Response to Community Comments on "Constraining net long term climate feedback from satellite observed internal variability possible by mid 2030s"

We appreciate the community' constructive feedbacks. Below, we provide our detailed responses and the modifications made in response to his comments and suggestions.

- **Community comment.**

This is an interesting and useful study. However, I see two significant technical shortcomings in the authors' derivation of their emergent constraint based estimates of net long term climate feedback and of equilibrium climate sensitivity (ECS). The first of these two shortcomings biases the climate feedback estimate, weakening its central value by 32%, and together the two shortcomings bias the ECS estimate upwards by some 70%

1. The authors derive an emergent constraint on net long term climate feedback (λ_{ab}) with a median value of -1.56 $Wm^{-2}K^{-1}$ from a linear regression fit between net internal variability feedback (λ_{it}) and λ_{ab}, as shown in Figure 1(f). I cannot see that the regression method used for this purpose is explicitly stated, but it appears to be standard ordinary least squares (OLS) regression of λ_{ab} on λ_{it}. Such OLS regression, using data points and CERES-ERBE-ERA5 (observational) λ_{it} of –1.28 $Wm^{-2}K^{-1}$ digitized from Figure 1(f), yields a –1.56 $Wm^{-2}K^{-1}$ central estimate for λ_{ab}, identical to that given in line 287.

   OLS regression assumes that the regressor variable is error free; if it is not then the regression slope will be biased towards zero ("regression dilution"). There is little uncertainty in the regressee variable, λ_{ab}, due to the high level of effective radiative forcing (ERF), and hence large changes in planetary net radiative balance (N) and surface temperature anomaly (ΔT), involved in abrupt4xCO2 simulations by atmosphere-ocean global climate models (GCMs). However, there is significant uncertainty in the regressor variable, λ_{it}, as shown by the horizontal error bars in Figure 1(f). Hence OLS regression of λ_{ab} on λ_{it} is unsuitable and will give a slope estimate biased towards zero.

However, as there is little uncertainty in λ_{ab}, OLS regression will give an almost unbiased estimation if the regressor and regressee variables are switched, with λ_{it} regressed on λ_{ab}.

Doing so gives a regression fit estimate of $\lambda_{it} = 0.187 + 0.637\ \lambda_{ab}$, which on rearranging implies $\lambda_{ab} = 1.570\ \lambda_{it} - 0.294$. The central estimate of λ_{ab}, based on the observed λ_{it} estimate of –1.28 $Wm^{-2}K^{-1}$, is then –2.30 $Wm^{-2}K^{-1}$.

In summary, the authors should use λ_{ab} rather than λ_{it} as the regressor, in order to avoid significant bias in the estimated linear fit between them, and should adopt the resulting observationally-constrained λ_{ab} estimate of –2.30 $Wm^{-2}K^{-1}$ in place of their –1.56 $Wm^{-2}K^{-1}$ estimate, which is seriously biased by regression dilution.

To objectively assess whether the OLS regression used to characterize the emergent relationship between internal variability and forced climate net feedbacks is appropriate or significantly affected by regression dilution, we evaluate the performance of the OLS regression model by analyzing its residuals using several diagnostic methods. These include examining the mean residuals, the probability density function (PDF) (Figure 1 in this document), the Q-Q plot (Figure 2 in this document), and the residuals versus fitted values plot (Figure 3 in this document).

The mean residual (1.95e-16) and the probability density function (PDF) suggest that the model's residuals are centered around zero. Additionally, the Q-Q plot shows only minor deviations from normality, and the residuals versus fitted values plot reveals no distinct residuals pattern, indicating constant variance and minimal heteroscedasticity.

Furthermore, we repeated the regressions using Orthogonal Distance Regression (ODR), a method that accounts for errors in both the independent and dependent variables. The comparison of the regression coefficients from both methods (Table 1 in this document) shows minimal differences.

These findings indicate that, despite minor biases and deviations from normality, the model's fit remains robust, with minimal impact from regression dilution on the slope approaching zero. This evaluation alleviates concerns about regression dilution in calculating the emergent relationship, suggesting that our OLS regression results are likely reliable. We have now included this comparison in our revised manuscript.

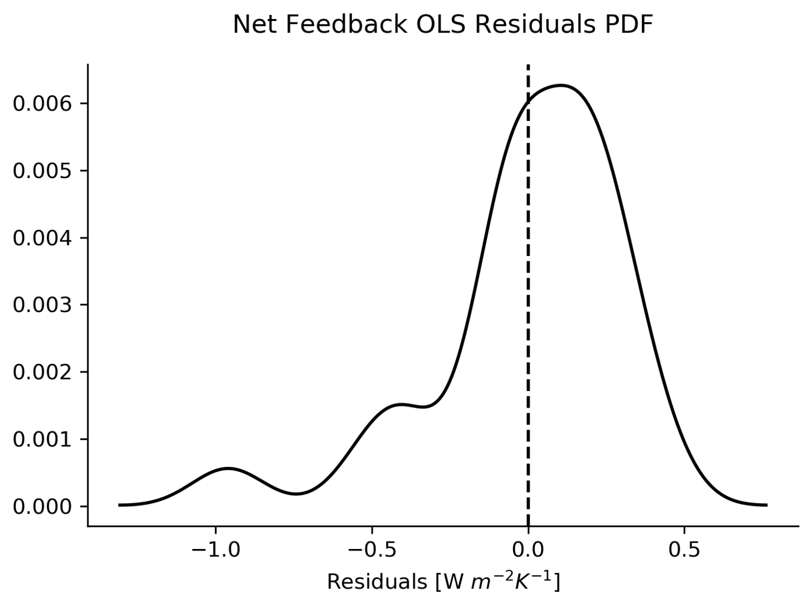

Figure 1: Probability density function (PDF) of residuals from the OLS regression between forced climate and internal variability net feedbacks. The distribution of residuals indicates a high likelihood close to zero, suggesting that the model's residuals are centered around zero, which supports the validity of the OLS regression model.

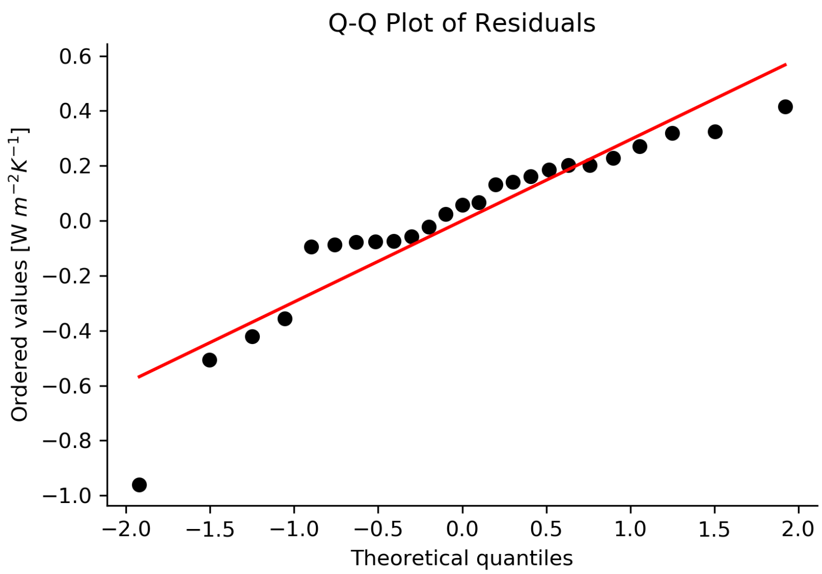

Figure 2: Q-Q Plot of Residuals from the OLS regression between forced climate and internal variability net feedback. The red line represents the theoretical quantiles, while the black markers show the sample quantiles. Minor deviations from the line indicate slight departures from normality.

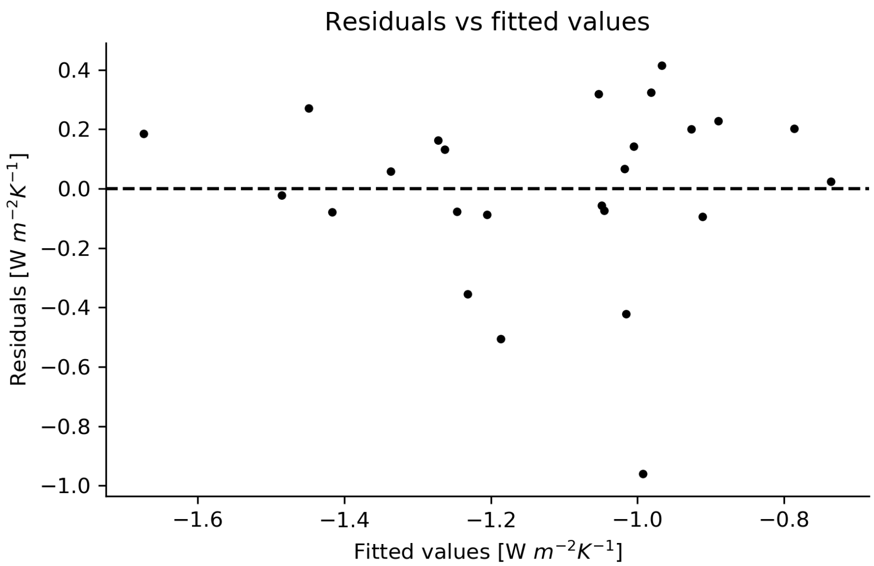

Figure 3: Residuals against predicted values from the OLS regression between forced climate and internal variability net feedback. The lack of a clear pattern suggests that the residuals are randomly distributed around zero, indicating that the model's assumptions of linearity and homoscedasticity are reasonably met.

Table 1: Forced climate and internal variability feedbacks regression coefficients depending on the choice of regression method.

	Longwave		Shortwave		Net	
	OLS	ODR	OLS	ODR	OLS	ODR
Slope	0.57	0.57	0.74	0.74	0.57	0.62
Intercept [$Wm^{-2}K^{-1}$]	-0.77	-0.76	-0.26	-0.26	-0.82	-0.80

2. The standard estimate of a GCM's ECS corresponds to the ΔT at which N = 0 when extending an OLS regression linear fit of annual mean N on ΔT over 150 years of its abrupt4xCO2 simulation. That ΔT mathematically equals minus the slope of the regression fit line, $-\lambda_{ab}$, divided into the value of N where the fit line intersects the ΔT = 0 axis (F4x_reg150). It follows that the standard estimate is ECS = F4x_reg150 / $-\lambda_{ab}$.

   For almost all GCMs, F4x_reg150 is significantly lower than the actual ERF from quadrupled CO2, as estimated from fixed SST simulations with a correction for land surface warming (F4x_SST-ts). The reason for this is simple. Net feedback is generally higher in the early part of 150 year abrupt4xCO2 simulations than in the much more numerous subsequent years, which have a dominant influence on linear estimation using OLS. As a result, in the earliest part of the N versus ΔT plot the regression fit lies below the actual N values, most significantly when ΔT = 0 at the start (taking F4x_SST-ts as the best estimate of the actual value of N at that point). As the standard estimate of ECS in GCMs (before scaling from 4x to 2x CO2) is F4x_reg150 / $-\lambda_{ab}$, ECS estimated as F4x_SST-ts / $-\lambda_{ab}$ will be biased upwards.

   It follows that deriving ECS, as the authors do for their median ECS estimate of 2.5 K, by dividing an observationally-constrained GCM-based estimate of $-\lambda_{ab}$ into estimated F2x_SST-ts will significantly over-estimate ECS. This point is illustrated and more fully explained in sections 4.1 and S1 of Lewis (2023). Although using F2x_SST-ts rather than F2x_reg150 as the numerator when using estimated λ_{ab} in the denominator is not uncommon when estimating ECS, doing so is unjustifiable: it is mathematically incorrect and causes significant overestimation of ECS.

   On average, F4x_reg150 is 16% lower than F4x_SST-ts in the 17 CMIP6 models for which Smith et al (2020) were able to derive F4x_SST-ts (see their Table S1 ERFreg150 and ERF_ts values). Those 17 models have an average F4x_SST-ts of 8.41 Wm^{-2}. This should be converted to a value for a doubling of CO2, F2x_SST-ts, by dividing F4x_SST-ts by 2.10, per the formula in Meinshausen et al. (2020) that was adopted in IPCC AR6, rather than using the popular but inaccurate method of simply halving the 4x CO2 ERF. Doing so gives a F2x_SST-ts value of 4.01 Wm^{-2} for the Smith et al (2020) CMIP6 mean. That is close to the 3.93 Wm^{-2} value of F2x_SST-ts derived in AR6 and used in the manuscript. By contrast, F2x_reg150, the similarly converted value of F4x_reg150, is only 3.37 Wm^{-2} for the 17 Smith et al (2020) models – almost identical to the 3.35 Wm^{-2} average that I calculate for a larger set of 30 CMIP6 models.

   The 2.10 ratio of quadrupled to doubled CO2 ERF is derived using detailed line-by-line radiation code, not simplified GCM radiation code, but the radiation code in CMIP6 models should be more accurate than that in earlier GCM generations. Moreover, I compute a similar (marginally higher) average 4x to 2x CO2 ERF ratio for the five GCMs analysed in Rugenstein et al (2020) with data from both abrupt4xCO2 and abrupt2xCO2 simulations, with the ratio being 2.10 for the only CMIP6 GCM included (CNRM-CM6). (That is based on estimating ERF by regression of N on ΔT over the first ten years after the abrupt CO2 increase, which provides a reasonable proxy for F4x_SST-ts in the Smith et al (2020) set of abrupt4xCO2 simulations.)

It follows that the authors should revise their ECS estimation formula to ECS = (F4x_reg150 / 2.10) /λ_{ab}, using the average regression-derived F4x_reg150 for the set of CMIP6 models used to constrain λ_{ab}. If that F2x_reg150 were the same as the 3.35 Wm^{-2} that I calculated for 30 CMIP6 models, then the revised median ECS estimate, based on the corrected central λ_{it} estimate of –2.30 Wm^{-2}K-1, would be 3.35 / 2.30 = 1.46 K. If the IPCC AR6 assessment of ECS is correct, then such a low ECS estimate may be considered unlikely to be accurate. If so, the correct, and important, conclusion to draw would then be that the relationship between λ_{ab} and λ_{it} in CMIP6 models does not provide a reliable emergent constraint on ECS.

We recognize the concern regarding the use of estimated radiative forcing to calculate ECS. However, our primary objective is to explore how internal variability feedbacks can aid to constrain forced climate feedbacks. The ECS estimate presented in the manuscript serves as an example of what could be achieved with further observations of the Earth's radiation budget. We have emphasized that our emergent constraint results should be interpreted cautiously, as they offer indicative insights rather than definitive observational evidence.

In response to the community's feedback, we have included his suggested alternative approach in our revised manuscript. In addition to using the IPCC AR6 radiative forcing estimate, we now compute radiative forcing from the model set by taking the y-intercept of the regression between TOA flux anomalies and surface temperature anomalies, then dividing by 2.1. Using this alternative radiative forcing estimate and the constrained forced feedback, we provide the corresponding ECS estimate.

The relationship of "true" ECS to that derived from regression over 150 years after a CO2 increase

Linear regression of N on ΔT over the first 150 years of an abrupt4xCO2 (or abrupt2xCO2) simulation, with ECS taken as the N = 0 intercept of the fit (ECSreg150) is the standard method for estimating the ECS of GCMs. Moreover, it is usual for observationally-constrained non-paleoclimate ECS studies to estimate that or another effective climate sensitivity measure. But, as noted in the Comment by Anonymous Referee #1, in GCMs the actual (true) ECS, as estimated from ultra-long abrupt CO2 increase forced simulations, generally exceeds ECSreg150. However, the 17% mean excess (for abrupt4xCO2 simulations) stated in the paper they cited, Rugenstein et al (2020), includes the FAMOUS model, which appears to be near to runaway at quadrupled CO2 – it warms almost four times as much as for doubled CO2. On a forcing-adjusted basis (dividing abrupt4xCO2 warming by 2.10), its 4x to 2x CO2 ECS ratio is 1.86, while for all the other models with both simulations that ratio lies in the range 1.00 to 1.10.

The average excess of estimated actual ECS over ECSreg150 in the Rugenstein et al (2020) models excluding the outlier FAMOUS abrupt4xCO2 simulation is 13.6%. The average ratio for the superset of those models included in Dunne et al (2020), ex FAMOUS, is almost identical.

Moreover, if true equilibrium ECS is to be estimated, it should logically be from ultra long abrupt2xCO2 simulations, as the definition of ECS is for a doubling, not a quadrupling, of preindustrial carbon dioxide concentration. The average ratio, for the Rugenstein et al (2020) models, of estimated true ECS for 2x CO2 to ECSreg150 derived by dividing abrupt4xCO2 data by 2.10, is slightly lower at 1.11x (or 1.06x when including FAMOUS).

We believe this community comment is referring to the main comment raised by Referee 1 and not specifically to our manuscript. Since we do not represent the referees' opinions directly in our work, we have chosen not to provide further commentary on this matter.