Climate energy balance models: two layers, orders, timescales, or regions?

B. B. Cael[1], Jonah Bloch-Johnson[2], and Nicholas Lutsko[3]

[1]National Oceanography Centre, Southampton, UK.

[2]National Centre for Atmospheric Science, Reading, UK.

[3]Scripps Institution of Oceanography, University of California, San Diego, USA

Corresponding author: B. B. Cael (cael@noc.ac.uk)

Key Points:

- Energy balance models adding a degree of complexity to a linear EBM (2-layer, order-2, 2-region, & 2-timescale) are compared
- The 2-region model fits historical global temperature and abrupt and gradual forcing simulations best, followed closely by the 2-layer model
- The 2-region model's best-fit parameter values are more plausible for the fit to historical temperatures than the 2-layer models'

Abstract

Climate energy balance models (EBMs) – simple energy-balance-based models of climate change – are widely used. The simplest "linear" EBM is deficient in capturing the behavior of complex climate models, so a "two-layer" model with an additional degree of complexity, i.e. two vertical layers, is typically used. Other additional degrees of complexity are equally plausible as well, however, and different approaches to add a degree complexity have not been compared quantitatively. Here we compare four types of EBMs - two-layer, order-two (temperature-dependent feedback), two-region (in space), and two-timescale (fast and slow climate responses) - specifically, their ability to capture historical temperature change and simulated temperature changes in abrupt (4x) and gradual (1%-ramp) forcing scenarios. The two-region model outperforms the others. The two-region model's best-fit parameters to historical temperatures are also more physically plausible than the next-best-fitting model, the two-layer model. We therefore conclude that the two-region model is the preferred climate EBM.

Plain Language Summary

State-of-the-art climate models are intensive to run, so for many applications their behavior is approximated by a simple climate "energy balance model" (EBM). By far the most common EBM is the "two-layer" model, which is one step up in complexity from the simplest possible EBM in that it represents the climate system as having two layers: a surface, and a deep ocean. There are other ways of stepping up complexity from the simplest possible model, however, which may work better at capturing the behavior of complex climate models or the historical evolution of global temperature. For example, one might instead represent two regions of the Earth's surface, or both a fast and a slow response to anthropogenic forcing, or that the baseline response changes as a function of temperature. Here we compare the ability of different "level-two" models, i.e. models each with a different single step up in complexity from the simplest possible model, to capture complex climate models and historical temperatures. We find that a two-region model works best, because it fits the data better than the others, and the next-best-fitting model, the two-layer model, has unrealistic parameter values for historical temperatures.

1 Introduction

Earth System Models (ESMs) are sophisticated representations of the climate system which are used to predict how the climate will evolve in response to anthropogenic radiative forcing (Eyring et al. 2015). ESMs are computationally intensive to run, so for many applications, from exploring implications of emissions trajectories (Dvorak et al. 2022) to estimating the economic consequences of climate change (Calel & Stainforth 2017), energy balance models (EBMs) are used instead. EBMs are simple representations of global climate which approximate the response of ESMs. A degree of complexity beyond a simple linear response to forcing is required to capture the behavior of ESMs; by far the most common way to include such complexity is with the "two-layer" model (Gregory 2000, Held et al. 2010, Geoffroy et al. 2013, Lutsko & Popp 2019; Section 2) which includes the effect of a deep ocean exchanging heat with the Earth's surface layer. This model has been used in countless applications and is a core component of contemporary understanding of Earth's climate.

However, other options to add such a degree of complexity exist, with associated physical rationales. Just as the two-layer model considers vertical variation in Earth's temperature, a "two-region" model can incorporate horizontal variation. Such horizontal variation could be particularly important to capture given the importance of horizontal variation in the climate feedback in setting the global temperature evolution (Armour et al. 2013, Zhou et al. 2017, Dong et al. 2019). While two-region and two-layer models are mathematically similar (Rohrschneider et al. 2019 – though see Section 2), they differ in their physical meaning and resulting implications for projections of warming.

Alternatively, one can consider multiple response timescales (van Hateren, 2013, Proistosescu & Huybers, Andrews et al. 2015) with a "two-timescale" model, as ESMs and Earth's climate respond to forcing over a continuum from fast to slow responses. This response is also likely dependent on the baseline climate state to which a forcing perturbation is applied, with colder or warmer climates having a higher climate sensitivity (Bloch-Johnson et al. 2021), so an "order-two" model may capture the temperature-dependence of the climate feedback. These various ways of adding complexity to an EBM have not been compared in terms of their ability to capture ESM behavior, or historical temperatures. One might of course include multiple of these degrees of complexity, but if only one can be or is chosen, as in many applications, it is worth knowing which one most accurately and plausibly captures historical and modelled temperature responses to radiative forcing.

Here we compare these four ways to incorporate additional complexity into an EBM by evaluating four corresponding "level-two" models' abilities to capture ESMs' responses to both abrupt and gradual forcing, as well as historical temperatures. We find that the two-region model outperforms the other models, and that the best-fit parameters for the two-region model's reproduction of historical temperatures are more physically plausible than those for the next-best-fitting model, the two-layer model. We thus conclude that the two-region model is the preferred EBM at this degree of complexity.

2 Energy balance models

EBMs generally take the form of energy balance models, which utilize conservation of energy to describe the warming of the Earth's surface. Arguably the simple such model is the linear energy balance model (Eq. 1):

$$c\frac{dT}{dt} = F - \lambda T - H$$

where c [W y m^{-2} K^{-1}] is the heat capacity of the surface layer whose temperature anomaly is T [K], F [W m^{-2}] is radiative forcing, λ [W m^{-2} K^{-1}] is the (global and total) climate feedback, and H[W m^{-2}] is ocean heat uptake. Note that here we use a sign convention of a positive λ being associated with a stable climate. H is often specified as equal to κT, with κ [W m^{-2} K^{-1}] being the ocean heat uptake efficiency. This model is generally understood not to be sufficiently complex to capture either the historical evolution of T or the T response of climate models to common forcing experiments. These are arguably respectably because λ has changed in recent decades thanks to the evolving spatial pattern of global warming setting off different feedback responses in different regions (Armour et al. 2013, Andrews et al. 2016, Andrews et al. 2022) and because the climate feedback changes over different response timescales in climate models (Andrews et al. 2015). It is therefore common to add a degree of complexity to Eq. 1 to capture this more complex behavior; by far the most commonly used model to do so is the two-layer model (Eq. 2; Gregory, 2000, Held et al. 2010, Geoffroy et al. 2013):

$$c\frac{dT}{dt} = F - \lambda T - \epsilon\gamma(T - T_D), \qquad c_D\frac{dT_D}{dt} = \gamma(T - T_D)$$

where c_D [W y m^{-2} K^{-1}] and T_D [K] represent a deep-ocean layer, which exchanges heat with the surface layer proportionally to γ [W m^{-2} K^{-1}]. ϵ is a dimensionless 'efficacy' parameter which is commonly included to capture changing feedbacks over time, but we note that this parameter is redundant and only notational; one may multiply both sides of the right-side equation above by ϵ and get the same dynamics with two free parameters, $\epsilon\gamma$ and ϵc_D. The two-layer model therefore has four free parameters ($c, \epsilon c_D, \epsilon\gamma, \lambda$ which we specify in this notation including ϵ here because it is more common.

The two-layer model is used so widely because it approximates well the behavior of ESMs (Held et al. 2010), but other models adding a different degree of complexity may do so just as well. We evaluate three other "level-two" models here. In each case, for simplicity of comparison we use a simple enough version of each model to have the same number of free parameters (four) as the two-layer model (see Section 4). One is an "order-two" model, which captures potential temperature-dependence of the linear EBM parameters (Eq. 3):

$$c(1 + \mu T)\frac{dT}{dt} = F - (\lambda + \kappa)(1 + \nu T)T$$

where ν [K^{-1}] captures the extent to which the climate feedback and ocean heat uptake efficiency may change with baseline climate state, and μ [K^{-1}] captures the same for the effective heat capacity. Note that it is mathematically equivalent whether κ and/or λ are specified as temperature-dependent; this only changes the interpretation of the parameter ν. Other potential order-two models are possible, e.g. with dependence on forcing, but these are less plausible mechanistically and in terms of their fit to historical records and ESM simulations (Cael et al. 2022, Bloch-Johnson et al. 2021). (Different combinations of forcing- and/or temperature-dependence were also tried and were found to perform worse than the above, so are not discussed further; see supporting information (SI) for other tested models.) Another model we consider is a "two-timescale" model which captures the fast and slow responses of the climate (Eq. 4):

$$T_n(t) = \alpha_n exp\left(\frac{-t}{\tau_n}\right) * F(t), \;\; T(t) = T_1(t) + T_2(t)$$

where $n = 1{,}2$, α_n [K m^2 W^{-1}] is the inverse of a climate feedback, and represents the convolution operation (Proistosescu & Huybers, 2017). This model is a direct representation of the multi-timescale response of the climate system; other versions of a multi-timescale response are possible, e.g. with a slow feedback term within a dynamical system (Goodwin & Cael 2021, Cael et al. 2022) or a superposition of temperature anomalies with multiple heat capacities, but Eq. 4 is the most common of such models and do not have the feature of Eq. 4 that the slow mode remembers past forcing. Slow-feedback terms that integrate temperature backwards in time as in Cael et al. (2022) or forcing as in Goodwin and Cael (2021), and a superposition of two temperatures with different heat capacities, were also tested and found to perform slightly worse than the above, so are not discussed further here; see SI for other tested models.

The last model we investigate is a "two-region" model which splits the surface layer into two regions (Eq. 5), one with a radiative feedback and one without:

$$c_1\frac{dT_1}{dt} = F - \frac{1}{\frac{1}{2}+\delta}\lambda T_1, \;\; c_2\frac{dT_2}{dt} = F, \quad T(t) = \left(\frac{1}{2}+\delta\right)T_1 + \left(\frac{1}{2}-\delta\right)T_2$$

where δ [dimensionless] is the excess fractional area in the radiatively active region (region 1), and c_1 and c_2 [W y m^{-2} K^{-1}] are the heat capacities of each region. The $\frac{1}{\frac{1}{2}+\delta}$term in the left-hand equation is included for ease of interpretation of the parameters, because it makes λ correspond to the globally averaged climate feedback; this term has no impact on the dynamics.

This model captures the spatial variation of two quantities: the climate feedback and the heat capacity of the ocean's mixed layer. Estimates made using a Green's function approach suggest that the feedbacks associated with surface temperature perturbations in regions of tropical ascent are an order of magnitude stronger than feedbacks elsewhere (Bloch-Johnson et al., 2023) such that these smaller feedbacks can be neglected in reconstructions of the global radiative response (Dong et al. 2019). Conversely, the Southern Ocean and the North Atlantic have far deeper mixed layers than the rest of the Earth (Manabe et al. 1991, Winton et al. 2010), such that their surface layers have far larger heat capacities than in other parts of the ocean. These are regions where radiative feedbacks are particularly weak (Wills et al. 2019).

This model therefore divides the Earth into a region that contains areas of tropical atmospheric convection with a radiative feedback and a small surface layer heat capacity (region 1), and a region that contains the high latitude areas of deep ocean heat uptake, with a negligible heat uptake and a large surface layer heat capacity, with the precise division between these regions allowed to vary as a free parameter δ. Our model is in some ways similar to the MT2 model from Gregory et al. (2023), though in Gregory et al. (2023) the radiative forcing in region 2 does not induce any surface warming. Several other versions of this model were tested (e.g. with a fixed δ, including a heat transport term between the two regions, or including multiple feedback terms; see SI for other tested models.) and were found to perform slightly worse than the above, so are not discussed further here. Note that we neglect energy transport between these two regions, which can be an important effect (Feldl and Roe 2013) but adds additional complexity beyond the four-parameter models considered in this paper. Note also that Rohrschneider et al.

(2019) showed that for every four-parameter two-layer model, there is an equivalent five-parameter (with nonzero λ in region two) two-region model with the same dynamics. This however does not make the model above equivalent to the two-layer model; it is not true that for every set of parameters for the above equation there is a set of parameters for the two-layer model giving the same dynamics, or vice versa.

Altogether Eqs. 2-5 represent four ways to add a degree of complexity to an EBM to capture the more complex response of climate models and historical records to historical forcing than Eq. 1 is capable of doing. Our question here is whether two layers, orders, timescales, or regions are best to approximate the warming response to radiative forcing.

3 Materials and Methods

We compare the four models in terms of their ability to capture three cases: i) historical temperatures over the past 170 years, and the temperature response over 150 years of ESMs in response to ii) abrupt forcing (i.e. an instantaneous quadrupling of CO_2 from preindustrial levels) and iii) gradual forcing (i.e. a 1% per year increase of CO_2 from preindustrial levels). For (i), we use the radiative forcing time series ensemble (2237 members) from the Intergovernmental Panel on Climate Change's Working Group I contribution to the Sixth Assessment Report (Smith et al. 2021), which is available through 2019, and the HadCRUT5 global surface temperature record (Morice et al. 2021). The temperature anomaly is defined relative to the period 1850-1900 and the radiative forcing is defined as an anomaly relative to the period 1759-1849. The preferred HadCRUT5 infilled version is used, and its 200-member ensemble is randomly resampled to generate an ensemble of the same size as that for radiative forcing. (Multiple random resamplings did not affect the results.) For (ii) and (iii), we use twenty-six ESMs' simulations from the CMIP6 ensemble (Eyring et al. 2015), available from github.com/markringer/cmip6 (accessed 13.1.2023). Note that the 4x and 1%-ramp simulations are the most standard CMIP simulations. For these twenty-six models the forcing associated with an instantaneous quadrupling of CO_2, F_{4x}, is calculated from the first 20 years of the simulation using the standard Gregory method (Gregory et al. 2004), and logarithmic forcing is assumed such that radiative forcing increases from zero to F_{4x} linearly over 140 years in the 1% per year simulations.

In each case, models are fit to each temperature-forcing pair by finding the parameter values that globally minimize the root-mean-square-error (RMSE, [°C]) of each models' fit to each temperature time-series when initialized with zero warming and then forced with the corresponding forcing time-series. This simple statistical metric is applicable because each model has the same number of free parameters, and is preferred over more complicated approaches because it gives an intuitive understanding of the statistical procedure and the meaning of the goodness-of-fit. These optimal parameter values are found iteratively via a random walk in parameter space with a Gaussian distribution of jump sizes (i.e. mostly small and occasional large jumps); the results are insensitive to altering the initial parameter guesses. In total this gives, for each model, and for each of 2237 ensemble estimates of the historical forcing and temperature as well as for each of 26 CMIP6 models' abrupt and gradual forcing simulations, a set of parameters and an associated goodness-of-fit statistic. All parameters except ν, μ, and α_2 are forced to be positive (these three may be negative, corresponding respectively to an increased climate sensitivity and decreased effective heat capacity at higher temperatures and an increased climate sensitivity on longer timescales). Not doing so only compounds the findings in the following section because for the two-layer model fit to historical temperatures, the best-fitting $\boldsymbol{\lambda}$ value is negative for most ensemble members (corresponding to an unstable climate in the sign convention used here).

For Figure 3, the historical cumulative radiative response is obtained by integrating the historical forcing ensemble in time and subtracting the mean ocean heat uptake estimate from Zanna et al. (2019).

4 Results and Discussion

We find that the two-region model capture the historical and model time-series better than the other models. Figure 1 shows the cumulative distribution function (CDF) of the RMSE for each model in each case. For historical temperatures and abrupt and gradual forcing simulations, the two-timescale model by and large has the worst fits. The order-two model has substantially worse fits for model time-series, and slightly worse fits for the historical time-series, than the two-region model. The differences in the RMSEs of the two-region and two-layer models are smaller. The two-region model has a lower RMSE than the two-layer model for 21/26 of the abrupt forcing simulations, 20/26 of the gradual forcing simulations, and >99% of historical temperature ensemble members, but the mean difference is 0.002°C or less in all three cases. The poorer performance of the two-timescale model is unsurprising in the context of Proistosescu and Huybers' (2017) finding that three timescales were required to capture models' responses to abrupt forcing, but a three-timescale model has six free parameters and is thus far more complex than the two-layer model or the two-region model described in Eq. 5. While the two-timescale model is conceptually similar to the other models that incorporate multiple timescales via multiple effective heat capacities (two-layer and -region), it appears these other formulations yield better performance. The order-two model's poorer performance is arguably more surprising given its superior reproduction of historical temperatures in Cael et al. (2022) relative to other models, albeit none of those considered here, which further underscores the importance of resolving multiple response timescales (n.b. the two-layer and two-region models implicitly capture a slower timescale response via a second heat capacity).

Thus the two-layer and two-region models are better able to reproduce historical temperatures and model responses to forcings, and have similar skill. Another essential aspect of this reproduction, however, is whether these EBMs are reproducing historical and model temperatures for the right reasons, i.e. with plausible parameter values. Significant differences in the parameter values for each case are to be expected given the substantial differences between historical temperatures and historical ESM simulations; nonetheless, the best-fit parameter values in each instance may or may not be consistent with basic reasoning about how the climate system functions. Figure 2 shows the CDF of each parameter's values for the two-layer (top row) and two-region (bottom row) models in each case. All four heat capacities are in the right order of magnitude to represent roughly a global surface ocean boundary layer (approximately the mixed layer), a global deep ocean layer, a surface ocean boundary layer in regions with shallow mixed layers, and a surface ocean boundary layer in regions with deeper mixed layers (Gregory et al. 2023, Cael et al. 2022).

For the two-region model, the fraction of the Earth without a climate feedback is variable but is ~1/3 of the Earth in the historical case and ~1/6$^{\text{th}}$ in the ESM cases. This is consistent with region 2 having a similar area fraction to deep ocean heat uptake regions (Manabe et al. 1991, Winton et al. 2010), suggesting that differences in heat capacity may play a larger role in setting $\boldsymbol{\delta}$ than differences in climate feedback (maps of spatial feedbacks from Dong et al. 2019 and Bloch-Johnson et al. 2023 suggest that if region 1 only included areas with large stabilizing feedbacks, then $\boldsymbol{\delta}$ would be negative). The two-region model's global total climate feedback $\boldsymbol{\lambda}$ is on the larger side of expectations (~2.5 W m^{-2} K^{-1}) relative to expectations in the historical case, consistent with the low observed historical warming relative to historical radiative forcing attributed to the pattern effect, and on the lower side in the model cases, consistent with the fairly high climate sensitivities in CMIP6 (Zelinka et al. 2020).

In contrast, the heat transport parameter between the surface and deep layers is ~5 W m^{-2} K^{-1}, an order of magnitude larger than expectations (Cael 2022). More importantly the global total climate feedback in the two layer case is effectively zero, being <0.05 W m^{-2} K^{-1} for 98% of historical ensemble members. This implies that radiative forcing has been almost entirely balanced by deep ocean heat uptake, whereas in reality most of the energy incident upon the top of the atmosphere is returned to space via climate feedbacks. The two-layer's fit to historical temperatures is therefore the result of a physically implausible set of parameters. This is illustrated in Figure 3: the cumulative radiative response, $\Re$ (i.e. $\boldsymbol{\lambda T}$ integrated over the surface of the Earth and then integrated in time), is not significantly different between the observations and the two-region model forced with historical $\boldsymbol{F}$ and using its historical parameter ensemble, but is almost zero for the two-layer model forced with historical $\boldsymbol{F}$ and using its historical parameter ensemble.

The two-region model therefore appears to be the preferred "level-two" EBM, because it fits historical and model temperature responses to radiative forcing best and with a physically plausible combination of parameters. That this model is preferred is consistent with the key role of the spatial pattern of warming, i.e. the 'pattern effect,' for the evolution of global temperatures. The version analyzed here is the simplest possible version; future work should consider what level of complexity is optimal for such a model – e.g. whether increasing complexity by combining some of the models given here or incorporating features like horizontal heat transport significantly improves the model's skill. One unsatisfying aspect of the two-region model presented here is that it does not have a well-defined equilibrium climate sensitivity, because the second region's temperature anomaly increases proportionally to forcing. Though equilibrium behavior is not relevant for most applications of EBMs, this issue is easily resolved regardless by including a second feedback parameter or a heat transport term between the two regions. It would also be valuable to identify a set of parameters that optimally fits all three cases considered here, to use e.g. for integrated assessment modelling, and to analyze the climate projections of such a model and parameter set. Doing so could improve calculations of everything from the timing of warming commitments (Dvorak et al. 2022) to the social cost of carbon (Rennert et al. 2022).

Acknowledgments
The authors have no conflicts to declare. Cael acknowledges support from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 820989 (project COMFORT). Bloch-Johnson acknowledges support from Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 786427 (project COUPLET). The work reflects only the authors' view; the European Commission and their executive agency are not responsible for any use that may be made of the information the work contains.

Open Research

Code is available for review purposes at github.com/bbcael/trol and will be given a Zenodo DOI should this manuscript be accepted for publication. Simulation time series are available at github.com/markringer/cmip6. Historical time series are available at https://github.com/chrisroadmap/ar6 and https://www.metoffice.gov.uk/hadobs/hadcrut5/.

References

Andrews, T., Gregory, J. M., & Webb, M. J. (2015). The dependence of radiative forcing and feedback on evolving patterns of surface temperature change in climate models. *Journal of Climate*, *28*(4), 1630-1648.

Andrews, T., Bodas-Salcedo, A., Gregory, J. M., Dong, Y., Armour, K. C., Paynter, D., et al. (2022). On the effect of historical SST patterns on radiative feedback. *Journal of Geophysical Research: Atmospheres*, 127, e2022JD036675.

Armour, K. C., Bitz, C. M., & Roe, G. H. (2013). Time-varying climate sensitivity from regional feedbacks. *Journal of Climate*, *26*(13), 4518-4534.

Bloch-Johnson, J., Rugenstein, M., Stolpe, M. B., Rohrschneider, T., Zheng, Y., & Gregory, J. M. (2021). Climate sensitivity increases under higher CO2 levels due to feedback temperature dependence. *Geophysical Research Letters*, *48*(4), e2020GL089074.

Bloch-Johnson, J., Maria A.A. Rugenstein, Marc J. Alessi, et al. (2013). The Green's Function Model Intercomparison Project (GFMIP) Protocol. *Authorea.* DOI: 10.22541/essoar.167839939.92474288/v2

Cael, B. B., Britten, G. L., Calafat, F. M., Bloch-Johnson, J., Stainforth, D., & Goodwin, P. (2022). Climate nonlinearities: selection, uncertainty, projections, and damages. *Environmental Research Letters*, *17*(8), 084025.

Cael, B. B. (2022). Ocean heat uptake efficiency increase since 1970. *Geophysical Research Letters*, *49*(19), e2022GL100215.

Calel, R., & Stainforth, D. A. (2017). On the physics of three integrated assessment models. *Bulletin of the American Meteorological Society*, *98*(6), 1199-1216.

Dong, Y., Proistosescu, C., Armour, K. C., & Battisti, D. S. (2019). Attributing historical and future evolution of radiative feedbacks to regional warming patterns using a Green's function approach: The preeminence of the western Pacific. *Journal of Climate*, *32*(17), 5471-5491.

Dvorak, M. T., Armour, K. C., Frierson, D. M. W., Proistosescu, C., Baker, M. B., & Smith, C. J. (2022). Estimating the timing of geophysical commitment to 1.5 and 2.0° C of global warming. *Nature Climate Change*, *12*(6), 547-552.

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

Feldl, N., & Roe, G. H. (2013). The nonlinear and nonlocal nature of climate feedbacks. *Journal of Climate*, *26*(21), 8289-8304.

Goodwin, P., & Cael, B. B. (2021). Bayesian estimation of Earth's climate sensitivity and transient climate response from observational warming and heat content datasets. *Earth System Dynamics*, *12*(2), 709-723.

Geoffroy, O., Saint-Martin, D., Olivié, D. J., Voldoire, A., Bellon, G., & Tytéca, S. (2013). Transient climate response in a two-layer energy-balance model. Part I: Analytical solution and parameter calibration using CMIP5 AOGCM experiments. *Journal of climate*, *26*(6), 1841-1857.

Gregory, J. M. (2000). Vertical heat transports in the ocean and their effect on time-dependent climate change. *Climate Dynamics*, *16*, 501-515.

Gregory, J., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., ... & Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical research letters*, *31*(3).

Gregory, J. M., and Andrews, T. (2016), Variation in climate sensitivity and feedback parameters during the historical period, *Geophys. Res. Lett.*, 43, 3911– 3920, doi:10.1002/2016GL068406.

Gregory, Jonathan, Jonah Bloch-Johnson, Matthew Couldrey, Eleftheria Exarchou, Stephen Griffies, Till Kuhlbrodt, Oleg Saenko, Tatsuo Suzuki, Quran Wu, Laure Zanna (in review). A new conceptual model of global ocean heat uptake, *Climate Dyanmics.*

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., & Vallis, G. K. (2010). Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *Journal of Climate*, *23*(9), 2418-2427.

Lutsko, N. J., & Popp, M. (2019). Probing the sources of uncertainty in transient warming on different timescales. *Geophysical Research Letters*, *46*(20), 11367-11377.

Proistosescu, C., & Huybers, P. J. (2017). Slow climate mode reconciles historical and model-based estimates of climate sensitivity. *Science Advances*, *3*(7), e1602821.

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., ... & Simpson, I. R. (2021). An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set. *Journal of Geophysical Research: Atmospheres*, *126*(3), e2019JD032361.

Rennert, K., Prest, B. C., Pizer, W. A., Newell, R. G., Anthoff, D., Kingdon, C., ... & Errickson, F. (2022). The social cost of carbon: advances in long-term probabilistic projections of population, GDP, emissions, and discount rates. *Brookings Papers on Economic Activity*, *2021*(2), 223-305.

Rohrschneider, T., Stevens, B., & Mauritsen, T. (2019). On simple representations of the climate response to external radiative forcing. *Climate Dynamics*, *53*, 3131-3145.

Smith, Christopher et al. “The earth’s energy budget, climate feedbacks, and climate sensitivity supplementary material”. In: *Climate Change: The Physical Science Basis* (2021), IPCC.

van Hateren, J. H. (2013). A fractal climate response function can simulate global average temperature trends of the modern era and the past millennium. *Clim. Dyn. 40*, 2651–2670.

Wills, R. C., Armour, K. C., Battisti, D. S., Proistosescu, C., & Parsons, L. A. (2021). Slow modes of global temperature variability and their impact on climate sensitivity estimates. *Journal of climate*, *34*(21), 8717-8738.

Winton, M., Takahashi, K., & Held, I. M. (2010). Importance of ocean heat uptake efficacy to transient climate change. *Journal of Climate*, *23*(9), 2333-2344.

Zanna, L., Khatiwala, S., Gregory, J. M., Ison, J., & Heimbach, P. (2019). Global reconstruction of historical ocean heat storage and transport. *Proceedings of the National Academy of Sciences*, *116*(4), 1126-1131.

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., ... & Taylor, K. E. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, *47*(1), e2019GL085782.

Zhou, C., Zelinka, M. D., & Klein, S. A. (2017). Analyzing the dependence of global cloud feedback on the spatial pattern of sea surface temperature change with a G reen's function approach. *Journal of Advances in Modeling Earth Systems*, *9*(5), 2174-2189.

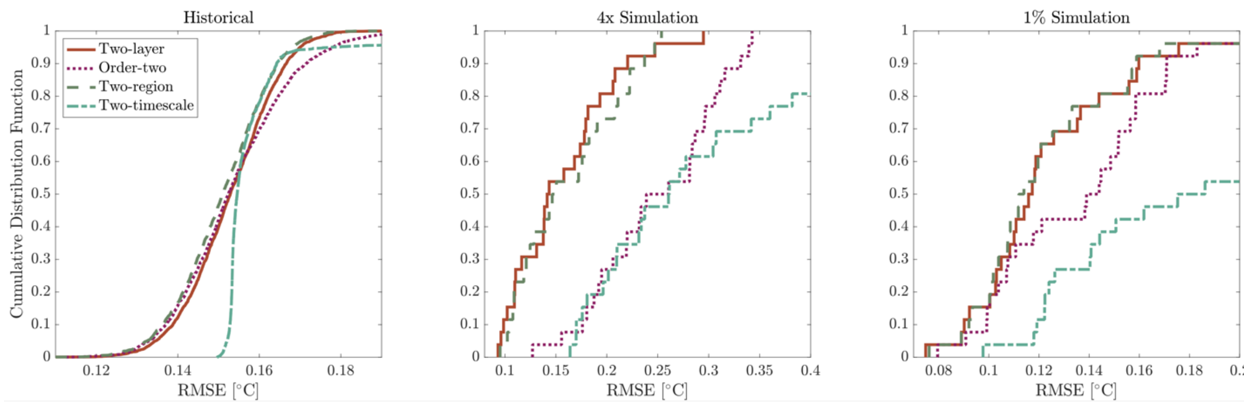

Figure 1. Cumulative distribution functions of the root-mean-square-error of each model's fit to each temperature time-series type.

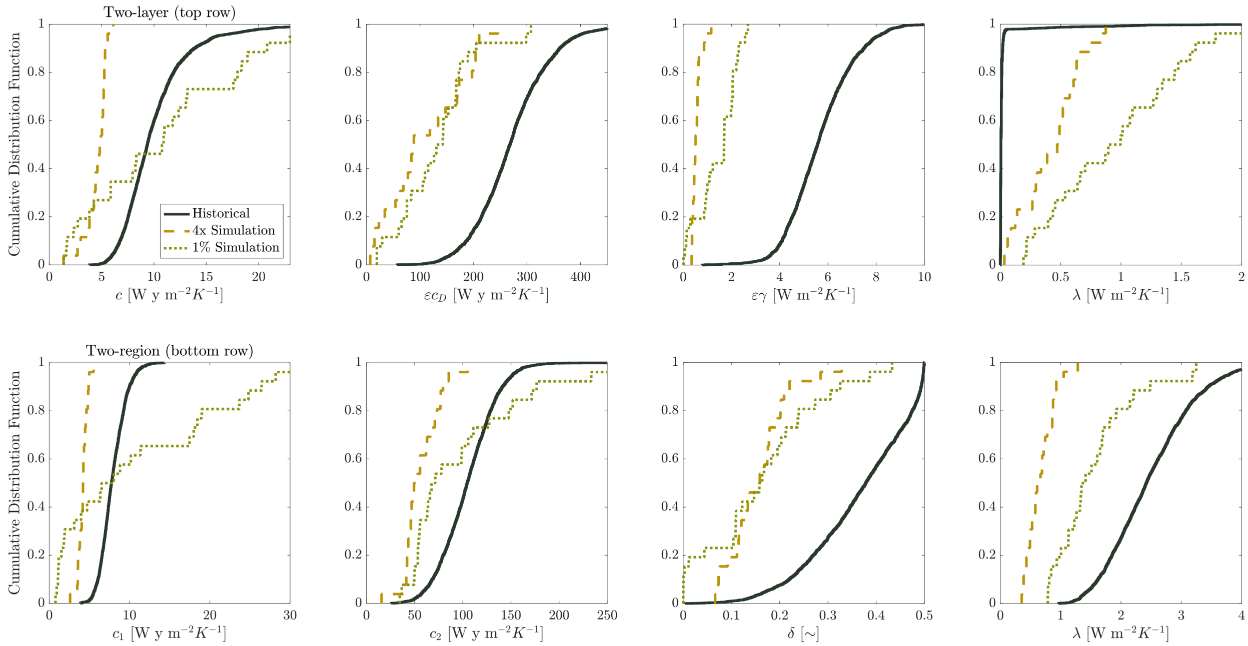

Figure 2. Cumulative distribution functions of the best-fit parameter values of each model's fit to each temperature time-series type.

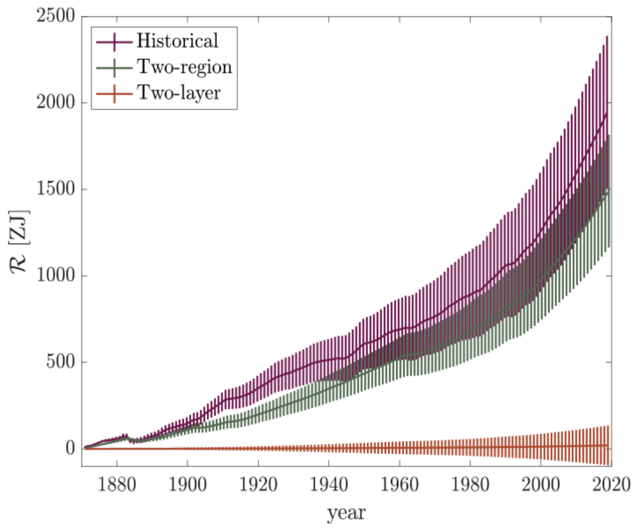

Figure 3. Cumulative radiative response $\Re$ (i.e. λT integrated over the surface of the Earth and then integrated in time) for the two-region and two-layer models with their historical parameter ensembles, compared to a historical estimate. Central lines correspond to ensemble means and error bars correspond to one ensemble standard deviation.