mu_{\mp} - (\lambda' - \hat{\varepsilon}\gamma')](\mu_{\mp} + \gamma'_d)}{\hat{\varepsilon}\gamma'} \right) e_{\pm,u} = 0$$

$$533 \quad \frac{\hat{\varepsilon}\gamma'\gamma'_d - [\mu_{\mp} - (\lambda' - \hat{\varepsilon}\gamma')](\mu_{\mp} + \gamma'_d)}{\hat{\varepsilon}\gamma'} e_{\pm,u} = 0, (\hat{\varepsilon}, \gamma' \neq 0) \therefore$$

534

$$535 \quad \{ \hat{\varepsilon}\gamma'\gamma'_d - [\mu_{\mp} - (\lambda' - \hat{\varepsilon}\gamma')](\mu_{\mp} + \gamma'_d) \} e_{\pm,u} = 0$$

536

$$537 \quad \{ \hat{\varepsilon}\gamma'\gamma'_d + [(\lambda' - \hat{\varepsilon}\gamma') - \mu_{\mp}](\gamma'_d + \mu_{\mp}) \} e_{\pm,u} = 0$$

$$538 \quad - \{ -\hat{\varepsilon}\gamma'\gamma'_d + [(\lambda' - \hat{\varepsilon}\gamma') - \mu_{\mp}](-\gamma'_d - \mu_{\mp}) \} e_{\pm,u} = 0$$

539

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

$$544 \quad e_{\pm,u} = \alpha$$

$$545 \quad e_{\pm,d} = -\frac{\mu_{\mp} + \gamma'_d}{\hat{\varepsilon}\gamma'} \alpha$$

546

547 or as a vector in the temperature basis

$$548 \quad \mathbf{e}_{\pm} = e_{\pm,u} \mathbf{e}_u + e_{\pm,d} \mathbf{e}_d$$

$$549 \quad \mathbf{e}_{\pm} = \alpha \mathbf{e}_u - \frac{\mu_{\mp} + \gamma'_d}{\hat{\epsilon} \gamma'} \alpha \mathbf{e}_d$$

550

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

$$553 \quad \mathbf{e}_{\pm} = \mathbf{e}_u - \frac{\mu_{\mp} + \gamma'_d}{\hat{\epsilon} \gamma'} \mathbf{e}_d$$

554

555 which is the same as the equation (A8).

556 Now, I can derive the expressions of the temperature basis vectors in terms of the two eigenvectors.

557 If one solves for e_u in equation (A8)

$$558 \quad \mathbf{e}_{\pm} + \frac{\mu_{\mp} + \gamma'_d}{\hat{\epsilon} \gamma'} \mathbf{e}_d = \mathbf{e}_u$$

559

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

$$561 \quad \mathbf{e}_- + \frac{\mu_+ + \gamma'_d}{\hat{\epsilon} \gamma'} \mathbf{e}_d = \mathbf{e}_+ + \frac{\mu_- + \gamma'_d}{\hat{\epsilon} \gamma'} \mathbf{e}_d$$

$$562 \quad \left(\frac{\mu_+ + \gamma'_d}{\hat{\epsilon} \gamma'} - \frac{\mu_- + \gamma'_d}{\hat{\epsilon} \gamma'} \right) \mathbf{e}_d = \mathbf{e}_+ - \mathbf{e}_-$$

$$563 \quad \frac{(\mu_+ + \gamma'_d) - (\mu_- + \gamma'_d)}{\hat{\epsilon} \gamma'} \mathbf{e}_d = \mathbf{e}_+ - \mathbf{e}_-$$

$$564 \quad \frac{\mu_+ - \mu_-}{\hat{\epsilon} \gamma'} \mathbf{e}_d = \mathbf{e}_+ - \mathbf{e}_-$$

$$565 \quad \mathbf{e}_d = \frac{\hat{\epsilon} \gamma'}{\mu_+ - \mu_-} (\mathbf{e}_+ - \mathbf{e}_-)$$

566

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

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

$$569 \quad \mathbf{e}_+ + \frac{\mu_- + \gamma'_d}{\hat{\epsilon}\gamma'} \mathbf{e}_d = \mathbf{e}_u$$

$$570 \quad \mathbf{e}_+ + \frac{\mu_- + \gamma'_d}{\hat{\epsilon}\gamma'} \frac{\hat{\epsilon}\gamma'}{\mu_+ - \mu_-} (\mathbf{e}_+ - \mathbf{e}_-) = \mathbf{e}_u$$

$$571 \quad \mathbf{e}_+ + \frac{\mu_- + \gamma'_d}{\mu_+ - \mu_-} (\mathbf{e}_+ - \mathbf{e}_-) = \mathbf{e}_u$$

$$572 \quad \left(1 + \frac{\mu_- + \gamma'_d}{\mu_+ - \mu_-}\right) \mathbf{e}_+ - \frac{\mu_- + \gamma'_d}{\mu_+ - \mu_-} \mathbf{e}_- = \mathbf{e}_u$$

$$573 \quad \frac{\mu_+ - \mu_- + \mu_- + \gamma'_d}{\mu_+ - \mu_-} \mathbf{e}_+ - \frac{\mu_- + \gamma'_d}{\mu_+ - \mu_-} \mathbf{e}_- = \mathbf{e}_u$$

$$574 \quad \frac{\mu_+ + \gamma'_d}{\mu_+ - \mu_-} \mathbf{e}_+ - \frac{\mu_- + \gamma'_d}{\mu_+ - \mu_-} \mathbf{e}_- = \mathbf{e}_u$$

575

576 and the temperature basis vectors in the eigenvector representation are

$$577 \quad \mathbf{e}_u = \frac{\mu_+ + \gamma'_d}{\mu_+ - \mu_-} \mathbf{e}_+ - \frac{\mu_- + \gamma'_d}{\mu_+ - \mu_-} \mathbf{e}_-$$

$$578 \quad \mathbf{e}_d = \frac{\hat{\epsilon}\gamma'}{\mu_+ - \mu_-} (\mathbf{e}_+ - \mathbf{e}_-) \quad (\text{A9})$$

579 **Matrix in the eigenvector representation. Solutions**

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

$$582 \quad \mathbf{B} = \begin{pmatrix} \mu_+ & 0 \\ 0 & \mu_- \end{pmatrix} \quad (\text{A10})$$

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

$$588 \quad \mathbf{e}_i = \Lambda_i^\alpha \mathbf{e}_\alpha$$

589

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

$$592 \quad \mathbf{e}_\alpha = \Theta_\alpha^i \mathbf{e}_i$$

593

594 where matrix Θ has in its rows the coordinates the eigenvector basis in the temperature represen-
595 tation. This means that

$$596 \quad \mathbf{e}_\alpha = \Theta_\alpha^i \mathbf{e}_i = \Theta_\alpha^i \Lambda_i^\beta \mathbf{e}_\beta$$

597

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

$$599 \quad \mathbf{e}_\alpha = \delta_\alpha^\beta \mathbf{e}_\beta = \mathbf{e}_\alpha$$

600

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

602 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
603 the temperature representation, then the action of the linear operator f should be $f(\mathbf{v}) = f(v^j \mathbf{e}_j) =$
604 $v^j f(\mathbf{e}_j)$. Then the action of f on a vector expressed in a given basis only depends on the action
605 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
606 rows the coordinates in the temperature representation of the action of f over each basis vector.
607 Once one understands what is happening under the hood, what we want is the matrix \mathbf{B} , which
608 is the representation of f in the eigenvector basis. Therefore, I begin with the basic relationship
609 in the temperature representation and introduce the change of representation using the alternative

610 representation of equations (A8) and (A9)

$$\begin{aligned}
611 \quad & f(\mathbf{e}_i) = \mathbf{A}_i^j \Lambda_j^\zeta \mathbf{e}_\zeta \\
612 \quad & f(\Lambda_i^\alpha \mathbf{e}_\alpha) = \mathbf{A}_i^j \Lambda_j^\zeta \mathbf{e}_\zeta \\
613 \quad & \Lambda_i^\alpha f(\mathbf{e}_\alpha) = \mathbf{A}_i^j \Lambda_j^\zeta \mathbf{e}_\zeta \\
614 \quad & (\Lambda^{-1})_\beta^i \Lambda_i^\alpha f(\mathbf{e}_\alpha) = (\Lambda^{-1})_\beta^i \mathbf{A}_i^j \Lambda_j^\zeta \mathbf{e}_\zeta \\
615 \quad & f(\mathbf{e}_\beta) = (\Lambda^{-1})_\beta^i \mathbf{A}_i^j \Lambda_j^\zeta \mathbf{e}_\zeta, f(\mathbf{e}_\beta) := \mathbf{B}_\beta^\zeta \mathbf{e}_\zeta \\
616 \quad & \mathbf{B}_\beta^\zeta = (\Lambda^{-1})_\beta^i \mathbf{A}_i^j \Lambda_j^\zeta \\
617 \quad &
\end{aligned}$$

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

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

621 First, note that $\mu_+ - \mu_- = \kappa$. One also looks at the following quantities that will help in the
622 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
623 first product, $\Lambda^{-1} \mathbf{A}$.

$$\begin{aligned}
624 \quad & \Lambda^{-1} \mathbf{A} = \begin{pmatrix} 1 & -\frac{\mu_- + \gamma'_d}{\hat{\varepsilon} \gamma'} \\ 1 & -\frac{\mu_+ + \gamma'_d}{\hat{\varepsilon} \gamma'} \end{pmatrix} \begin{pmatrix} \hat{\lambda} + \gamma'_d & \gamma'_d \\ \hat{\varepsilon} \gamma' & -\gamma'_d \end{pmatrix} \\
625 \quad & = \begin{pmatrix} \hat{\lambda} + \gamma'_d - \mu_- - \gamma'_d & \left(1 + \frac{\mu_- + \gamma'_d}{\hat{\varepsilon} \gamma'}\right) \gamma'_d \\ \hat{\lambda} + \gamma'_d - \mu_+ - \gamma'_d & \left(1 + \frac{\mu_+ + \gamma'_d}{\hat{\varepsilon} \gamma'}\right) \gamma'_d \end{pmatrix} \\
626 \quad & = \begin{pmatrix} \hat{\lambda} - \mu_- & \frac{\hat{\varepsilon} \gamma' + \mu_- + \gamma'_d}{\hat{\varepsilon} \gamma'} \gamma'_d \\ \hat{\lambda} - \mu_+ & \frac{\hat{\varepsilon} \gamma' + \mu_+ + \gamma'_d}{\hat{\varepsilon} \gamma'} \gamma'_d \end{pmatrix} \\
627 \quad & = \begin{pmatrix} \mu_+ & \frac{\hat{\varepsilon} \gamma' + \mu_- + \gamma'_d}{\hat{\varepsilon} \gamma'} \gamma'_d \\ \mu_- & \frac{\hat{\varepsilon} \gamma' + \mu_+ + \gamma'_d}{\hat{\varepsilon} \gamma'} \gamma'_d \end{pmatrix} \\
628 \quad &
\end{aligned}$$

629 and multiply the result by Λ

$$\begin{aligned}
630 \quad \Lambda^{-1} \mathbf{A} \Lambda &= \begin{pmatrix} \mu_+ & \frac{\hat{\varepsilon} \gamma' + \mu_- + \gamma'_d}{\hat{\varepsilon} \gamma'} \gamma'_d \\ \mu_- & \frac{\hat{\varepsilon} \gamma' + \mu_+ + \gamma'_d}{\hat{\varepsilon} \gamma'} \gamma'_d \end{pmatrix} \begin{pmatrix} \frac{\mu_+ + \gamma'_d}{\mu_+ - \mu_-} & -\frac{\mu_- + \gamma'_d}{\mu_+ - \mu_-} \\ \frac{\hat{\varepsilon} \gamma'}{\mu_+ - \mu_-} & -\frac{\hat{\varepsilon} \gamma'}{\mu_+ - \mu_-} \end{pmatrix} \\
631 &= \frac{1}{\kappa} \begin{pmatrix} \mu_+^2 + \mu_+ \gamma'_d + \hat{\varepsilon} \gamma' \gamma'_d + \mu_- \gamma'_d + \gamma'_d{}^2 & -\mu_+ \mu_- - \mu_+ \gamma'_d - \hat{\varepsilon} \gamma' \gamma'_d - \mu_- \gamma'_d - \gamma'_d{}^2 \\ \mu_- \mu_+ + \mu_- \gamma'_d + \hat{\varepsilon} \gamma' \gamma'_d + \mu_+ \gamma'_d + \gamma'_d{}^2 & -\mu_-^2 - \mu_- \gamma'_d - \hat{\varepsilon} \gamma' \gamma'_d - \mu_+ \gamma'_d - \gamma'_d{}^2 \end{pmatrix} \\
632 &= \frac{1}{\kappa} \begin{pmatrix} \mu_+^2 + (\hat{\lambda} + \hat{\varepsilon} \gamma' + \gamma'_d) \gamma'_d & -\mu_+ \mu_- - (\hat{\lambda} + \hat{\varepsilon} \gamma' + \gamma'_d) \gamma'_d \\ \mu_- \mu_+ + (\hat{\lambda} + \hat{\varepsilon} \gamma' + \gamma'_d) \gamma'_d & -\mu_-^2 - (\hat{\lambda} + \hat{\varepsilon} \gamma' + \gamma'_d) \gamma'_d \end{pmatrix} \\
633 &= \frac{1}{\kappa} \begin{pmatrix} \mu_+^2 - \mu_+ \mu_- & \lambda' \gamma'_d - \lambda' \gamma'_d \\ -\lambda' \gamma'_d + \lambda' \gamma'_d & -\mu_-^2 + \mu_+ \mu_- \end{pmatrix} = \frac{1}{\kappa} \begin{pmatrix} \mu_+ \kappa & 0 \\ 0 & \mu_- \kappa \end{pmatrix} = \begin{pmatrix} \mu_+ & 0 \\ 0 & \mu_- \end{pmatrix} \\
634 &
\end{aligned}$$

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

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

$$\begin{aligned}
637 \quad \dot{\mathbf{T}} &= \mathbf{F}' + \mathbf{T} \mathbf{B} \\
638 &
\end{aligned} \tag{A11}$$

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

641 The equations are

$$\begin{aligned}
642 \quad \dot{T}_{\pm} &= F'_{\pm} + \mu_{\pm} T_{\pm} \\
643 &
\end{aligned}$$

644 and the solutions of a generic initial value problem are

$$\begin{aligned}
645 \quad T_{\pm} &= \left(T_{\pm,0} + \int_{t_0}^t F'_{\pm} e^{-\mu_{\pm}(\tau-t_0)} d\tau \right) e^{\mu_{\pm}(t-t_0)} \\
646 &
\end{aligned} \tag{A12}$$

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

$$\begin{aligned}
649 \quad T_{\pm,0} &= \pm \frac{1}{\mu_+ - \mu_-} [(\mu_{\pm} + \gamma'_d) T_{u,0} + \hat{\varepsilon} \gamma' T_{d,0}] \\
650 &
\end{aligned}$$

651 the forcing components are

$$652 \quad F'_{\pm} = \pm \frac{\mu_{\pm} + \gamma'_d}{\mu_+ - \mu_-} F'$$

653

654 and the solutions in the temperature representation are

$$655 \quad T_u = T_+ + T_-$$

$$656 \quad T_d = -\frac{\mu_- + \gamma'_d}{\hat{\epsilon}\gamma'} T_+ - \frac{\mu_+ + \gamma'_d}{\hat{\epsilon}\gamma'} T_-$$

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

$$658 \quad T_u = T_+ + T_- \tag{A13}$$

$$659 \quad T_d = -\frac{\hat{\lambda} + 2\gamma'_d}{2\hat{\epsilon}\gamma'} (T_+ + T_-) + \frac{\kappa}{2\hat{\epsilon}\gamma'} (T_+ - T_-)$$

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

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

$$665 \quad T_d = -\frac{\mu_- + \gamma'_d}{\hat{\epsilon}\gamma'} T_+ - \frac{\mu_+ + \gamma'_d}{\hat{\epsilon}\gamma'} T_-$$

$$666 \quad = -\frac{1}{\hat{\epsilon}\gamma'} \left[\left(\frac{\hat{\lambda} - \kappa}{2} + \gamma'_d \right) T_+ + \left(\frac{\hat{\lambda} + \kappa}{2} + \gamma'_d \right) T_- \right]$$

$$667 \quad = -\frac{1}{\hat{\epsilon}\gamma'} \left[\left(\frac{\hat{\lambda} + 2\gamma'_d}{2} - \frac{\kappa}{2} \right) T_+ + \left(\frac{\hat{\lambda} + 2\gamma'_d}{2} + \frac{\kappa}{2} \right) T_- \right]$$

$$668 \quad = -\frac{1}{2\hat{\epsilon}\gamma'} \left[(\hat{\lambda} + 2\gamma'_d)(T_+ + T_-) - \kappa(T_+ - T_-) \right]$$

669

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

671 **Planetary imbalance**

672 Now, I will find an expression for the planetary imbalance in terms of the equations (A13). The
 673 mathematical expression that I should expand is $N = N_u + N_d = C_u \dot{T}_u + C_d \dot{T}_d$

$$\begin{aligned}
 674 \quad C_u \dot{T}_u &= C_u \dot{T}_s \\
 675 \quad C_d \dot{T}_d &= -C_d \frac{\hat{\lambda} + 2\gamma'_d}{2\hat{\epsilon}\gamma'} \dot{T}_s + C_d \frac{\kappa}{2\hat{\epsilon}\gamma'} \dot{T}_a \therefore \\
 676 \quad N &= C_u \dot{T}_s - C_d \frac{\hat{\lambda} + 2\gamma'_d}{2\hat{\epsilon}\gamma'} \dot{T}_s + C_d \frac{\kappa}{2\hat{\epsilon}\gamma'} \dot{T}_a \\
 677 \quad &= \left(C_u - C_d \frac{\hat{\lambda} + 2\gamma'_d}{2\hat{\epsilon}\gamma'} \right) \dot{T}_s + C_d \frac{\kappa}{2\hat{\epsilon}\gamma'} \dot{T}_a \\
 678 \quad &= C_s \dot{T}_s + C_a \dot{T}_a \\
 679
 \end{aligned}$$

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

$$\begin{aligned}
 681 \quad \dot{T}_s &= \mu_+ T_+ + \mu_- T_- + (F'_+ + F'_-) = \mu_+ T_+ + (\mu_+ - \kappa) T_- + (F'_+ + F'_-) \\
 682 \quad &= \mu_+ T_s - \kappa T_- + (F'_+ + F'_-) = \mu_+ T_s - \frac{\kappa}{2} (T_s - T_a) + (F'_+ + F'_-) \\
 683 \quad &= \frac{\hat{\lambda}}{2} T_s + \frac{\kappa}{2} T_a + (F'_+ + F'_-) = \frac{\hat{\lambda}}{2} T_s + \frac{\kappa}{2} T_a + F' \\
 684 \quad \dot{T}_a &= \mu_+ T_+ - \mu_- T_- + (F'_+ - F'_-) = \mu_+ T_+ - (\mu_+ - \kappa) T_- + (F'_+ - F'_-) \\
 685 \quad &= \mu_+ T_a + \kappa T_- + (F'_+ - F'_-) = \mu_+ T_a + \frac{\kappa}{2} (T_s - T_a) + (F'_+ - F'_-) \\
 686 \quad &= \frac{\kappa}{2} T_s + \frac{\hat{\lambda}}{2} T_a + (F'_+ - F'_-) = \frac{\kappa}{2} T_s + \frac{\hat{\lambda}}{2} T_a + \frac{\hat{\lambda} + 2\gamma'_d}{\kappa} F' \therefore \\
 687 \quad N &= \frac{1}{2} (\hat{\lambda} C_s + \kappa C_a) T_s + \frac{1}{2} (\hat{\lambda} C_a + \kappa C_s) T_a + \left(C_s + C_a \frac{\hat{\lambda} + 2\gamma'_d}{\kappa} \right) F' \\
 688
 \end{aligned}$$

Further expanding the coefficients

$$\begin{aligned}\hat{\lambda}C_s + \kappa C_a &= \hat{\lambda}C_u - \frac{C_d}{2\hat{\epsilon}\gamma'}(\hat{\lambda}^2 + 2\gamma'_d\hat{\lambda} - \kappa^2) = \hat{\lambda}C_u - \frac{C_d}{2\hat{\epsilon}\gamma'}(\hat{\lambda}^2 + 2\gamma'_d\hat{\lambda} - \hat{\lambda}^2 - 4\gamma'_d\lambda') \\ &= 2\frac{C_u}{\hat{\epsilon}}\left(\lambda' + \frac{\hat{\epsilon} - 1}{2}\hat{\lambda}\right)\end{aligned}$$

$$\hat{\lambda}C_a + \kappa C_s = \kappa C_u - \frac{C_d}{2\hat{\epsilon}\gamma'}(\kappa\hat{\lambda} + 2\gamma'_d\kappa - \kappa\hat{\lambda}) = \kappa C_u - \frac{C_u}{\hat{\epsilon}}\kappa = \kappa\frac{C_u}{\hat{\epsilon}}(\hat{\epsilon} - 1)$$

$$C_s + C_a\frac{\hat{\lambda} + 2\gamma'_d}{\kappa} = C_u - \frac{C_d}{2\hat{\epsilon}\gamma'}(\hat{\lambda} + 2\gamma'_d - \hat{\lambda} - 2\gamma'_d) = C_u$$

then the imbalance is

$$N = \frac{C_u}{\hat{\epsilon}}\left[\hat{\epsilon}F' + \left(\lambda' + \frac{\hat{\epsilon} - 1}{2}\hat{\lambda}\right)T_s + \kappa\frac{\hat{\epsilon} - 1}{2}T_a\right] \quad (\text{A14})$$

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 dN/dT_u . By virtue of the chain rule, $dN/dT_u = \dot{N}(dt/dT_u)$. In a neighborhood where $T_u(t)$ is injective, $dt/dT_u = 1/\dot{T}_u$. Therefore, the slope dN/dT_u is the ratio of two total derivatives: \dot{N} and \dot{T}_u .

We know that $T_u = T_s$, then $\dot{T}_u = \dot{T}_s$. 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

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

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

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

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

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

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

$$\frac{\dot{N}}{\dot{T}_u} = \left\{ -\frac{C_u}{|\lambda|} \frac{\dot{F}'}{\dot{T}_s} + \frac{\hat{\varepsilon} + 1}{2 \hat{\varepsilon}} \left(1 + \frac{\hat{\varepsilon} - 1}{\hat{\varepsilon} + 1} \frac{C_u \kappa}{|\lambda|} \left[\left(\hat{\varepsilon} + \frac{C_u}{C_d} \right) \frac{\gamma}{C_u \kappa} - \frac{\dot{T}_a}{\dot{T}_s} \right] \right) \right\} \lambda \quad (\text{A15})$$

736 The term in square brackets in equation (A15) is the key term that provides a NT -diagram with
 737 evolving slope when the forcing is constant. The second part of this term provides the temporal

738 evolution, whereas the first part is a constant term that sets the base enhancement of the slope.
 739 Interestingly, this first part contains in particular the thermal capacities of the system.

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

$$741 \quad \frac{\dot{N}}{\dot{T}_u} = \left\{ -\frac{C_u \dot{F}'}{|\lambda| \dot{T}_s} + \frac{\hat{\varepsilon} + 1}{2\hat{\varepsilon}} + \frac{\hat{\varepsilon} - 1}{2\hat{\varepsilon}} \frac{C_u \kappa}{|\lambda|} \left[\left(\frac{\hat{\varepsilon}}{C_u} + \frac{1}{C_d} \right) \frac{\gamma}{\kappa} - \frac{\dot{T}_a}{\dot{T}_s} \right] \right\} \lambda \quad (A16)$$

742

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

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

$$750 \quad \lim_{\hat{\varepsilon} \rightarrow 1} \frac{\dot{N}}{\dot{T}_u} = \lambda, \quad F = \text{const}$$

751

752 **Symmetric and antisymmetric modes**

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

778 $\cosh x \pm \sinh x$. The factors within square brackets in the last two terms can be thought as
 779 $e^x I_+ \pm e^{-x} I_-$, where I_{\pm} are the corresponding integrals. Using the expression of the exponential
 780 function in terms of the hyperbolic functions, I expand $e^x I_+ \pm e^{-x} I_- = (\cosh x + \sinh x) I_+ \pm (\cosh x -$
 781 $\sinh x) I_- = (I_+ \pm I_-) \cosh x + (I_+ \mp I_-) \sinh x$. Then, I overcome the limitation and now the two
 782 terms are written with hyperbolic functions. The coefficients of the hyperbolic functions are simple
 783 combinations of the integrals which can be also expanded easily. I do that now

$$\begin{aligned}
 784 \quad I_+ + I_- &= \int_{t_0}^t F' e^{-\mu_+(\tau-t_0)} d\tau + \int_{t_0}^t F' e^{-\mu_-(\tau-t_0)} d\tau = \int_{t_0}^t F' [e^{-\mu_+(\tau-t_0)} + e^{-\mu_-(\tau-t_0)}] d\tau \\
 785 \quad &= \int_{t_0}^t F' e^{-(\hat{\lambda}/2)(\tau-t_0)} [e^{-(\kappa/2)(\tau-t_0)} + e^{(\kappa/2)(\tau-t_0)}] d\tau \\
 786 \quad &= 2 \int_{t_0}^t F' e^{-(\hat{\lambda}/2)(\tau-t_0)} \cosh \left[\frac{\kappa}{2} (\tau - t_0) \right] d\tau \\
 787 \quad I_+ - I_- &= \int_{t_0}^t F' e^{-\mu_+(\tau-t_0)} d\tau - \int_{t_0}^t F' e^{-\mu_-(\tau-t_0)} d\tau = \int_{t_0}^t F' [e^{-\mu_+(\tau-t_0)} - e^{-\mu_-(\tau-t_0)}] d\tau \\
 788 \quad &= \int_{t_0}^t F' e^{-(\hat{\lambda}/2)(\tau-t_0)} [e^{-(\kappa/2)(\tau-t_0)} - e^{(\kappa/2)(\tau-t_0)}] d\tau \\
 789 \quad &= -2 \int_{t_0}^t F' e^{-(\hat{\lambda}/2)(\tau-t_0)} \sinh \left[\frac{\kappa}{2} (\tau - t_0) \right] d\tau \\
 790
 \end{aligned}$$

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

$$793 \quad T_s = \frac{e^{(\hat{\lambda}/2)(t-t_0)}}{\kappa} \left\{ C_1 \cosh \left[\frac{\kappa}{2} (t - t_0) \right] + C_2 \sinh \left[\frac{\kappa}{2} (t - t_0) \right] \right\} \quad (\text{A17})$$

$$794 \quad T_a = \frac{e^{(\hat{\lambda}/2)(t-t_0)}}{\kappa} \left\{ C_2 \cosh \left[\frac{\kappa}{2} (t - t_0) \right] + C_1 \sinh \left[\frac{\kappa}{2} (t - t_0) \right] \right\} \quad (\text{A18})$$

796 where

$$\begin{aligned}
 797 \quad C_1 &= \kappa T_{u,0} \\
 798 \quad &- (\hat{\lambda} + 2\gamma'_d) \int_{t_0}^t F' e^{-(\hat{\lambda}/2)(\tau-t_0)} \sinh \left[\frac{\kappa}{2} (\tau - t_0) \right] d\tau + \kappa \int_{t_0}^t F' e^{-(\hat{\lambda}/2)(\tau-t_0)} \cosh \left[\frac{\kappa}{2} (\tau - t_0) \right] d\tau \\
 799 \quad C_2 &= (\hat{\lambda} + 2\gamma'_d) T_{u,0} + 2\hat{\epsilon}\gamma'_d T_{d,0} \\
 800 \quad &+ (\hat{\lambda} + 2\gamma'_d) \int_{t_0}^t F' e^{-(\hat{\lambda}/2)(\tau-t_0)} \cosh \left[\frac{\kappa}{2} (\tau - t_0) \right] d\tau - \kappa \int_{t_0}^t F' e^{-(\hat{\lambda}/2)(\tau-t_0)} \sinh \left[\frac{\kappa}{2} (\tau - t_0) \right] d\tau \\
 801
 \end{aligned}$$

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

806 **Forced response to constant forcing**

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

$$808 \quad C_1 = F'_c \left\{ -(\hat{\lambda} + 2\gamma'_d) \int_{t_0}^t e^{-(\hat{\lambda}/2)(\tau-t_0)} \sinh \left[\frac{\kappa}{2}(\tau - t_0) \right] d\tau + \kappa \int_{t_0}^t e^{-(\hat{\lambda}/2)(\tau-t_0)} \cosh \left[\frac{\kappa}{2}(\tau - t_0) \right] d\tau \right\}$$

$$809 \quad C_2 = F'_c \left\{ (\hat{\lambda} + 2\gamma'_d) \int_{t_0}^t e^{-(\hat{\lambda}/2)(\tau-t_0)} \cosh \left[\frac{\kappa}{2}(\tau - t_0) \right] d\tau - \kappa \int_{t_0}^t e^{-(\hat{\lambda}/2)(\tau-t_0)} \sinh \left[\frac{\kappa}{2}(\tau - t_0) \right] d\tau \right\} \blacksquare$$

811 where the integrals are easily computed

$$812 \quad \int_{t_0}^t e^{-(\hat{\lambda}/2)(\tau-t_0)} \sinh \left[\frac{\kappa}{2}(\tau - t_0) \right] d\tau = \frac{e^{-(\hat{\lambda}/2)(t-t_0)}}{\lambda' \gamma'_d} \left\{ \frac{\kappa}{2} \cosh \left[\frac{\kappa}{2}(t - t_0) \right] + \frac{\hat{\lambda}}{2} \sinh \left[\frac{\kappa}{2}(t - t_0) \right] \right\} - \frac{\kappa}{2\lambda' \gamma'_d}$$

$$813 \quad \int_{t_0}^t e^{-(\hat{\lambda}/2)(\tau-t_0)} \cosh \left[\frac{\kappa}{2}(\tau - t_0) \right] d\tau = \frac{e^{-(\hat{\lambda}/2)(t-t_0)}}{\lambda' \gamma'_d} \left\{ \frac{\hat{\lambda}}{2} \cosh \left[\frac{\kappa}{2}(t - t_0) \right] + \frac{\kappa}{2} \sinh \left[\frac{\kappa}{2}(t - t_0) \right] \right\} - \frac{\hat{\lambda}}{2\lambda' \gamma'_d} \blacksquare$$

815 and, upon reduction, the C_i are

$$816 \quad C_1 = \frac{F'_c}{\lambda'} e^{-(\hat{\lambda}/2)(t-t_0)} \left\{ -\kappa \cosh \left[\frac{\kappa}{2}(t - t_0) \right] + (2\lambda' - \hat{\lambda}) \sinh \left[\frac{\kappa}{2}(t - t_0) \right] + \kappa e^{(\hat{\lambda}/2)(t-t_0)} \right\}$$

$$817 \quad C_2 = \frac{F'_c}{\lambda'} e^{-(\hat{\lambda}/2)(t-t_0)} \left\{ -(2\lambda' - \hat{\lambda}) \cosh \left[\frac{\kappa}{2}(t - t_0) \right] + \kappa \sinh \left[\frac{\kappa}{2}(t - t_0) \right] + (2\lambda' - \hat{\lambda}) e^{(\hat{\lambda}/2)(t-t_0)} \right\}$$

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

$$821 \quad T_s = \frac{F_c}{\lambda} \left\{ e^{(\hat{\lambda}/2)(t-t_0)} \left(\cosh \left[\frac{\kappa}{2}(t - t_0) \right] + \frac{2\lambda' - \hat{\lambda}}{\kappa} \sinh \left[\frac{\kappa}{2}(t - t_0) \right] \right) - 1 \right\} \quad (\text{A19})$$

$$822 \quad T_a = \frac{F_c}{\lambda} \left\{ e^{(\hat{\lambda}/2)(t-t_0)} \left(\frac{2\lambda' - \hat{\lambda}}{\kappa} \cosh \left[\frac{\kappa}{2}(t - t_0) \right] + \sinh \left[\frac{\kappa}{2}(t - t_0) \right] \right) - \frac{2\lambda' - \hat{\lambda}}{\kappa} \right\} \quad (\text{A20})$$

823

824 where $F'_c := F_c/C_u$. I can also obtain the explicit time derivatives of both modes. We take the time
 825 derivative both equations (A19) and (A20)

$$\begin{aligned}
 826 \quad \dot{T}_s &= \frac{F_c}{\lambda} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \frac{\hat{\lambda}}{2} \left(\cosh \left[\frac{\kappa}{2}(t-t_0) \right] + \frac{2\lambda' - \hat{\lambda}}{\kappa} \sinh \left[\frac{\kappa}{2}(t-t_0) \right] \right) \right. \\
 827 &\quad \left. + \frac{\kappa}{2} \left(\frac{2\lambda' - \hat{\lambda}}{\kappa} \cosh \left[\frac{\kappa}{2}(t-t_0) \right] + \sinh \left[\frac{\kappa}{2}(t-t_0) \right] \right) \right\} \\
 828 &= \frac{F_c}{\lambda} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \lambda' \cosh \left[\frac{\kappa}{2}(t-t_0) \right] + \frac{\lambda' \hat{\lambda} + 2\gamma'_d \lambda'}{\kappa} \sinh \left[\frac{\kappa}{2}(t-t_0) \right] \right\} \\
 829 &= \frac{F_c}{C_u} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \cosh \left[\frac{\kappa}{2}(t-t_0) \right] + \frac{\hat{\lambda} + 2\gamma'_d}{\kappa} \sinh \left[\frac{\kappa}{2}(t-t_0) \right] \right\} \\
 830 \quad \dot{T}_a &= \frac{F_c}{\lambda} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \frac{\hat{\lambda}}{2} \left(\frac{2\lambda' - \hat{\lambda}}{\kappa} \cosh \left[\frac{\kappa}{2}(t-t_0) \right] + \sinh \left[\frac{\kappa}{2}(t-t_0) \right] \right) \right. \\
 831 &\quad \left. + \frac{\kappa}{2} \left(\cosh \left[\frac{\kappa}{2}(t-t_0) \right] + \frac{2\lambda' - \hat{\lambda}}{\kappa} \sinh \left[\frac{\kappa}{2}(t-t_0) \right] \right) \right\} \\
 832 &= \frac{F_c}{\lambda} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \frac{\lambda' \hat{\lambda} + 2\gamma'_d \lambda'}{\kappa} \cosh \left[\frac{\kappa}{2}(t-t_0) \right] + \lambda' \sinh \left[\frac{\kappa}{2}(t-t_0) \right] \right\} \\
 833 &= \frac{F_c}{C_u} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \frac{\hat{\lambda} + 2\gamma'_d}{\kappa} \cosh \left[\frac{\kappa}{2}(t-t_0) \right] + \sinh \left[\frac{\kappa}{2}(t-t_0) \right] \right\} \\
 834 &
 \end{aligned}$$

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

$$\begin{aligned}
 836 \quad \dot{T}_s &= \frac{F_c}{C_u} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \cosh \left[\frac{\kappa}{2}(t-t_0) \right] + \frac{\hat{\lambda} + 2\gamma'_d}{\kappa} \sinh \left[\frac{\kappa}{2}(t-t_0) \right] \right\} \\
 837 \quad \dot{T}_a &= \frac{F_c}{C_u} e^{(\hat{\lambda}/2)(t-t_0)} \left\{ \frac{\hat{\lambda} + 2\gamma'_d}{\kappa} \cosh \left[\frac{\kappa}{2}(t-t_0) \right] + \sinh \left[\frac{\kappa}{2}(t-t_0) \right] \right\} \\
 838 &
 \end{aligned}$$

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

$$\begin{aligned}
 841 \quad \frac{\dot{T}_a}{\dot{T}_s} &= \frac{\frac{\hat{\lambda} + 2\gamma'_d}{\kappa} \cosh \left[\frac{\kappa}{2}(t-t_0) \right] + \sinh \left[\frac{\kappa}{2}(t-t_0) \right]}{\cosh \left[\frac{\kappa}{2}(t-t_0) \right] + \frac{\hat{\lambda} + 2\gamma'_d}{\kappa} \sinh \left[\frac{\kappa}{2}(t-t_0) \right]} \\
 842 &= \frac{\frac{\hat{\lambda} + 2\gamma'_d}{\kappa} + \tanh \left[\frac{\kappa}{2}(t-t_0) \right]}{1 + \frac{\hat{\lambda} + 2\gamma'_d}{\kappa} \tanh \left[\frac{\kappa}{2}(t-t_0) \right]} \\
 843 &
 \end{aligned}$$

844 Formally, above result have the alternative form

$$845 \quad \frac{\dot{T}_a}{\dot{T}_s} = \tanh \left[\frac{\kappa}{2}(t - t_0) + \operatorname{arctanh} \left(\frac{\hat{\lambda} + 2\gamma'_d}{\kappa} \right) \right]$$

846

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

$$848 \quad \left| \frac{\hat{\lambda} + 2\gamma'_d}{\kappa} \right| \leq 1$$

$$849 \quad \frac{\hat{\lambda}^2 + 4\gamma'_d\hat{\lambda} + 4\gamma'^2_d}{\hat{\lambda}^2 + 4\gamma'_d\lambda'} \leq 1$$

$$850 \quad \hat{\lambda}^2 + 4\gamma'_d\hat{\lambda} + 4\gamma'^2_d \leq \hat{\lambda}^2 + 4\gamma'_d\lambda'$$

$$851 \quad \hat{\lambda} + \gamma'_d \leq \lambda'$$

$$852 \quad -\hat{\varepsilon}\gamma' \leq 0$$

853

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

$$855 \quad \frac{\dot{T}_a}{\dot{T}_s} = \tanh \left[\frac{\kappa}{2}(t - t_0) + \operatorname{arctanh} \left(\frac{\hat{\lambda} + 2\gamma'_d}{\kappa} \right) \right] \quad (\text{A21})$$

856

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

$$863 \quad t_1 = t_0 + \frac{2}{\kappa} \operatorname{arctanh} \left| \frac{\hat{\lambda} + 2\gamma'_d}{\kappa} \right|$$

864

865 **Variation of the climate feedback parameter**

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

$$867 \frac{\dot{N}}{\dot{T}_u} = \frac{\hat{\varepsilon} + 1}{2\hat{\varepsilon}} \left(1 + \frac{\hat{\varepsilon} - 1}{\hat{\varepsilon} + 1} \frac{C_u \kappa}{|\lambda|} \left[\left(\hat{\varepsilon} + \frac{C_u}{C_d} \right) \frac{\gamma}{C_u \kappa} - \tanh \left(\frac{\kappa}{2} (t - t_0) + \operatorname{arctanh} \left(\frac{\hat{\lambda} + 2\gamma'_d}{\kappa} \right) \right) \right] \right) \lambda \quad (\text{A22})$$

868

869 The factor is composed of terms that are positive except for the ratio term coming from equation
 870 (A21). The negative ratio for $t \in [t_0, t_1)$ clearly generates a more negative slope, whereas for
 871 $t \in (t_1, \infty)$ makes it less negative. At the start one can get the slope

$$872 \frac{\dot{N}}{\dot{T}_u} = \left(1 + (\hat{\varepsilon} - 1) \frac{\gamma}{|\lambda|} \right) \lambda, \quad t = t_0$$

873

874 and at the time of sign reversal

$$875 \frac{\dot{N}}{\dot{T}_u} = \frac{\hat{\varepsilon} + 1}{2\hat{\varepsilon}} \left(1 + \frac{\hat{\varepsilon} - 1}{\hat{\varepsilon} + 1} \left(\hat{\varepsilon} + \frac{C_u}{C_d} \right) \frac{\gamma}{|\lambda|} \right) \lambda, \quad t = t_1$$

876

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

$$878 \lim_{t \rightarrow \infty} \frac{\dot{N}}{\dot{T}_u} = \frac{\hat{\varepsilon} + 1}{2\hat{\varepsilon}} \left(1 + \frac{\hat{\varepsilon} - 1}{\hat{\varepsilon} + 1} \frac{C_u \kappa}{|\lambda|} \left[\left(\hat{\varepsilon} + \frac{C_u}{C_d} \right) \frac{\gamma}{C_u \kappa} - 1 \right] \right) \lambda$$

879

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

884 **APPENDIX B**

885 **Feedbacks and pattern effect in a non-linear planetary budget**

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

$$887 N = (1 - \alpha)S + G - \epsilon \sigma (fT_u)^4 - \dot{C}T_u \quad (\text{B1})$$

888

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

894 If I compute the total derivative of N then

$$895 \quad \dot{N} = [(1 - \alpha)\dot{S} + \dot{G}] - S\dot{\alpha} - \sigma(fT_u)^4\dot{\epsilon} - 4\epsilon\sigma(fT_u)^3(\dot{f}T_u + f\dot{T}_u) - \dot{C}\dot{T}_u - T_u\ddot{C}$$

$$896 \quad = [(1 - \alpha)\dot{S} + \dot{G}] - \mathcal{R}$$

897

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

$$902 \quad \mathcal{R} = S\dot{\alpha} + \sigma(fT_u)^4\dot{\epsilon} + 4\epsilon\sigma(fT_u)^3(\dot{f}T_u + f\dot{T}_u) + \dot{C}\dot{T}_u + T_u\ddot{C}$$

$$903 \quad = 4\epsilon f\sigma(fT_u)^3\dot{T}_u \left[\frac{S}{4\epsilon f\sigma(fT_u)^3} \frac{\dot{\alpha}}{\dot{T}_u} + \frac{T_u}{4\epsilon} \frac{\dot{\epsilon}}{\dot{T}_u} + \frac{T_u}{f} \frac{\dot{f}}{\dot{T}_u} + 1 + \frac{\dot{C}}{4\epsilon f\sigma(fT_u)^3} + \frac{T_u}{4\epsilon f\sigma(fT_u)^3} \frac{\ddot{C}}{\dot{T}_u} \right]$$

904

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

$$907 \quad \frac{\dot{N}}{\dot{T}_u} = \left[(1 - \alpha) \frac{\dot{S}}{\dot{T}_u} + \frac{\dot{G}}{\dot{T}_u} \right]$$

$$908 \quad - \left[1 + \frac{S}{4\epsilon f\sigma(fT_u)^3} \frac{\dot{\alpha}}{\dot{T}_u} + \frac{T_u}{4\epsilon} \frac{\dot{\epsilon}}{\dot{T}_u} + \frac{T_u}{f} \frac{\dot{f}}{\dot{T}_u} + \frac{\dot{C}}{4\epsilon f\sigma(fT_u)^3} + \frac{T_u}{4\epsilon f\sigma(fT_u)^3} \frac{\ddot{C}}{\dot{T}_u} \right] 4\epsilon f\sigma(fT_u)^3$$

909

910 The first contribution in the \mathcal{R}/\dot{T}_u term is 1, representing the Planck feedback. The second
911 contribution is the planetary albedo feedback. It includes the surface albedo feedback as well as
912 the short-wave cloud feedback. The third contribution is the emissivity feedback, to which mainly
913 contributes the traditional water-vapor feedback. The fourth contribution is a representation of the
914 lapse-rate feedback. The fifth and sixth contributions are not atmospheric feedbacks but the effect

915 of the evolving planetary thermal capacity provided by the atmosphere-ocean interaction and the
916 ocean circulation.

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

921 **References**

- 922 Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor, 2012: Forcing, feedbacks and climate
923 sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophys. Res. Lett.*, **39 (9)**,
924 L09 712, <https://doi.org/10.1029/2012GL051607>.
- 925 Armour, K. C., 2017: Energy budget constraints on climate sensitivity in light of inconstant climate
926 feedbacks. *Nat. Clim. Change*, **7 (5)**, 331–335, <https://doi.org/10.1038/nclimate3278>.
- 927 Armour, K. C., C. M. Bitz, and G. H. Roe, 2013: Time-Varying Climate Sensitivity from Regional
928 Feedbacks. *J. Climate*, **26 (13)**, 4518–4534, <https://doi.org/10.1175/JCLI-D-12-00544.1>.
- 929 Arrhenius, S., 1896: On the Influence of Carbonic Acid in the Air upon the Temperature of the
930 Ground. *Philosophical Magazine and Journal of Science*, **41**, 237–276.
- 931 Budyko, M., 1969: The effect of solar radiation variations on the climate of the Earth. *Tellus*,
932 **21 (5)**, 611–619, <https://doi.org/10.1111/j.2153-3490.1969.tb00466.x>.
- 933 Callendar, G. S., 1938: The artificial production of carbon dioxide and its influence on temperature.
934 *Quarterly Journal of the Royal Meteorological Society*, **64 (275)**, 223–240, <https://doi.org/10.1002/qj.49706427503>.
- 935
- 936 Ceppi, P., and J. M. Gregory, 2017: Relationship of tropospheric stability to climate sensitivity
937 and Earth’s observed radiation budget. *Proc. Natl. Acad. Sci. (USA)*, **114 (50)**, 13 126–13 131,
938 <https://doi.org/10.1073/pnas.1714308114>.
- 939 Datseris, G., J. Blanco, O. Hadas, S. Bony, R. Caballero, Y. Kaspi, and B. Stevens, 2022: Minimal
940 recipes for global cloudiness. *Earth and Space Science Open Archive*, **14**, <https://doi.org/10.1002/essoar.10510797.2>.
- 941

- 942 Dong, Y., C. Proistosescu, K. C. Armour, and D. S. Battisti, 2019: Attributing Historical and
943 Future Evolution of Radiative Feedbacks to Regional Warming Patterns using a Green's Func-
944 tion Approach: The Preeminence of the Western Pacific. *J. Climate*, **32** (17), 5471–5491,
945 <https://doi.org/10.1175/JCLI-D-18-0843.1>.
- 946 Fourier, J.-B. J., 1827: Mémoire sur les Températures du Globe Terrestre et des Espaces Planétaires.
947 *Mémoires d l'Académie Royale des Sciences de l'Institute de France*, **7**, 570–604.
- 948 Geoffroy, O., D. Saint-Martin, G. Bellon, A. Voltaire, D. J. L. Olivié, and S. Tytéca, 2013a:
949 Transient Climate Response in a Two-Layer Energy-Balance Model. Part II: Representation of
950 the Efficacy of Deep-Ocean Heat Uptake and Validation for CMIP5 AOGCMs. *J. Climate*, **26** (6),
951 1859–1876, <https://doi.org/10.1175/JCLI-D-12-00196.1>.
- 952 Geoffroy, O., D. Saint-Martin, D. J. L. Olivié, A. Voltaire, G. Bellon, and S. Tytéca, 2013b:
953 Transient Climate Response in a Two-Layer Energy-Balance Model. Part I: Analytical Solution
954 and Parameter Calibration Using CMIP5 AOGCM Experiments. *J. Climate*, **26** (6), 1841–1857,
955 <https://doi.org/10.1175/JCLI-D-12-00195.1>.
- 956 Gregory, J. M., R. J. Stouffer, S. C. B. Raper, P. A. Stott, and N. A. Rayner, 2002: An Observationally
957 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](https://doi.org/10.1175/1520-0442(2002)015<3117:AOBEOT>2.0.CO;2).
- 959 Gregory, J. M., and Coauthors, 2004: A new method for diagnosing radiative forcing and climate
960 sensitivity. *Geophys. Res. Lett.*, **31** (3), L03 205, <https://doi.org/10.1029/2003GL018747>.
- 961 Hansen, J. E., R. A. Ruedy, M. Sato, and K.-W. K. Lo, 2010: Global surface temperature change.
962 *Reviews of Geophysics*, **48** (4), RG4004, <https://doi.org/10.1029/2010RG000345>, URL <https://doi.org/10.1029/2010RG000345>.
- 964 Hansen, J. E., G. Russell, A. Lacis, I. Fung, D. Rind, and P. Stone, 1985: Climate Response
965 Times: Dependence on Climate Sensitivity and Ocean Mixing. *Science*, **229** (4716), 857–859,
966 <https://doi.org/10.1002/2017MS001208>.
- 967 Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis, 2010: Probing the
968 Fast and Slow Components of Global Warming by Returning Abruptly to Preindustrial Forcing.
969 *J. Climate*, **23** (9), 2418–2427, <https://doi.org/10.1175/2009JCLI3466.1>.

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

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

974 Jiménez-de-la-Cuesta, D., 2022b: Global and tropical band averages for a selection of CMIP5 and
975 CMIP6 models: piControl and abrupt-4xCO2 experiments. *Dataset in Zenodo*, [https://doi.org/](https://doi.org/10.5281/zenodo.6531208)
976 [10.5281/zenodo.6531208](https://doi.org/10.5281/zenodo.6531208).

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

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

982 Mauritsen, T., 2016: Clouds cooled the Earth. *Nat. Geosci.*, **9** (12), 865–867, [https://doi.org/](https://doi.org/10.1038/ngeo2838)
983 [10.1038/ngeo2838](https://doi.org/10.1038/ngeo2838).

984 Meraner, K., T. Mauritsen, and A. Voigt, 2013: Robust increase in equilibrium climate
985 sensitivity under global warming. *Geophys. Res. Lett.*, **40** (2), 5944–5948, [https://doi.org/](https://doi.org/10.1002/2013GL058118)
986 [10.1002/2013GL058118](https://doi.org/10.1002/2013GL058118).

987 Newsom, E., L. Zanna, S. Khatiwala, and J. M. Gregory, 2020: The Influence of Warming Patterns
988 on Passive Ocean Heat Uptake. *Geophys. Res. Lett.*, **47** (18), e2020GL088 429, [https://doi.org/](https://doi.org/10.1029/2020GL088429)
989 [10.1029/2020GL088429](https://doi.org/10.1029/2020GL088429).

990 Otto, A., and Coauthors, 2013: Energy budget constraints on climate response. *Nat. Geosci.*, **6** (6),
991 415–416, <https://doi.org/10.1038/ngeo1836>.

992 Rohrschneider, T., B. Stevens, and T. Mauritsen, 2019: On simple representations of the climate
993 response to external radiative forcing. *Climate Dyn.*, **3** (5-6), 3131–3145, [https://doi.org/10.](https://doi.org/10.1007/s00382-019-04686-4)
994 [1007/s00382-019-04686-4](https://doi.org/10.1007/s00382-019-04686-4).

- 995 Rugenstein, M. A. A., and K. C. Armour, 2021: Three Flavors of Radiative Feedbacks and
996 Their Implications for Estimating Equilibrium Climate Sensitivity. *Geophys. Res. Lett.*, **48** (15),
997 e2021GL092983, <https://doi.org/10.1029/2021GL092983>.
- 998 Senior, C. A., and J. F. B. Mitchell, 2000: The time-dependence of climate sensitivity. *Geophys.*
999 *Res. Lett.*, **27** (17), 2685–2688, <https://doi.org/10.1029/2000GL011373>.
- 1000 Talley, L. D., 2015: Closure of the Global Overturning Circulation Through the Indian, Pacific,
1001 and Southern Oceans: Schematics and Transports. *Oceanography*, **26** (1), 80–97, [https://doi.org/](https://doi.org/10.5670/oceanog.2013.07)
1002 [10.5670/oceanog.2013.07](https://doi.org/10.5670/oceanog.2013.07).
- 1003 Wills, R. C. J., Y. Dong, C. Proistosescu, K. C. Armour, and D. S. Battisti, 2022: Systematic
1004 Climate Model Biases in the Large-Scale Patterns of Recent Sea-Surface Temperature and
1005 Sea-Level Pressure Change. *Geophys. Res. Lett.*, **49** (17), e2022GL100011, [https://doi.org/](https://doi.org/10.1029/2022GL100011)
1006 [10.1029/2022GL100011](https://doi.org/10.1029/2022GL100011).
- 1007 Winton, M., K. Takahashi, and I. M. Held, 2010: Importance of Ocean Heat Uptake Effi-
1008 cacy to Transient Climate Change. *J. Climate*, **23** (9), 2333–2344, [https://doi.org/10.1175/](https://doi.org/10.1175/2009JCLI3139.1)
1009 [2009JCLI3139.1](https://doi.org/10.1175/2009JCLI3139.1).
- 1010 Zhou, C., M. D. Zelinka, and S. A. Klein, 2016: Impact of decadal cloud variations on the Earth’s
1011 energy budget. *Nat. Geosci.*, **9** (12), 871–874, <https://doi.org/10.1038/ngeo2828>.