Reviewer comments are in black and our reply is in blue.

McBride et al. present a physics informed statistical model of temperature change induced by anthropogenic and natural forcing. They use a comprehensive set of observational datasets to inform the nine parameters of their model. They conclude that the observed attributable anthropogenic warming rate, and climate sensitivity, is not only significantly lower than the model mean of CMIP6, but also, for climate sensitivity, significantly lower than our recent assessment of multiple lines of evidence on climate sensitivity.

I believe the stark contrast between the current study and the recent assessment by Sherwood et al. (2020) is partially an artefact of data choices, most significantly the choice of the HadCRUT dataset for global temperature, which has poor coverage of quickly rising Arctic temperatures. The inclusion of older estimates of ocean heat uptake has a secondary effect. I think that the paper will be a useful contribution to debate around climate sensitivity after major revisions.

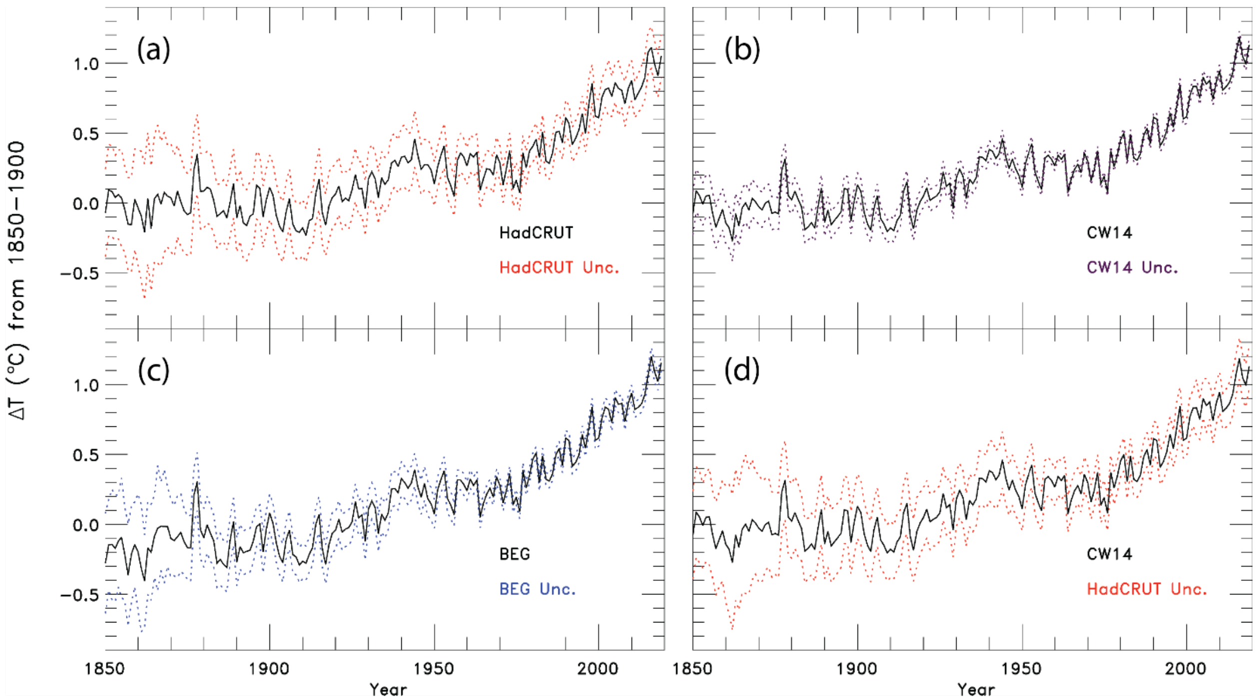
> We thank the reviewer for taking the time to carefully read our manuscript and suggest useful changes. Upon revision, we will make changes to address all of these comments, as detailed below each comment.

1) The HadCRUT dataset is known to underestimate recent warming. The model mean is computed without compensating for the fact that Artic temperatures are underrepresented, leading to an underestimate of warming (Cowtan and Way, 2014, Cowtan, 2017). Even with corrected sea surface temperature, this still leads to under-reported warming compared to a record with global coverage (Cowtan, 2017). The HadCRUT data should be replaced by the Cowtan record.

> We will address comment 1 and comment 2 together just below, since they are both related to the data records used in the EM-GC framework.

2) According to the IPCC's SROCC report, older estimates of ocean heat uptake have biases that may lead to an underestimate of ocean heat uptake (Bindoff, 2019, p.457). Carton et al (2018), whose record was derived with data assimilation, indicated that previous estimates with data assimilation (possibly like Balmaseda's record), may have contained errors that have prevented them from being sufficiently. Similarly, Cheng et al (2017) and Ishii et al (2017) can be considered superior to the old standard of Levitus (2012).

> We plan, upon revision, to show model results for the use of Cowtan and Way (2014) (hereafter CW14) data record for global mean near-surface temperature together with the Cheng et al. (2017) record for ocean heat content (OHC).
>
> When we ran the EM-GC using the CW14 temperature record, one issue that arose is the relatively small published uncertainties between 1850-1900 associated with this record. The EM-GC was not able to calculate any good fits as defined by the computation of a value of the reduced chi-squared parameter for GMST, χ^2_{ATM}, being less than 2 upon using the CW14 temperature time series and the published uncertainties. The New Fig. S4 below compares the uncertainties associated with the Hadley Centre Climate Research Unit (HadCRUT), Berkeley Earth Group (BEG), and CW14.

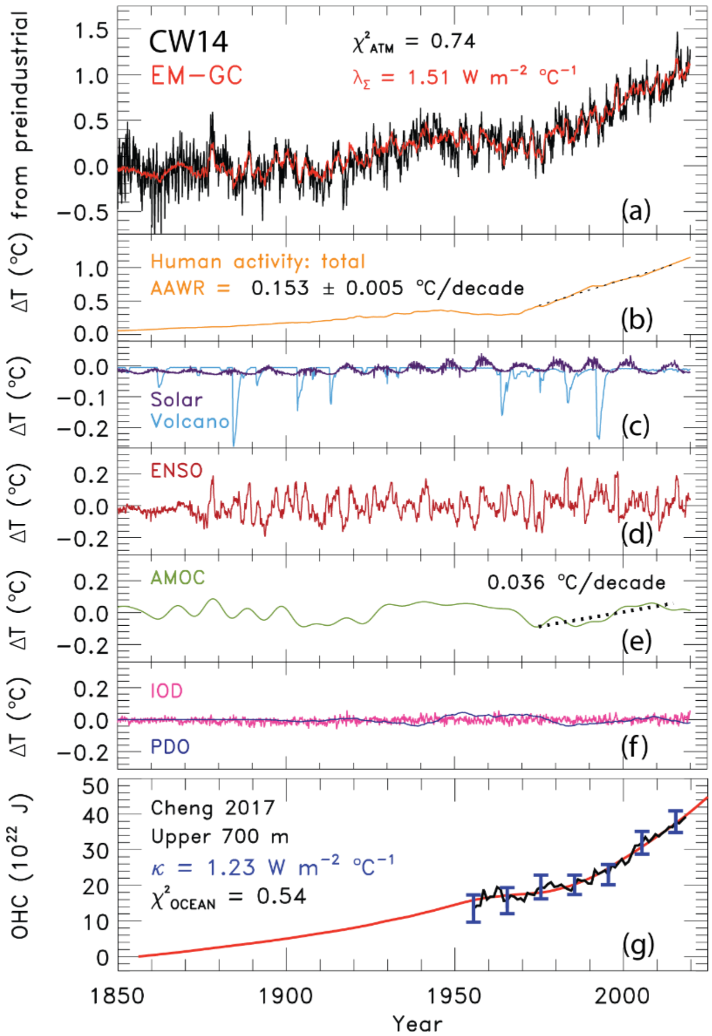


New Figure S4. GMST anomaly relative to pre-industrial over time. (a) HadCRUT with the HadCRUT uncertainties. (b) CW14 with the CW14 uncertainties. (c) BEG with the BEG uncertainties. (d) CW14 with the HadCRUT uncertainties.

The uncertainties for CW14 (New Fig. S4b) are much smaller than those for the HadCRUT (New Fig. S4a) and BEG (New Fig. S4c) temperature records. The small values of CW14 uncertainties, especially from 1850-1900, cause the EM-GC to not be able to achieve good fits to this temperature record. We have two choices for use of the CW14 record; either relax the constraint for χ^2_{ATM} (i.e., run with $\chi^2_{ATM} \leq 4$), or modify the CW14 uncertainties.

Upon revision, we propose to show model results for which we combine the uncertainties from HadCRUT with the data values for GMST from the CW14 record (New Fig. S4d), since CW14 is based upon the HadCRUT temperature record. Upon use of this combination of data and uncertainty, we are able to find good fits to the CW14 temperature record that look reasonable.

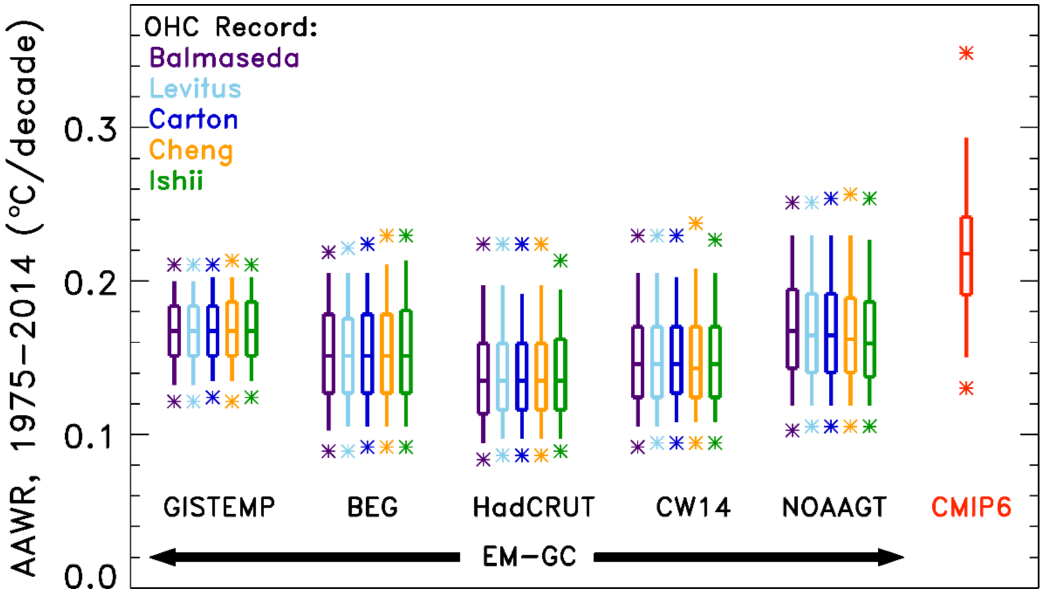
When the CW14 temperature record is used instead of HadCRUT, we see modest changes in the climate feedback parameter, λ_Σ, the attributable anthropogenic warming rate, AAWR, equilibrium climate sensitivity, ECS, and our projected global mean surface temperature (GMST) in 2100, ΔT_{2100}, as detailed below. The New Fig. S3 below shows a plot similar to Fig. 1 of the submitted paper, for which we use the CW14 temperature record along with the Cheng et al. (2017) record for OHC.



New Figure S3. Measured and modeled GMST anomaly (ΔT) relative to a pre-industrial (1850-1900) baseline. (a) Observed (black) and modeled (red) ΔT from 1850-2019. This panel also displays the values of λ_Σ and χ^2_{ATM} (see text) for this best-fit simulation. (b) Contributions from total human activity. This panel also denotes the numerical value of the attributable anthropogenic warming rate from 1975-2014 (black dashed) as well as the 2σ uncertainty in the slope. (c) Solar irradiance (light blue) and major volcanoes (purple). (d) Influences from ENSO on ΔT. (e) Contributions from AMOC to ΔT and to observed warming from 1975-2014. (f) Influences from PDO (blue) and IOD (pink) on ΔT. (g) Measured (black) and modeled (red) ocean heat content (OHC) as a function of time for the Cheng 2017 OHC record, the value of χ^2_{OCEAN} for this run, as well as the ocean heat uptake efficiency, κ, needed to provide the best-fit to the OHC record. The error bars (blue) denote the uncertainty in OHC used in this analysis (see Sect. 2.2.8).

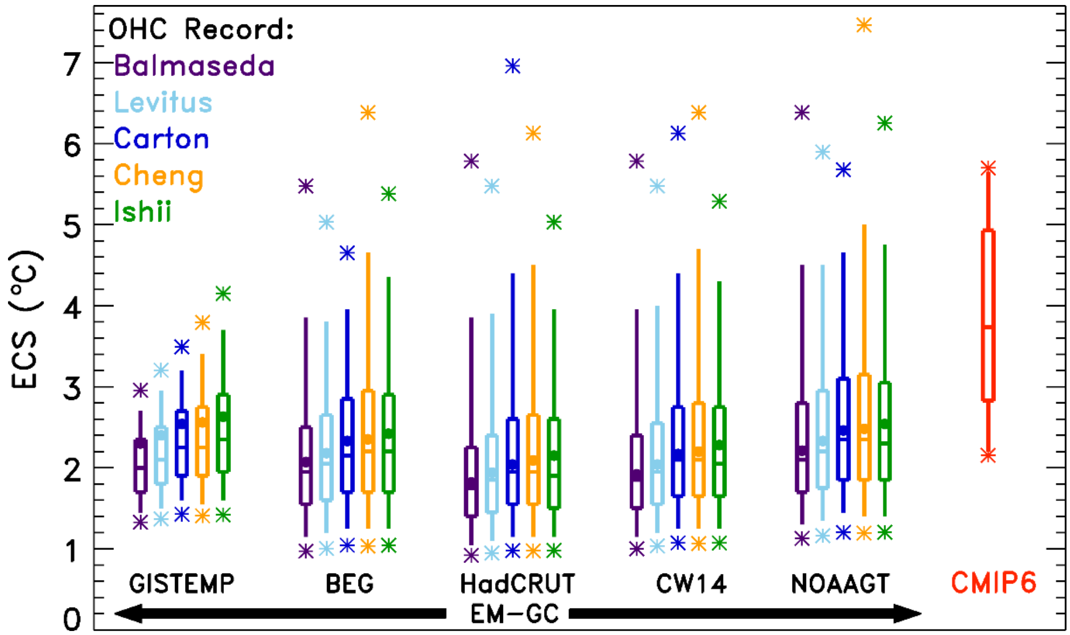
Our estimate of λ_Σ upon use of the best estimate of aerosol radiative forcing in 2011 (AER RF_{2011} of −0.9 W m^{-2}) increases from 1.38 W m^{-2} °C^{-1} (submitted paper) to 1.51 W m^{-2} °C^{-1} (New Fig S3) upon use of CW14 for GMST and the Cheng et al. 2017 record for OHC. The estimate of the value of AAWR increases from 0.144 °C/decade to 0.153°C/decade. This sensitivity of λ_Σ and AAWR to data choice for GMST and OHC will be highlighted in the revised paper. We propose to add New Figure S3 to the supplement to document the effect on λ_Σ of data choice, and to add model results for the CW14/Cheng et al. 2017 data combination to three additional figures in the Main paper, as detailed below.

Proposed update to Fig. 6 contains the estimates of AAWR for the four original GMST data records and the CW14 record. The use of the CW14 GMST results in slightly higher values of AAWR than HadCRUT, more in line with the values of AAWR from BEG. This result shows that utilizing a global temperature record does result in a modest increase in AAWR.



Proposed update to Figure 6. AAWR from the EM-GC and CMIP6 multi-model ensemble for 1975-2014. Five temperature data sets and five ocean heat content records are used to compare values of AAWR computed from the EM-GC. The box represents the 25th, 50th, and 75th percentiles, the whiskers denote the 5th and 95th percentiles, and the stars show the minimum and maximum values of AAWR from the EM-GC based upon the aerosol weighting method described in Sect. 2.5. The red box labeled "CMIP6" shows the 25th, 50th, and 75th percentiles, the whiskers represent the 5th and 95th percentiles, and the stars denote the minimum and maximum values of AAWR from the 50 member CMIP6 multi-model ensemble.

Proposed update to Fig. 7 contains the values of ECS utilizing the EM-GC for the four original GMST data sets as well as the CW14 record. The estimates of ECS using the CW14 GMST data set are larger than the estimates of ECS using the HadCRUT record. Similar to the values of AAWR, the values of ECS using the CW14 record are more in line with the BEG record than the HadCRUT record.

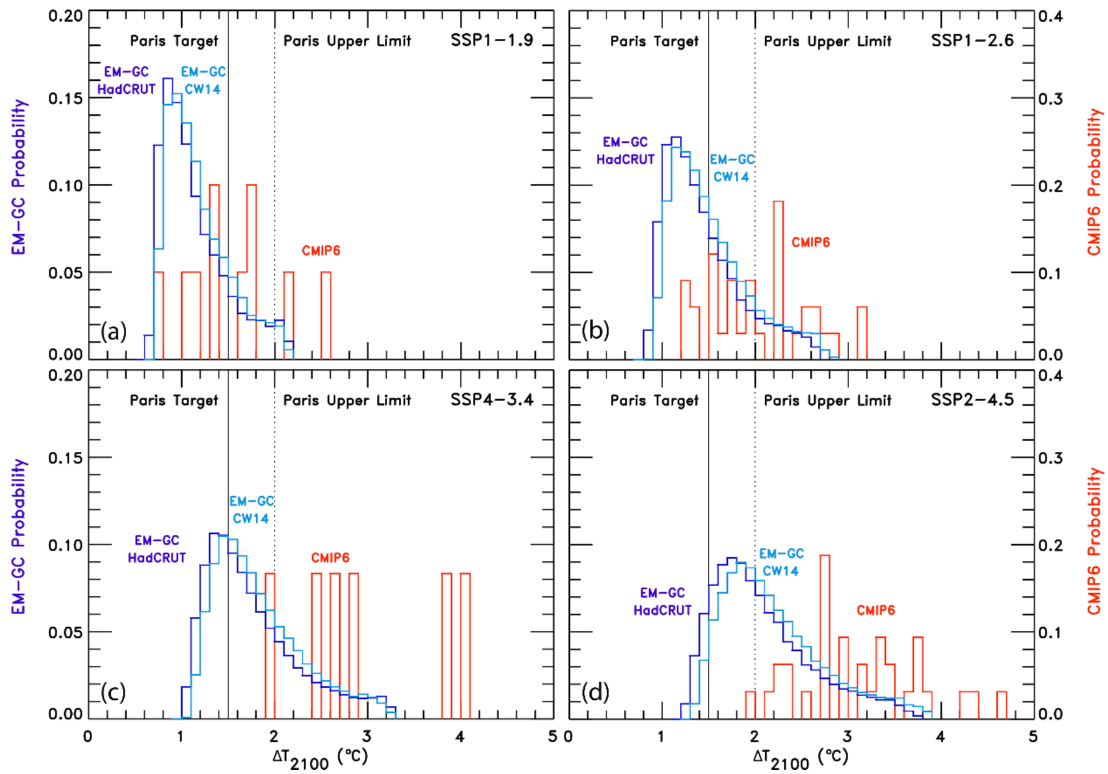


Proposed update to Figure 7. ECS from the EM-GC and the CMIP6 multi-model ensemble. Five GMST data sets and five ocean heat content records are used to compare values of ECS computed from the EM-GC. The box represents the 25th, 50th, and 75th percentiles, the whiskers denote the 5th and 95th percentiles, and the stars indicate the minimum and maximum values of ECS using the EM-GC based upon the weighting method described in Sect. 2.5. The circles denote the value of ECS associated with the best estimate of AER RF_{2011} of −0.9 W m^{-2}. The red box labeled "CMIP6" represents the 25th, 50th, and 75th percentiles, the whiskers denote the 5th and 95th percentiles, and the stars indicate the minimum and maximum values of ECS from the 28 member CMIP6 multi-model ensemble.

Proposed update to Fig. 12 contains the probability density function (PDF) for estimates of ΔT_{2100} using both the HadCRUT/Average OHC record and the CW14/Cheng record. The use of the CW14/Cheng record results in a PDF that is shifted towards higher values of ΔT_{2100}, which results in lower probabilities of achieving the Paris Agreement target (1.5°C) and upper limit (2.0°C), as shown in Proposed Update to Table 1.

The addition of the CW14 GMST data set combined with the Cheng OHC record makes our results more robust and shows the modest changes in AAWR, ECS, and ΔT_{2100} that occur upon changing the GMST record.

To address the second comment made by the reviewer, current Fig. S4 shows the five OHC records we use in our analysis normalized to 1986. This figure shows how the various OHC records relate to each other. All of the records follow the same overall trend, of increasing OHC from the beginning of the data record to the end. We would like to include all five data records for completeness. Our Proposed update to Fig. 6 and Fig. 7 illustrates that the choice of OHC record does have a slight impact on the values of AAWR and ECS but are not as important as the uncertainty in AER RF_{2011}. Upon revision, we include the combination of the CW14 GMST data set and the Cheng OHC record to show the effect of the OHC record that results in one of our highest kappa values, which leads to the most warming of all the OHC data records.



Proposed update to Figure 12. Probability density functions (PDF) for ΔT_{2100} found using the EM-GC with the HadCRUT temperature record (dark blue), the EM-GC with the CW14 temperature record (light blue) and Cheng 2017 OHC data set, and CMIP6 multi-model results. (a) PDF for EM-GC (blue) results and CMIP6 multi-model (red) results for SSP1-1.9. The left-hand y-axis is for EM-GC probabilities and the righthand y-axis is for the CMIP6 multi-model ensemble probabilities. (b) PDF for SSP1-2.6. (c) PDF for SSP4-3.4. (d) PDF for SSP2-4.5.

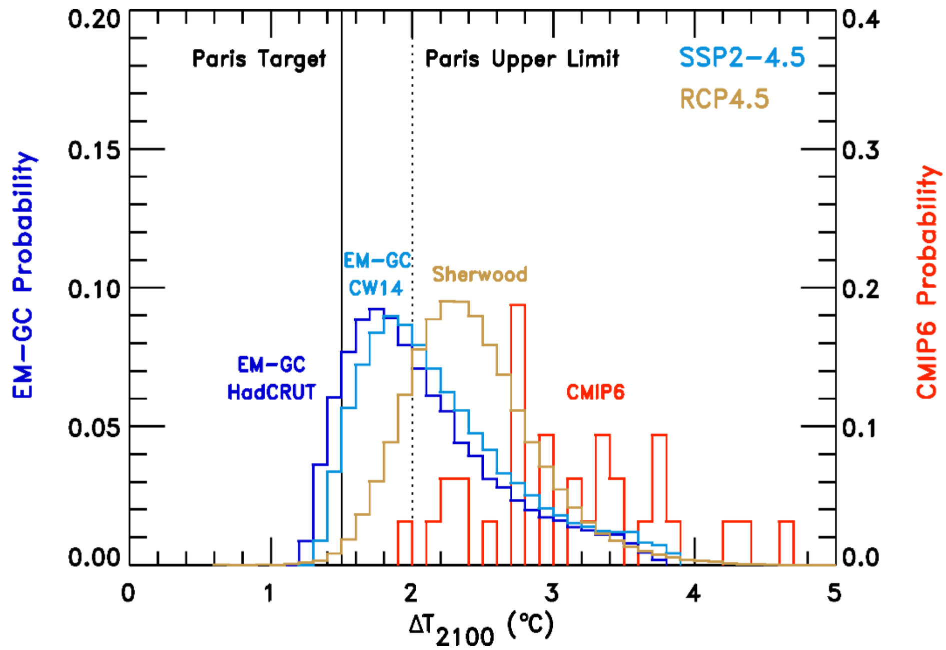
Proposed Update to Table 1. List of SSP scenarios analyzed in this study and the probabilities of achieving the Paris Agreement target or upper limit based on the EM-GC using the HadCRUT4 temperature record and average of the five OHC records and the CMIP6 multi-model ensemble. The second half of the table shows the probabilities of achieving the Paris Agreement target or upper limit based on the EM-GC using the CW14 temperature record and Cheng 2017 OHC record. The probabilities using the EM-GC are computed using the aerosol weighting method. The probabilities using the CMIP6 models are computed by calculating how many of the models for that scenario are below the temperature limits compared to the total number of models.

	Probability of Staying at or Below 1.5°C		**Probability of Staying at or Below 2.0°C**	
	EM-GC	**CMIP6**	**EM-GC**	**CMIP6**
SSP1-1.9	84.1%	50.0%	96.7%	80.0%
SSP1-2.6	64.8%	15.2%	88.4%	51.5%
SSP4-3.4	37.6%	0.0%	74.0%	16.7%
SSP2-4.5	10.5%	0.0%	53.1%	3.1%
SSP4-6.0	0.6%	0.0%	26.6%	0.0%
SSP3-7.0	0.0%	0.0%	1.3%	0.0%
SSP5-8.5	0.0%	0.0%	0.0%	0.0%
	Using CW14 and Cheng OHC Record			
SSP1-1.9	82.4%		97.5%	

SSP1-2.6	57.0%	85.5%
SSP4-3.4	28.1%	69.6%
SSP2-4.5	4.2%	43.2%
SSP4-6.0	0.0%	17.4%
SSP3-7.0	0.0%	0.0%
SSP5-8.5	0.0%	0.0%

We conclude this section by addressing the statement made by the reviewer, in the second (introductory) paragraph, that the stark contrast between our results and the recent assessment by Sherwood et al. (2020) may partially be an artefact of our data choices. The figure below, Response Fig. 1, shows the probability density function (PDF) for the rise in GMST in year 2100 (ΔT_{2100}) found using our model trained by GMST from HadCRUT (dark blue) and CW14 and OHC from Cheng et al. (2017) (light blue), and from the CMIP6 GCMs (red lines). All three of these lines are for SSP2-4.5. We also show a PDF for RCP-4.5 from Fig. 23 of Sherwood et al. (2020) (brown). Since projections found using our EM-GC are based on SSP2-4.5 and the Sherwood projection is based on RCP4.5, the comparison is not exactly "like to like", but all of these ΔT_{2100} projections are for GHG scenarios designated to reach a rise of about 4.5 W m^{-2} by end of century (relative to pre-industrial). Note the large region of overlap between our projections of ΔT_{2100} and those of Sherwood et al. shown in this figure. The results in Response Fig. 1 show that differences between our projections of ΔT_{2100} and those from Sherwood et al. (2020) are not, in fact, ***primarily*** due to our use of data for GMST from HadCRUT rather than CW14. However, our use of data from CW14 does move our projections closer to those of Sherwood et al. (2020), so the reviewer is correct that data choice does in fact contribute ***partially*** to this difference. Projections of ΔT_{2100} found using our approach, and those given by Sherwood et al. (2020), both fall far short of the CMIP6 projections. The Sherwood et al. (2020) analysis lies in between our projections and those of the CMIP6 GCMs because, as noted starting on line 721 of our submitted paper, "they rely on a determination that the feedback due to clouds is moderately to strongly positive …". Text on lines 721 to 740 of our submitted paper provides further explanation of the fundamental difference between results found using our approach and that of Sherwood et al., which is primarily due to uncertainty in cloud feedback.

We will be happy to include Response Fig. 1 as a figure in the supplement of our revised paper and add the text in the preceding paragraph to the paper (Main and Supplement) to highlight the comparison between our results and those from Sherwood et al. (2020), if so directed either by the reviewer or editor.



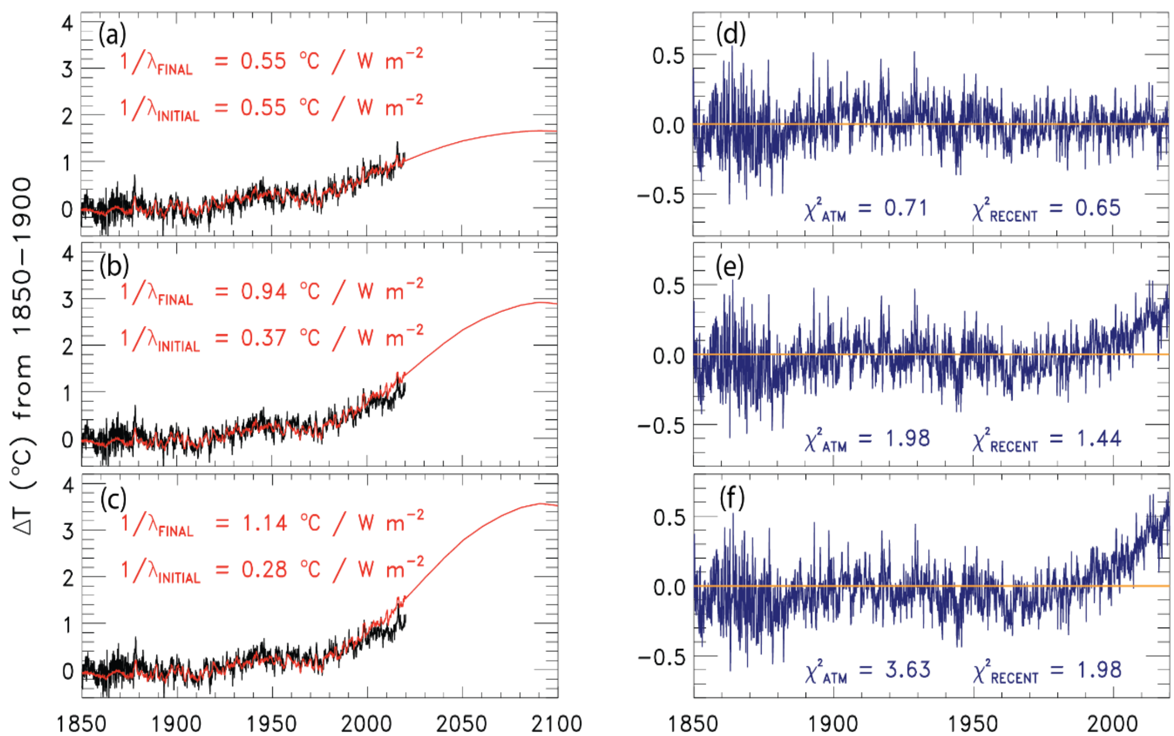
Response Figure 1. Probability density functions (PDF) for ΔT_{2100} found using the EM-GC with the HadCRUT temperature record (dark blue), the EM-GC with the CW14 temperature record (light blue), the CMIP6 multi-model results (red), and results for RCP4.5 from Fig. 23 from Sherwood et al. (2020).

3) Armour (2017) showed that climate sensitivity estimates from energy budgets can be reconciled with climate models by treating models as observations: if you estimate climate sensitivity of models using only data up to the present, your climate model sensitivity will be underestimated. The reason is that most models show increased sensitivity over time. The climate feedback parameter in McBride et al is assumed to be time-constant without justification. On timescales longer than the 150 years of the Gregory method, positive feedbacks are set to increase even further (Rugenstein, 2020).

Excellent point that we plan to address upon revision, as detailed below.

The assumption of the time-constant λ_Σ is the simplest assumption one can make. The fact the climate record can be fit so well based upon this assumption does provide support for the validity of this assumption. Nonetheless, as stated by the reviewer, many GCMs indicate that climate feedback varies over time. In our EM-GC framework, we are able to conduct calculations allowing the value of λ_Σ to vary over time, and to project future temperature with such an assumption. We therefore propose to add a new figure, which would be New Fig. 6, to address this concern raised and to add a new Section 2.6 where we describe model results for both constant climate feedback, and time varying climate feedback.

Our proposed New Fig. 6 shows the change in observed and modeled GMST under several assumptions. The first assumption is that the value of λ_Σ is constant over time. Of course, as noted in the submitted paper, we are able to fit the climate record over the past 170 years (χ^2_{ATM}) and past 80 years (χ^2_{RECENT}) quite well using this assumption. If we allow the value of λ_Σ to rise over time so that the value of χ^2_{ATM} is always less than or equal to 2, we obtain the result shown in New Fig. 6b. The modeled change in GMST starts to deviate from the observations around year 2000. This deviation is seen in the residual between modeled and observed GMST in New Fig. 6e. If we allow the value of λ_Σ to vary over time so that the value of χ^2_{RECENT} is less than or equal to 2, we get the result shown in New Fig. 6c. The modeled change in GMST starts to deviate dramatically from observations around year 1990. This stark deviation is seen in the residual between modeled and observed GMST in New Fig. 6f. The χ^2_{ATM} value in New Fig. 6f is 3.63, which does not satisfy our reduced chi-squared constraints, and interestingly appears to resemble the behavior of some CMIP6 GCMs. Upon revision, we would show New Fig. 6 and highlight the sensitivity of our projections of the rise in GMST by year 2100, for the SSP4-3.4 scenario, to the assumption of whether or not climate feedback is constant over time. Appropriate words would be added to the abstract and conclusions, such that the reader will be well aware that our baseline projections are based on time-invariant climate feedback, and that if the true climate feedback actually rises over time as suggested by some of the CMIP6 GCMs, our projections of global warming would be strongly affected.



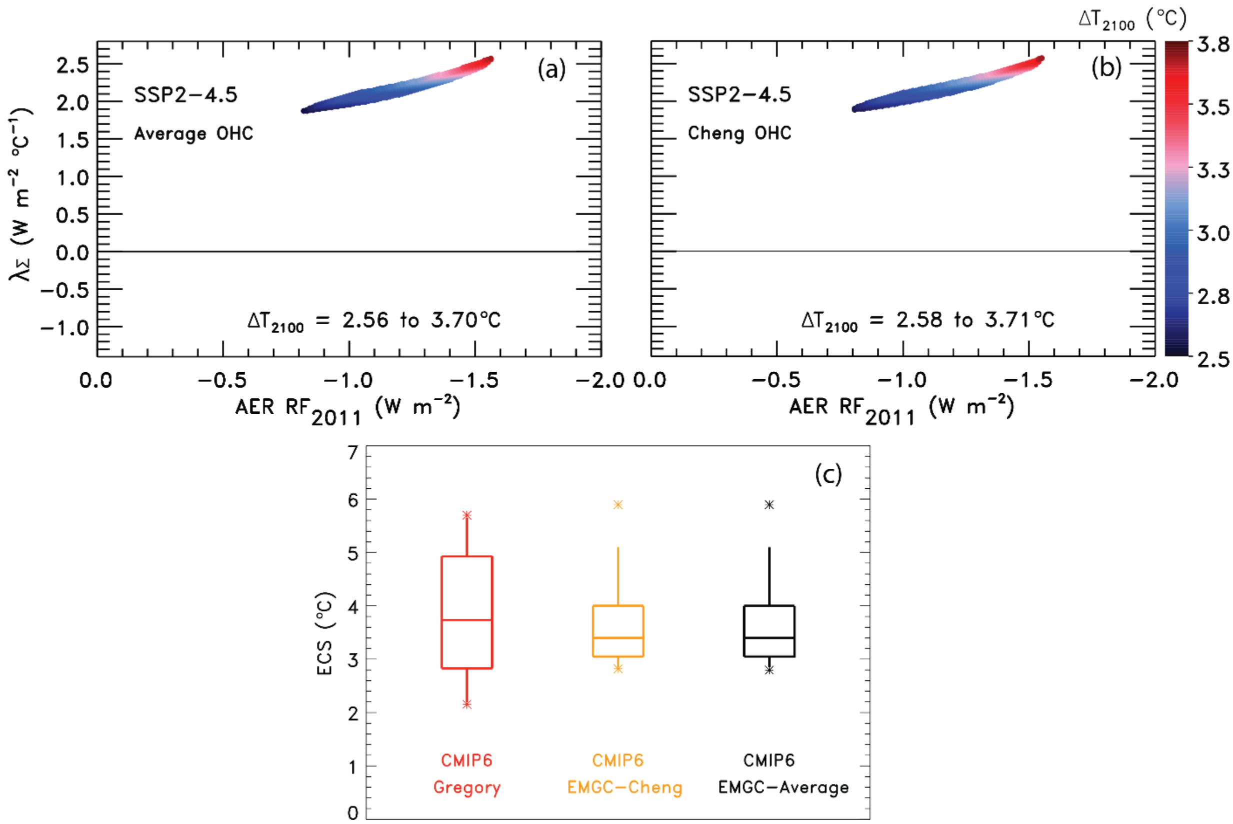
New Figure 6. Change in GMST from 1850-2019 for observed from HadCRUT (black) and 1850-2100 for modeled (red) using SSP4-3.4 and the residual between modeled and observations. (a) Rise in GMST assuming a constant value of λ_Σ. (b) Rise in GMST allowing λ_Σ to vary while the value of χ^2_{ATM} is kept below 2. (c) Rise in GMST allowing λ_Σ to vary while the value of χ^2_{RECENT} is kept below 2. (d) Residual between modeled and observed rise in GMST from 1850-2019 for constant λ_Σ. (e) Same as (d) but for varying λ_Σ while the value of χ^2_{ATM} is kept below 2. (f) Same as (e) but for varying λ_Σ while the value of χ^2_{RECENT} is kept below 2.

To address the reviewer's second point, we utilize the Gregory et al. (2004) method to calculate ECS from the CMIP6 GCMs because this procedure is preferred by Eyring et al. (2016) for the use in CMIP6. Our use of this method results in our calculated values of ECS being consistent with what GCM groups utilize for ECS, as has been published in numerous recent papers and will almost certainly be shown in the upcoming IPCC report. We agree with the reviewer that the Gregory et al. (2004) method may underestimate ECS, as shown in Rugenstein et al. (2020). Nonetheless, our paper will have greater value to the community if we use calculations for ECS that are consistent with the primary method employed by CMIP6. Upon revision, we will add more clarification to Sect. 2.4 about why we have chosen to use the Gregory et al. (2004) method, and we will note that this method may underestimate the value of ECS as suggested by the reviewer, with a citation to the Rugenstein et al. (2020) paper.

4) Considering the previous comment, I would like to be convinced the simple method can be used to estimate climate sensitivity when applied to climate models. Is this method able to give accurate predictions of climate sensitivity of climate models in contrast to previous energy balance methods?

Thanks for another excellent point that we plan to address upon revision, as detailed below.

Based on this comment, we have for the first time used our approach in the EM-GC framework to calculate ECS for the CMIP6 models. To use the EM-GC with the CMIP6 output, we calculated the CMIP6 multi-model mean change in GMST from 1850-2100 using the SSP2-4.5 scenario, because it reaches a doubling of preindustrial CO_2 by the end of the century. We used the standard deviation of the CMIP6-multi model mean to represent the uncertainty in the rise in GMST for our reduced chi-squared calculations. We trained the EM-GC from 1850-2100, included the CMIP6 prescribed values of SAOD and TSI, and did not include any natural variability. Our results are shown below in the proposed New Figure S13.



New Figure S13. GMST anomaly in 2100 relative to pre-industrial (ΔT_{2100}) as a function of climate feedback parameter and AER RF_{2011} and the values of ECS from the CMIP6 GCMs using three methods. (a) ΔT_{2100} for SSP2-4.5 using the CMIP6 multi-model mean and the average of the five OHC records. (b) ΔT_{2100} for SSP2-4.5 using the CMIP6 multi-model mean and the Cheng 2017 OHC record. (c) Values of ECS found using the Gregory et al. (2004) method (red), CMIP6 multi-model mean using the Cheng 2017 OHC (orange), and the CMIP6 multi-model mean using the average of the five OHC records. The box represents the 25th 50th, and 75th percentiles of the values of ECS and the whiskers denote the 5th and 95th percentiles. The stars indicate the minimum and maximum values of ECS.

The New Fig. S13 shows the values of λ_Σ and AER RF_{2011} for which the EM-GC calculates good fits to the CMIP6 multi-model mean for SSP2-4.5 from 1850-2100, using the average of five OHC record to represent the amount of heat going into the ocean (New Fig. S13a) or the Cheng 2017 OHC record (New Fig. S13b). Use of either the average of five OHC record or the Cheng 2017 OHC record results in similar values of λ_Σ, AER RF_{2011}, and ΔT_{2011}. We use the aerosol weighting method described in Sect. 2.5 of the submitted paper to calculate the box and whiskers shown in New Fig. S13c.

The box and whisker plots in New Fig. S13c using the EM-GC approach are similar to the box and whisker plot using the Gregory et al. (2004) method. The EM-GC approach does not yield values of ECS as low as some of the CMIP6 GCMs exhibit upon use of the Gregory et al. (2004) method, but the median values are relatively similar. The EM-GC approach is able to obtain values of ECS around the maximum value reported using the Gregory et al. (2004) method. The reviewer has raised an excellent point, and we contend that the inclusion of New Fig. S13 in supplement, complemented by a new paragraph in Section 3.2 of the Main paper, constitutes an important revision to our paper that will be of great interest to the community.

5) I could not quite understand the computations behind the TCRE: how are uncertainties in the carbon cycle taken into account? This is important for the 66% and 95% likelihood estimates.

Thanks for yet another excellent point that we also plan to address upon revision, as detailed below.

We had not taken the carbon cycle into account. To do so, we examined Friedlingstein et al. (2014) and Murphy et al. (2014) and found that the uncertainty in their estimates of atmospheric CO_2 concentration from emissions driven runs from the CMIP5 coupled carbon cycle models is about 10% (1-sigma). Upon revision, we propose to use this 10% value that relates uncertainty in CO_2 to carbon emissions, to represent uncertainty in the global carbon cycle for our estimates of the remaining carbon budget, before certain temperature thresholds are passed. We propose to update Table 2 as shown below and update the corresponding text in Sect. 3.3.4 and the conclusions to reflect this addition of uncertainty in the global carbon cycle, to our estimates of total cumulative carbon emissions.

Proposed Update to Table 2. Total cumulative and future carbon emissions that will lead to crossing the Paris temperature thresholds based on the EM-GC. Estimates of $\Sigma CO_2^{EMISSIONS}$ that would cause global warming to stay below indicated thresholds for 95%, 66%, and 50% probabilities. The values in the top half of the table are the estimates of total cumulative carbon emissions that will lead to crossing the Paris Agreement thresholds with the 10% uncertainty. The values in the bottom half of the table are the estimates of future cumulative carbon emissions after 2019 that will lead to crossing the Paris Agreement thresholds. The range of years given represents when the Paris Agreement thresholds will be passed based upon the rate of emissions from SSP5-8.5 or continuing the current rate of emissions of 11.7 Gt C yr^{-1}.

Total $\Sigma CO_2^{EMISSIONS}$ since 1870 from the EM-GC			
	95%	**66%**	**50%**
1.5°C	746 ± 75 Gt C	906 ± 91 Gt C	974 ± 97 Gt C
2.0°C	933 ± 93 Gt C	1203 ± 120 Gt C	1323 ± 132 Gt C
Future $\Sigma CO_2^{EMISSIONS}$ (assuming 638 Gt C released between 1870-2019)			
	95%	**66%**	**50%**
1.5°C	108 ± 75 Gt C (2022[a]-2032[a]) (2021[b]-2034[b])	268 ± 91 Gt C (2032-2042) (2034-2049)	336 ± 97 Gt C (2036-2045) (2039-2056)
2.0°C	295 ± 93 Gt C (2033[a]-2043[a]) (2036[b]-2052[b])	565 ± 120 Gt C (2046-2056) (2057-2077)	685 ± 132 Gt C (2051-2061) (2066-2088)

[a] Year the 1.5°C target or 2.0°C upper limit will be exceeded assuming the rate of emission inferred from SSP5-8.5 and the 1-sigma uncertainty

[b] Year the 1.5°C target or 2.0°C upper limit will be exceeded assuming current rate of emission of 11.7 Gt C yr^{-1} and the 1-sigma uncertainty

6) The paper is quite long, and I think that it will become more convincing after a good look at the prose. My minor comments will give further suggestions.

We thank the reviewer for this helpful suggestion, and we will take a careful look at the prose upon revision.

Minor comments:
76. Remove the word "active", as that implies a dynamic ocean, which is not what the model has

Thank you, we will fix

82. The paper uses many capitalised abbreviations, which is inevitable. However, words like months and obs can be written in lowercase to make reading more pleasant.

Thank you, we will fix

101. Maybe repeat what delta TMDL means

Thank you, we will add

132. I'm not sure whether your definition of lambda can also be called a climate feedback parameter. It would be confusing to have two different parameters with the same name

The literature is littered with numerous definitions of "climate feedback", and various names for our lambda. We have called this quantity the "climate feedback parameter" in our prior papers (Canty et al., 2013; Hope et al., 2017), so for the sake of conforming to past precedent, we prefer to retain this description of lambda in the current paper. The only other time we talk about another climate feedback parameter is when we describe that our quantity uses a different formalism from Bony et al., 2006 and Gregory, 2000, on page six of the submitted paper. We had hoped this text would illuminate, rather than confuse. This is a minor point and we will be happy to remove the sentence in question, if so directed.

136. This sentence can be removed, as it's not providing any information relevant to the study.

We would like to keep this sentence because it gives support to the relationship of λ_Σ and γ used in our analysis. Again, this is a minor point and we will be happy to remove the sentence in question, if so directed.

188. Normally the reduced chi-squared parameter is denoted $\chi\nu 2$ to differentiate from normal chi-squared.

The reviewer is correct; a superscript of "nu" is used in formal mathematics literature, for reduced chi-squared. Again, we have omitted this subscript in our prior papers (Canty et al., 2013; Hope et al., 2017) and for the sake of conforming to past precedent as well as for readability (i.e., less clutter), we prefer to not add this subscript. However, we will be delighted to add the subscript, if so directed by either the reviewer or editor.

Figure 1: The AAWR in panel b is different from the lead, which one is correct?

Thanks for this great catch! The value of AAWR shown in panel b is specifically for an aerosol radiative forcing pathway with a value of radiative forcing equal to −0.9 W m^{-2}. The value of AAWR reported in the abstract is the median value. We had used the wrong wording in the abstract and referred to the median value as the best estimate; upon revision, we will fix this mistake.

239-243. In this study, the datasets are referred to by the name of the institutions responsible for them, but they have specific names. Could you replace CRU with Had-CRUT, GISS with GISTEMP and so forth.

Great suggestion. We will replace CRU with HadCRUT, GISS with GISTEMP, NCEI with NOAAGT, and add CW14 for Cowtan and Way, 2014.

242: Typo. Berkley=Berkeley

Great catch; much thanks. We will fix.

247. Transformation usually means adjusting the mean and variance, where you're only adjusting the mean

Thank you we will replace "transformation" with "adjust".

257. The baseline is defined as no mitigation, so this sentence would be corrected if you remove that word

We are referring to Figure 1 of O'Neill et al. (2014) which describes the "challenges" space spanned by the SSPs. These are socioeconomic challenges for adaptation and socioeconomic challenges for mitigation. O'Neill et al. (2014) defines the challenges to mitigation as "socioeconomic factors that would make the mitigation task easier or harder for any given target and mitigation policy". We would like to edit the sentence to read "The baseline pathways follow specific narratives for factors such as population, education, economic growth, and technological developments of sources of renewable energy (Calvin et al., 2017; Fricko et al., 2017; Fujimori et al., 2017; Kriegler et al., 2017; van Vuuren et al., 2017) to represent several possible futures spanning different challenges for adaptation and mitigation to climate change as illustrated in Fig. 1 of O'Neill et al. (2014)." to draw the reader to this important paper that describes the creation of the SSPs.

263-269. Description of the tiers is unnecessary for this study, consider dropping it.

We would like to keep a short description of the tiers in this study, because the two different tiers were the basis for choosing which SSPs we highlighted in the main paper.

We propose eliminating the detailed description of the two tiers and will just describe which SSP scenarios are part of tier 1 and tier 2.

272. Add “the” (“in the supplement”)

Thank you, we will fix.

279-281. Which equation comes from which source?

All of the equations used to calculate radiative forcing come from Myhre et al. (1998). We use the updated values of preindustrial concentrations from Myhre et al. (2013) and the updated radiative efficiencies from WMO (2018). Upon revision, we will update the text to better describe from which source we obtain information for the radiative forcing calculations.

289. Remove brackets around Myhre

Thank you, we will fix.

295. Upon -> on?

Thank you, we will fix.

320. Are you sure it’s not perfectly identical?

The values of aerosol radiative forcing provided by the SSP Database are the same in 2005 and are slightly different in 2010, hence our wording of nearly identical. We would like to change the text has written to, “The database provides AER RF from 2005-2100, with values for all SSPs nearly identical until about 2010”. Upon examination of the aerosol radiative forcing time series on the SSP database, the time series start to deviate around 2015.

323. Remove “described above”, it’s unnecessary

Thank you we will fix.

363-369. If I understand it correctly, three different time series are appended. Would it not be easier to derive the entire time series yourself? That would also be easier to describe.

Yes, we did append three time series to create the MEI.v2 time series for ENSO in our model framework. While it may be easier to derive the entire time series ourselves, we want to train to data whenever possible. Hence our decision to append the two MEI time series together. We created the MEI time series ourselves from 1860-1871 because the MEI.ext did not extend past 1871. We propose to rework the text in the first paragraph of Sect. 2.2.6 to clarify our method for arriving at the full MEI.v2 time series.

415-416. Normalization involves both the mean and standard deviation, offsets are always additive. Maybe rewrite as “the five datasets are all set to zero in 1986 by applying an offset”

Thank you, we will change as suggested.

433. I didn't understand which standard deviation of the mean was taken

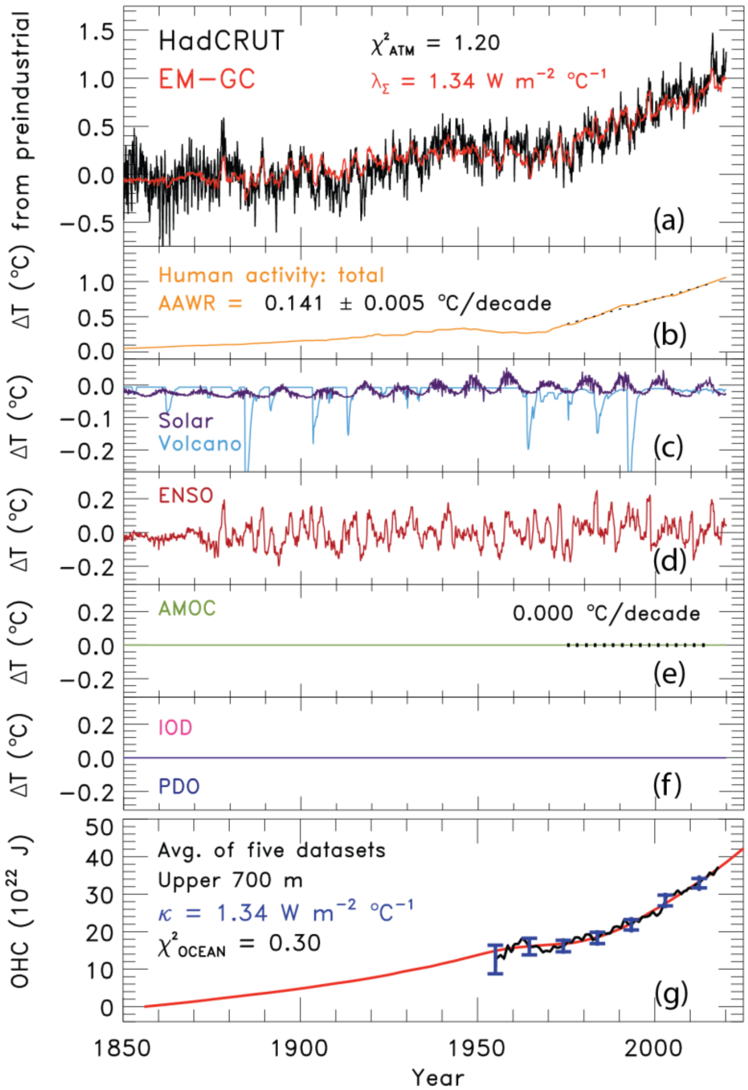
We are referring to the 1 sigma standard deviation of the average of the five OHC records. We will clarify this in the text.

463. Replace "based upon" by "using", remove "shown below"?

Thank you, we will fix.

477-480. I didn't understand why AAWR is not affected at all, as regression variables, such as lambda, are surely influenced by the inclusion of AMOC.

The value of AAWR and λ_Σ are only slightly affected by not including AMOC. We get a worse fit to the climate record when not including AMOC. Response Fig. 2 shows a plot similar to Fig. 1 from the paper, found upon not including AMOC, IOD, or PDO in the regression. As shown in Response Fig. 2a, the value of λ_Σ has decreased from 1.38 W m^{-2} $°C^{-1}$ (submitted paper) to 1.34 W m^{-2} $°C^{-1}$. The value of AAWR declines from 0.144°C/decade to 0.141°C/decade. The decrease in the value of AAWR is well within the 2σ uncertainty of AAWR. Our estimate of λ_Σ and AAWR are much more sensitive to the value of AER RF_{2011}, as shown in Fig. 3 and Fig. 6. Our value of χ^2_{ATM} increases from 0.71 to 1.20. The inclusion of AMOC, as well as PDO and IOD, is important to achieve a better fit to the climate record, allowing more of the parameter space for climate feedback and aerosol RF of climate to be considered upon use reduced of our chi-squared ≤ 2 filtering.



Response Figure 2. Measured and modeled GMST (ΔT) relative to pre-industrial (1850-1900) baseline. (a) Observed (black) and modeled (red) ΔT from 1850-2019. (b) Contributions from total human activity. This panel also denotes the numerical value of AAWR from 195-2014 (black dashed) as well as the 2σ uncertainty in the slope. (c) Solar irradiance (light blue) and major volcanoes (purple). (d) Influences from ENSO on ΔT. (e) Influence from AMOC is set to zero. (f) Influences from PDO (blue) and IOD (pink) are set to zero. (g) Measured (black) and modeled (red) OHC as a function of time for the average of five data sets.

519. if I understand it correctly, these equations assume there is no uncertainty at all in the radiative forcing at the doubling of CO2, which is inconsistent with definitions of radiative forcing and with CMIP6 models.

> We do not include uncertainty in the RF of climate due to a doubling of CO_2, and prefer to retain this formality, because this uncertainty is so small compared to so many other uncertainties. We are co-authors of a paper entitled "Reduced Complexity Model Intercomparison Project Phase 2: Synthesising Earth system knowledge for probabilistic climate projections", led by Zebedee Nicholls of the University of Melbourne, that was

submitted to the AGU journal Earth's Future on 12 Nov 2020. About half of the Reduced Complexity Climate Models used in this paper considered this uncertainty for projections of global warming out to 2300; our model along with about the other half did not consider this uncertainty. The author team of this Nicholls et al. paper decided to proceed without restricting uniformity for the treatment of this uncertainty, since it has such a small impact. That said, if so directed by either the reviewer or editor, we will gladly consider this uncertainty, but we note it will add a considerable computational burden for completion of our proposed revision, and we are certain it will not change our results in any meaningful manner.

550-552. I did not understand what an asymmetric Gaussian was, could you explain?

In our study, we use the term asymmetric Gaussian to refer to the fact that the distribution of the likely range and 5^{th} and 95^{th} percentiles of the values of AER RF_{2011} from Myhre et al. (2013) are not distributed symmetrically from the best estimate of AER RF_{2011} = −0.9 W m^{-2}. The likely range of AER RF_{2011} is given as −0.4 W m^{-2} and −1.5 W m^{-2}. The −0.4 W m^{-2} value is 0.5 W m^{-2} from the best estimate, while the value of −1.5 W m^{-2} is 0.6 W m^{-2} from the best estimate. The 5^{th} and 95^{th} percentiles of AER RF_{2011} are given as −0.1 and −1.9 W m^{-2}, which are 0.8 W m^{-2} and 1.0 W m^{-2} from the best estimate, respectively. We fit a Gaussian to the best estimate, likely range, and 5^{th} and 95^{th} percentiles. Because the likely range and 5^{th} and 95^{th} percentiles are not arranged symmetrically around the best estimate, the Gaussian is asymmetric.

Upon revision, we propose to add clarifying text to help readers understand this concept, to our method Section 2.5.

649. Remove "as indicated"

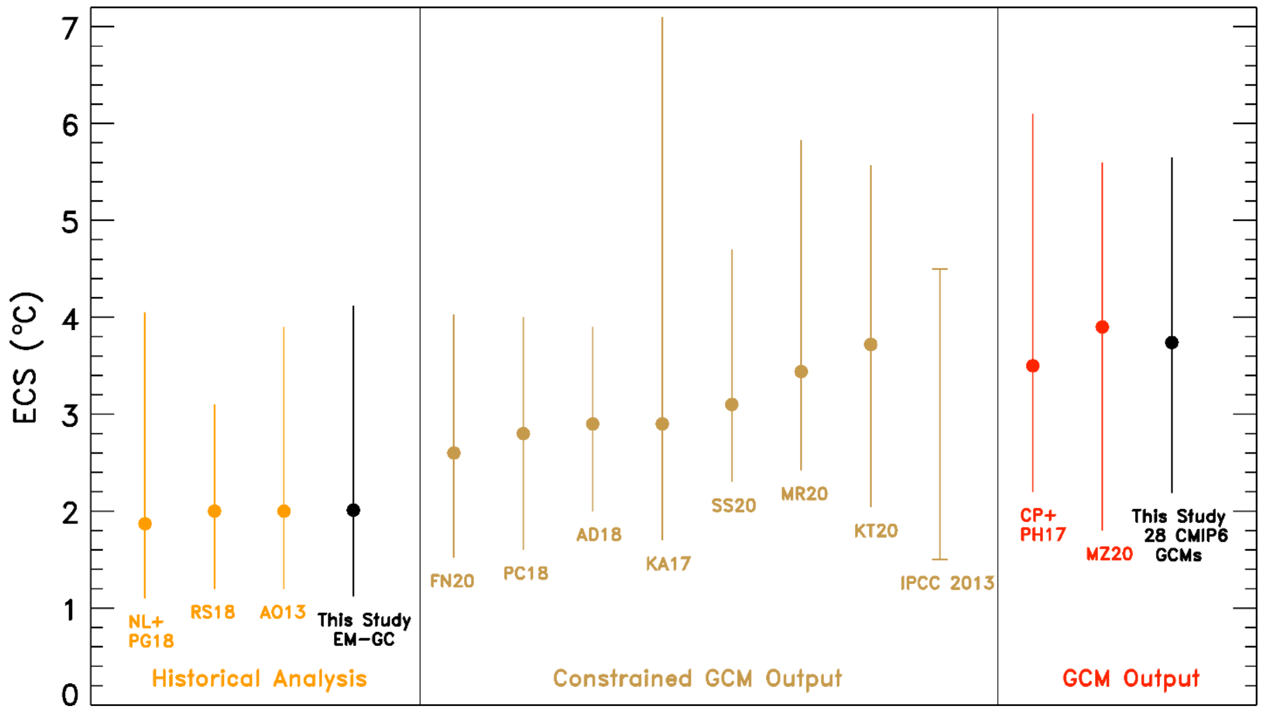
Thank you we will fix.

675. The value of 1.85 contradicts the value in the next paragraph of 2.01. Which one is correct?

Thanks for noting this contradiction. Both values are correct. The value of 1.85°C is the median value of ECS found using the HadCRUT temperature record. The value of 2.01°C is the best estimate value of ECS found using the HadCRUT temperature record. The best estimate corresponds to an AER RF_{2011} value of −0.9 W m^{-2}, whereas the median does not. We will add a sentence that describes the best estimate of ECS refers to the specific value of AER RF_{2011} of −0.9 W m^{-2}.

Figure 8: This figure only uses studies with low climate sensitivity and compares them to assessments of climate sensitivity (Sherwood/IPCC). Either explain the selection criteria, or add some studies to make this figure more balanced (the carbon brief provides an excellent overview: https://www.carbonbrief.org/explainer-how-scientists-estimateclimate-sensitivity)

We propose to update this figure to include more estimates of ECS, as shown below. The new figure is now divided into three categories: studies that used a historical analysis, studies that examined output from the GCMs and constrained it in some way, and studies that examined output from the GCMs and did not use any constraints. The studies that are included are from manuscripts published in the last few years. We will update the modified text, accordingly.



Proposed Update to Figure 8. Values of ECS from the EM-GC (black), CMIP6 multi-model ensemble (black), and 13 other studies grouped by type of study. The studies are listed by first author first initial of their first name and first initial of their last name and the year of publication, unless there are only two others, in which case both authors are listed. Historical analysis includes Lewis and Grünwald (2018) as NL+PG18, Otto et al. (2013) as AO13, and Skeie et al. (2018) as RS18. Constrained GCM output includes Armour (2017) as KA17, Cox et al. (2018) as PC18, Dessler et al. (2018) as AD18, Nijsse et al. (2020)as FN20, Rugenstein et al. (2020) as MR20, Sherwood et al. (2020) as SS20, Stocker et al. (2013) as IPCC 2013, Tokarska et al. (2020) as KT20. GCM output includes Proistosescu and Huybers (2017) as CP+PH17 and Zelinka et al. (2020) as MZ20.

689. The word "yet" implies a contradiction. However, with the very wide uncertainty specified by the IPCC, these probably overheating models are still within range.

Thank you for pointing out this apparent contradiction. We will replace the word "yet" with "and" and add "analyzed in this study" after CMIP6 GCMs to remove the implied contradiction.

698. rm the word "actual"?

Thank you, we will remove "actual".

704-705. Consider deleting "ninety-five…multi-model ensemble" as I think it is an unnecessary detail. Presenter 713. Remove "then"?

Thank you we will remove this sentence and the "then" later on in the paragraph.

739. Remove information between brackets, repetition of information within paragraph.

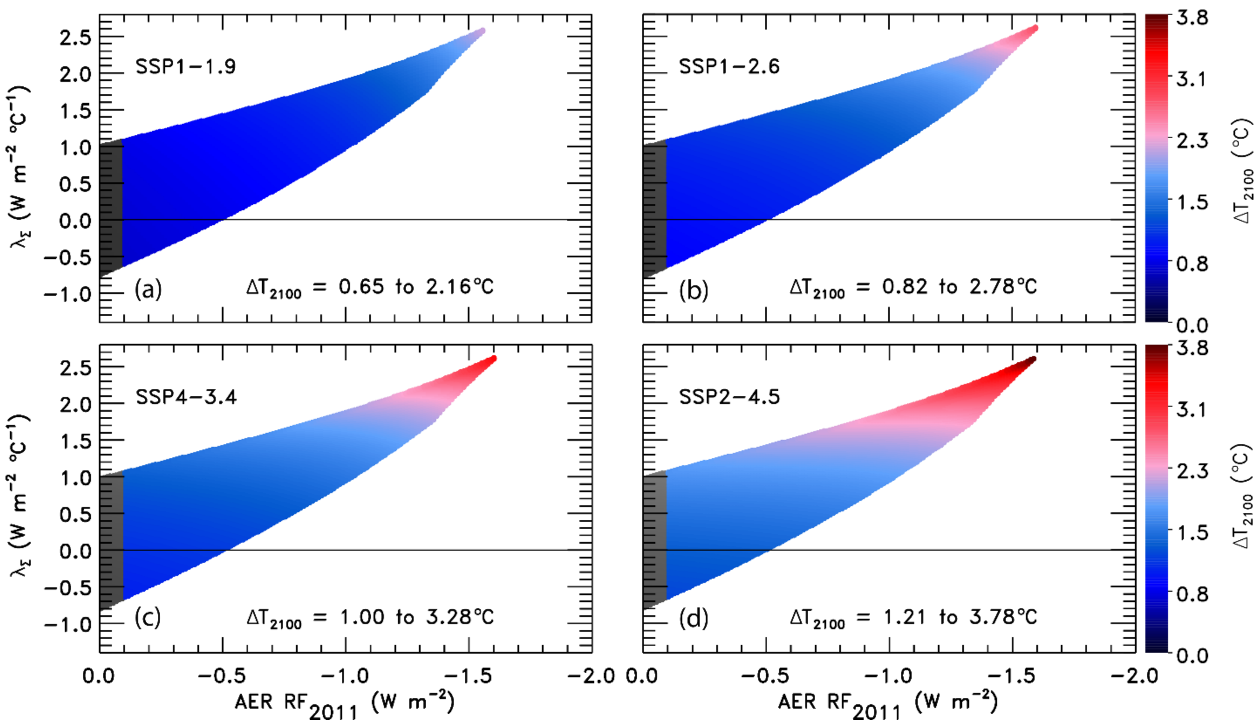
Thank you we will remove the 5th percent confidence interval that is written between the brackets.

782. Bifurcation has a specific meaning within mathematics, consider replacing by bimodality. If more models are added, check whether it's still true.

Thank you, we will change "bifurcation" to "bimodality".

Figure 10: Use different colour scheme. The rainbow colour scheme has false perceptual thresholds or hides real ones: https://www.nature.com/articles/519291d.

Thank you for pointing out this issue with the rainbow color scheme. We will replace the color scheme in Fig. 10, Fig. S1, and Fig. S8 with a blue-red color scheme. Below is an example of the revised color scheme for Fig. 10.



811. Replace "will" with "is set", we don't know the future.

Excellent point, we will change as suggested.

931. Replace "since" with "after".

Thank you, we will make this change.

669. Insert dioxide after carbon

We think the reviewer is referring to line 969; "carbon" will be changed to "carbon dioxide".

1002. Unnecessary to show all these percentiles, remove 25 and 75.

Thank you, we will change as suggested.

1009. 2017 was not an El Niño-year and non-El Niño-years 2018 and 2019 were comparable in temperature.

Thank you we will fix by clarifying the El Niño event ended in 2016.

1012. Similar, summarise, so do not show all percentiles

Thank you, we will fix.

1056. Similar, summarise.

Thanks again, we will fix.

1061. Replace "will" with "will not"

Great catch; thanks! We'll of course fix.

1071. Repetition of the information in 1061

We prefer to retain this sentence, to emphasize a key result from our analysis, but upon revision will gladly remove if so directed by the reviewer or editor.

1073. What is a literal interpretation? The model democracy interpretation?

We would like to remove the phrase "a literal interpretation of" so the sentence reads "We suggest there is slightly more time to achieve these steep reductions than indicated by the CMIP6 multi-model mean".

1074. Modeling is not the only source of information on warming of 1.5 degrees, many studies extrapolate current trends.

The purpose of this concluding paragraph is to mention the value of the CMIP6 GCMs. We are unsure exactly what studies the reviewer is referring to in their comment and would ask the reviewer to provide some examples if they believe it is important for us to mention these studies in this paragraph.

We would like to propose rewording the sentence to "The incredibly valuable output of the CMIP6 GCMs is important for determining the consequences for society of 1.5°C, 2.0°C, or even larger rises in GMST". This rewording removes the phrase "rely entirely on" that may have been the cause of the issue for the reviewer.

Figure S1: Replace the rainbow colour scheme.

Thank you, we will address by no longer using the rainbow color scheme, using instead the color scheme shown below:

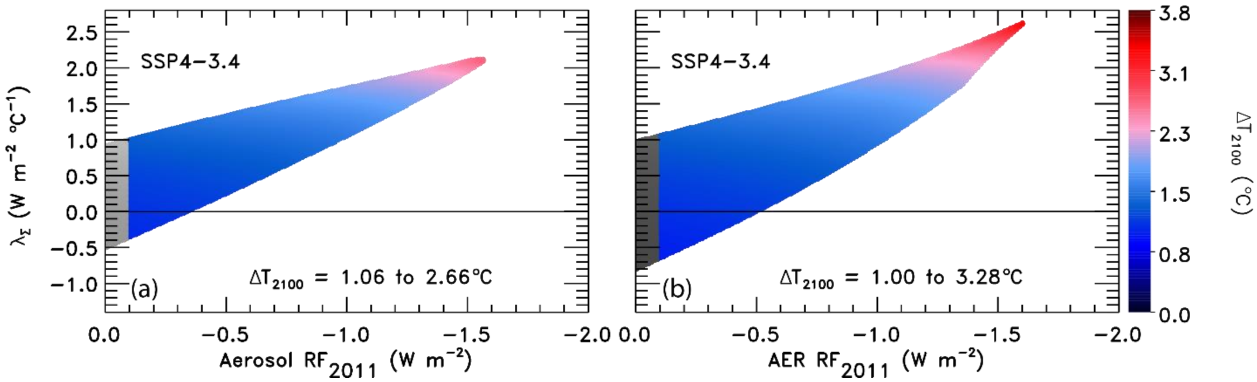


Figure S7: Caption should indicate that it's the unweighted one.

Thank you, we will fix.