

Summary of Relevant Changes to “The Fractional Energy Balance Equation for Climate projections through 2100” by Roman Procyk et. al. and Responses to All Referees

Roman Procyk et. al.

We would like to begin by thanking all three anonymous referees for their reviews. The following is a summary of the major changes to the paper following their suggestions:

- Overall improvements on written style (including citation style) to make the paper more readable.
- Inclusion of further references and expansion on the relevance of the FEBE in current literature about energy balance models.
- Several figures have been redone and corrected to improve clarity.
- The Bayesian inference method has been expanded, clarifying how the method is implemented.
- A discussion on approximating the joint prior probability density functions with a multivariate Gaussian along with how the covariance between the parameters used in the joint prior distribution is determined. Along with a further discussion on how projections are made using said distribution and Monte Carlo methods.
- New sections 4.1, 4.2, 4.3, 4.4 have been included further expanding on the novelty of the FEBE: its reliability, ability to recreate the amplitude of internal forcing and realistic fluctuations, and its accuracy in hindprojecting the 21st century slowdown (“hiatus”).
- Several tables summarizing key results have now been included (tables 4, 5, 6)
- The conclusion has now been split into two sections: Discussion and Conclusion with the hope this improves readability and clarity.

Interactive comment on “The Fractional Energy
Balance Equation for Climate projections
through 2100” by Roman Procyk et. al.

Roman Procyk et. al.

Figures 9 to 11 show the current and future evolutions of the global temperature as compared with the observations and the CMIP5 simulations. These projections are provided as ensembles in order to cover the range of uncertainties related to the model settings. An observational uncertainty is also provided. It is obvious that the spread of the ensemble is much smaller for the FEBE model than for the CMIP5 ensemble in Figure 9, and very often the observation band is not falling into the range of predicted projections of FEBE, while it is the case for the CMIP5. This suggests that the projections are unreliable in the sense that it does not cover all possible situations that can be observed. In weather and climate sciences, this aspect is key when making forecasts, predictions and projections using ensembles. This unreliability should be first acknowledged in the manuscript. Second this unreliability is maybe related to the use of a weak stochastic forcing. Some experiments with larger stochastic forcing would be desirable in order to clarify the under-dispersion of the ensemble. Stronger stochastic forcing could also maybe lead to larger climate sensitivity. This should be checked too.

Author: The basic purpose of Figures 9 to 11 was to compare the FEBE and GCM projections. In both cases, the projections are deterministic but with uncertainty limits due to their respective model uncertainties. Both yielded an estimate of the forced response but with qualitatively different uncertainty bounds. In the case of GCMs, the uncertainty is termed “structural” while for the FEBE it is parametric uncertainty. Unlike a probabilistic forecast, the results cannot be interpreted with the help of stochastic forecast notions such as reliability.

We can see why the referee may have misunderstood, since we compared the forced and internal components with the observed temperature series. This was intended as a quick visual validation of the forced component but since it contains the (stochastic) internal variability, strictly speaking, it should not be directly compared to the forced component.

In order to make a proper comparison with data, we first removed the ob-

servations from the figures so as to only compare the forced component (new figs. 10, 11 to replace the original figs. 10, 11). All the forced components are deterministic (the internal variability has been averaged out), for the GCMs the uncertainty is structural uncertainty, while for the FEBE it is parametric uncertainty. We then replaced the original figure 9 with one that represents the ensemble average over all the parametric and internal variability of the FEBE and the mean observational temperature (shown in new figs. 9a, 9b). This was achieved since the statistical dependence of the internal forcing and the parametric uncertainty are independent: the errors therefore add in quadrature. For this, it is sufficient to take the globally averaged yearly temperature anomaly ($\approx \pm 0.11C$) and combined it with the annual resolution parametric forced component from figs. 10, 11 over the historical period (1880-2020). The new figs. 9a,9b shows this result. This can also be done at monthly resolution following the same procedure but using the globally averaged monthly temperature anomaly ($\approx \pm 0.14C$), shown in figs. 9c, 9d.

The temperature observations do indeed fall within the 90% confidence limits of the FEBE historical reconstruction (i.e. the ensemble average of the response to both internal and external forcing). In both figures at annual resolution shown below (figs. 9a,9b), the historical mean temperature (red) is within the 90% CI of the FEBE forced response (with internal variability added) 92% of the years using the RCP scenario, and 94% using the SSP scenario. At the monthly resolution shown in figs. 9c, 9d, the historical mean temperature (red) is within the 90% CI of the FEBE forced response (with internal variability added) 90% of the months using the RCP scenario or the SSP scenario. The uncertainty is therefore compatible with the data. This is not the same as the reliability but it is an analogous validation of probabilistic aspects of the projection.

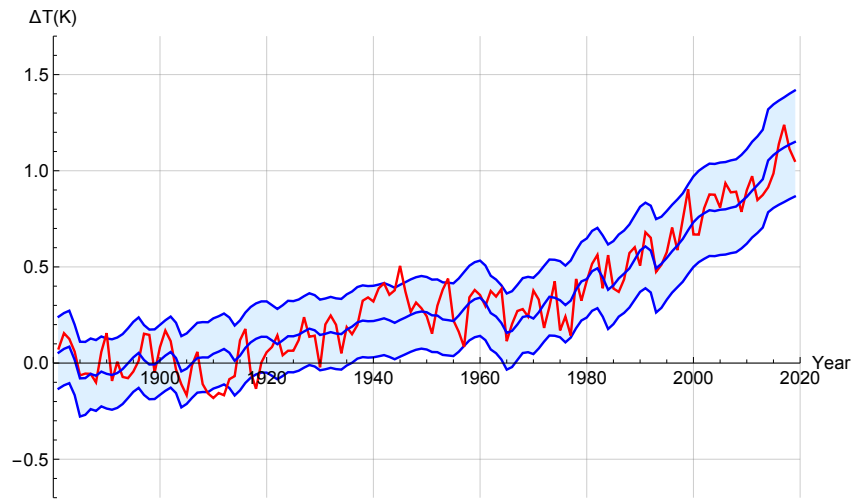


Fig. 9a: The historical reconstruction (forced temperature response and internal variability) of the FEBE, with parameters calibrated using F_{AerRCP} (blue) alongside mean of 5 observational temperature series (red) at yearly resolution; 90% CI (due to parametric uncertainty and internal variability) are indicated (shaded).

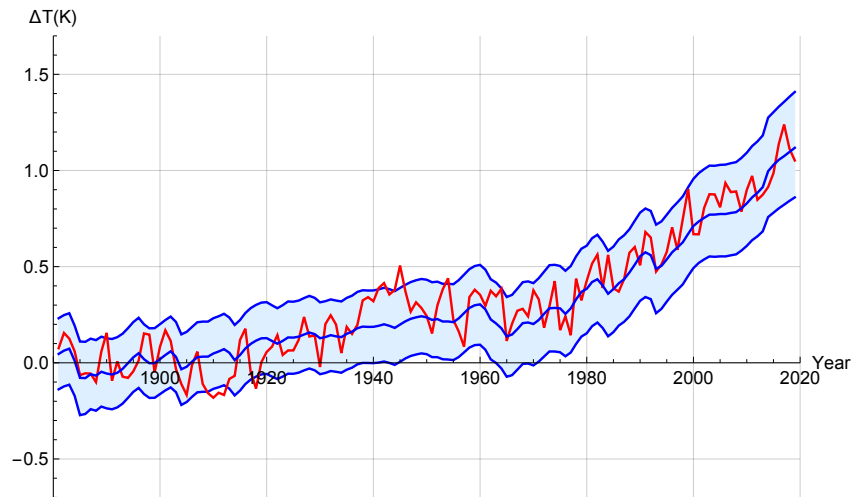


Fig. 9b: Same as fig. 9a except using F_{AerSSP} parameters and forcing.

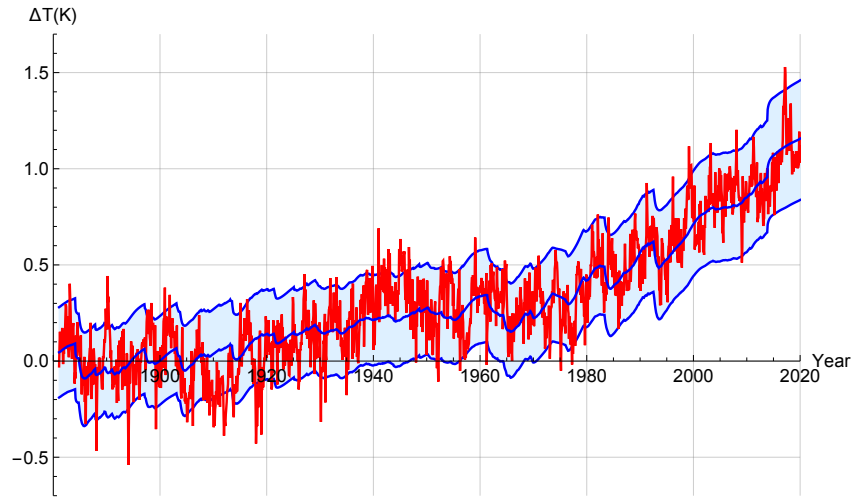


Fig. 9c: The historical reconstruction (forced temperature response and internal variability) of the FEBE, with parameters calibrated using F_{AerRCP} (blue) alongside mean of 5 observational temperature series (red) at monthly resolution; 90% CI (due to parametric uncertainty and internal variability) are indicated (shaded).

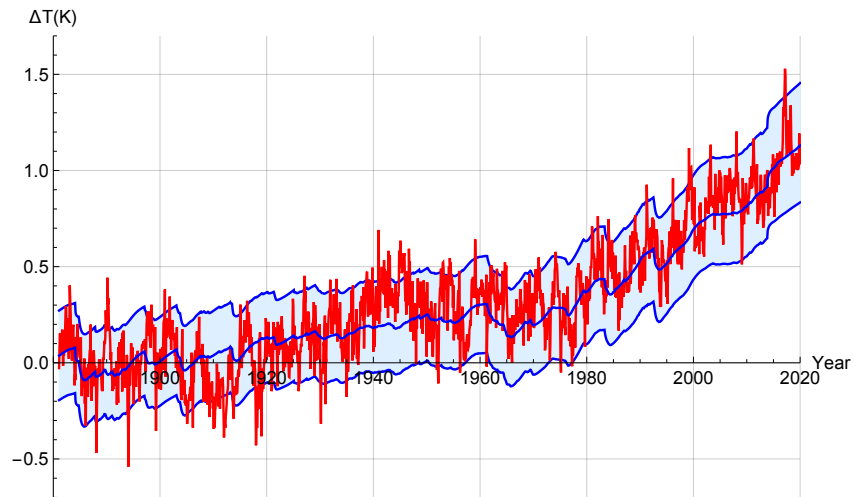


Fig. 9d: Same as fig. 9c except using F_{AerSSP} parameters and forcing.

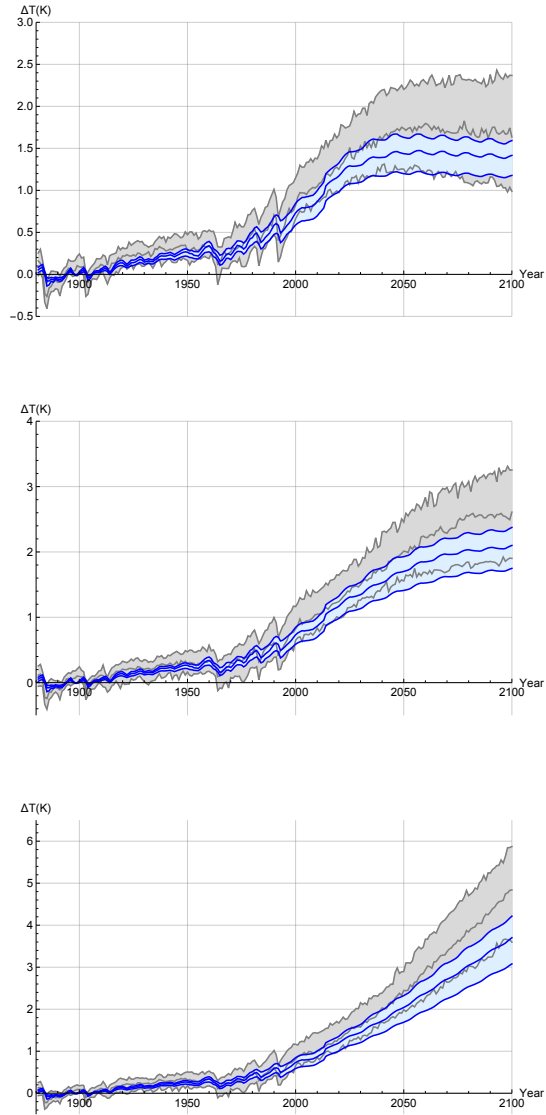


Fig. 10: The deterministic forced temperature response projected using the FEBE, with parameters calibrated using F_{AerRCP} (blue) compared with the CMIP5 MME projection (black); 90% CI from the parametric uncertainty are indicated (shaded). The projections until 2100, for RCP 2.6 (top), RCP 4.5 (middle) and RCP 8.5 (bottom), are shown.

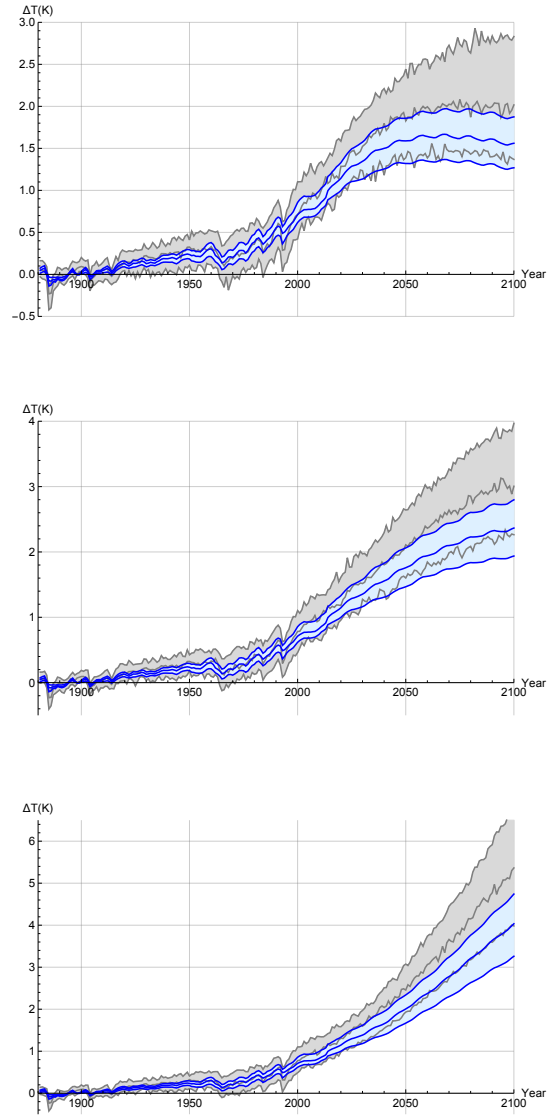


Fig. 11: The deterministic forced temperature response projected using the FEBE, with parameters calibrated using F_{AerSSP} (blue) compared with the CMIP6 MME projection (black); 90% CI from the parametric uncertainty are indicated (shaded). The projections until 2100, for SSP 126 (top), SSP 245 (middle) and SSP 585 (bottom), are shown.

At page 20, line 421, the authors claim that the hiatus is better represented in the FEBE than in the CMIP5 ensemble. Well to me this is not true as the observations are most of the time out of the range of FEBE. The CMIP5 looks better at capturing the observations. So I suggest to modify these comments and try to be more objective in the comparison, maybe by using measures of reliability.

Author: We discussed the reliability in the previous responses, by showing the observed temperature series along with the projection of the forced response plus the internal variability response. With the internal variability we expect the data to lie within the 90% CI, 90% of the time. We also discussed how historical reconstructions were made (figs. 9a, 9b, 9c, 9d). We now discuss the latter over the hiatus period. In the insets of figs. 1a, 1b, we show the comparison of the forced median FEBE projection (blue) (using F_{AerRCP} in fig. 1a, and F_{AerSSP} in fig. 1b), the CMIP5/6 MME median (black) and the mean of 5 observational temperature series (red) with the 90% CI over the historical period with the inset showing the hiatus period (a blow-up of 1998-2015). We see that indeed, the FEBE median forced component in both cases captures the hiatus rather accurately (see Lovejoy (2015) for a stochastic forecast with a similar high frequency limit).

A quantitative comparison between the amount of time the FEBE median response is within the bounds of the observational temperature series 90% CI and the same for the CMIP5/6 MME was performed at the annual resolution data. The amount of time the median FEBE forced component using F_{AerRCP} is within the 90% CI of the observational temperature series over the whole historic period is 47% and over the hiatus is 70% in comparison to the CMIP5 MME median which is within the whole historic period 39% and over the hiatus is 17%. When using the median FEBE forced component using F_{AerSSP} similar results are found, over the whole period: 45% and over the hiatus: 35%, in comparison to the CMIP6 MME median which is within the whole historic period 39% and over the hiatus is 30%. It can be seen in both cases that the CMIP MME is generally warmer than the FEBE forced component notably over the period of the hiatus.

Lovejoy, S. (2015), Using scaling for macroweather forecasting including the pause, *Geophys. Res. Lett.*, 42, 7148–7155, doi:10.1002/2015GL065665.

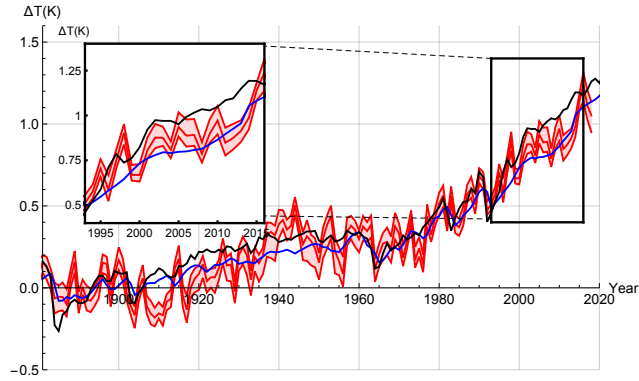


Fig. 1a: The median historical forced component of the FEBE, with parameters calibrated using F_{AerRCP} (blue), and the median of the CMIP5 MME (black) alongside mean of 5 observational temperature series (red) with their 90% CI indicated (shaded).

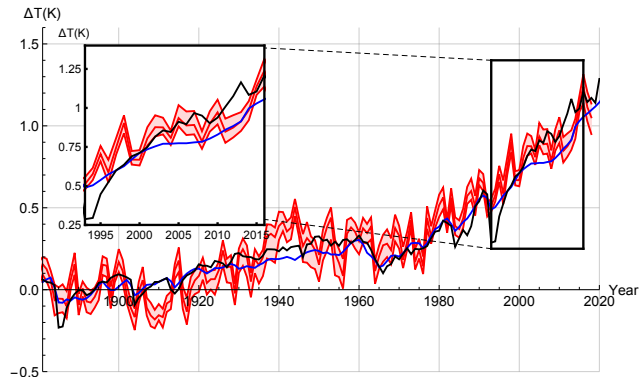


Fig. 1b: The median historical forced component of the FEBE, with parameters calibrated using F_{AerSSP} (blue), and the median of the CMIP6 MME (black) alongside mean of 5 observational temperature series (red) with the 90% CI indicated (shaded).

One key conclusion of the manuscript is the lower sensitivity of the FEBE model on long time scales as shown for instance in Figure 10. What is happening when the FEBE model is fitted with a stronger noise that could maybe help in increasing the spread of the ensemble and its reliability? The results and the conclusions should then be revisited once these experiments are done and the impact of the stochastic noise clarified.

Author: The the amplitude of the noise is not an adjustable parameter, it is determined from the data. Although the empirical amplitude of the internal variability does not alter the projections of the forced temperature response, it does affect the uncertainties of the parameters (once these have been estimated, the resulting projection is purely deterministic).

Equation 14: Please clarify what is the variance of gamma(t)

Author: The variance of $\gamma(t)$ is the amplitude of the internal forcing assumed to be a Gaussian white noise. The internal variability of the observational temperature is equal to the observed series with the forced temperature response removed. If we take the global annually averaged monthly temperature anomaly to be $\sigma_{T,\tau_r} \approx \pm 0.14C$, we can determine the variance of $\gamma(t)$ from Lovejoy et. al (2021):

$$K_h = \sqrt{\frac{\pi}{2\cos(\pi(h - \frac{1}{2}))\Gamma(-1 - 2h)}},$$

$$\sigma_{f,\tau_r} = \frac{\sigma_{T,\tau_r} K_h}{s} \left(\frac{\tau}{\tau_r}\right)^h.$$

Where K_h is a standard normalization constant chosen for convenience, τ is the relaxation time, τ_r is the resolution (taken to be monthly in this case), s is the climate sensitivity parameter, σ_{T,τ_r} is standard deviation of the globally averaged monthly temperature anomaly at resolution τ_r , h the scaling exponent of the temperature fluctuations, and $\sigma_{f,\tau_r} = \gamma(t)$ is the standard deviation of the internal forcing. Using our F_{RCP} (and F_{SSP}) parameter estimates, we find a mean estimate of the variance of $\gamma(t)$ to be ~~1.15~~ $3.2Wm^{-2}$ (~~1.29~~ $3.3Wm^{-2}$) and 90% CI of ~~[0.89, 1.42]~~ $[2.1, 4.2]Wm^{-2}$ (~~[0.99, 1.65]~~ $[2.3, 4.3] Wm^{-2}$). If we introduce a white noise forcing, $\gamma(t)$, with the variance calculated above to the FEBE we will be able to recreate the the amplitude of the internal temperature variability response. This will be included in the revision.

Harries et. al. (2010) sets out to examine the net energy flux balance at the top of atmosphere (TOA) measured using observations from polar-orbiting spacecraft. The early observations, using the Nimbus experiments, show an internal variability of the $4.1 \pm 4.0Wm^{-2}$, while more modern measurements (CERES) in the 2000s show variability of between ± 2 and $\pm 4 Wm^{-2}$ generally laying a few Wm^{-2} of zero. Thus our estimate of the internal forcing variability

is within estimates of the TOA net energy flux balance.

Lovejoy S, Procyk R, Hébert R, Rio Amador L. The fractional energy balance equation. QJR Meteorol Soc. 2021;1–25. <https://doi.org/10.1002/qj.4005>

Figure 6: The response of FEBE is compared with an IPCC two-box model. What is this model? Maybe I missed the place where it is described. Please describe this model in more details in the text

Author: The two-box model we are referring to is the classical linear two-layer energy-balance model described in Held et al. (2010) and found in IPCC AR5 (2013, section 8.SM.11.2):

$$C \frac{dT}{dt} = F - \lambda T - \gamma(T - T_0),$$
$$C_0 \frac{dT}{dt} = \gamma(T - T_0).$$

We graphically showed the comparison in fig. 6. The parameters for the two-box model are the best estimates from Geoffroy et al. (2013).

Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis, (2010): Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *J. Climate*, 23, 2418–2427. <https://doi.org/10.1175/2009JCLI3466.1>

Geoffroy O, Saint-Martin D, Olivié DJ, 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. *J Clim* 26:1841–1857. <https://doi.org/10.1002/env.2140>

Figure 12: Please correct the colors of the curves. I cannot figure out what is plotted

Author: Thank you for pointing this out, the colours will be changed as to be easily readable.

Interactive comment on “The Fractional Energy
Balance Equation for Climate projections
through 2100” by Roman Procyk et. al.

Roman Procyk et. al.

An assessment of scientific soundness is difficult at this point, in particular since information necessary for the reproduction of the results is not provided. There are errors in some of the figures (listed in detail below) and their quality as well as that of the general sentence structure could be improved for easier comprehension by the reader.

Author: The FEBE projections are deterministic i.e. for a given set of parameters, our projection represents the average over an infinite ensemble. However, the parameters are uncertain. Therefore, the projections have uncertainty even though each is deterministic. In order to numerically estimate the ensemble mean projection and the uncertainty in the projections it is convenient to use Monte Carlo methods. However it should be clear that this is simply a convenient numerical technique that does not imply any stochasticity in the projections. However, using Monte Carlo methods to determine the projections introduces standard Monte Carlo numerical uncertainty, but this is made quite small by using large numbers of Monte Carlo realizations.

For the FEBE, an ensemble of realizations is used throughout the paper. The ensemble for the FEBE calibrated using either the CMIP5/6 is included in the revision to allow for the reproduction of results; the full FEBE code will be publicly released upon publication and can be shared immediately with the reviewer.

There are a lot of gaps in the discussion of the model and results. Limitations of FEBE are not described at all, nor whether FEBE can be used for studies other than those concerning modern and future global mean temperature. The potential for future studies or extensions are also skimmed over.

Author: The fundamental limitations of the FEBE are “structural” (is the model a good model?), and parametric (how well have we estimated the parameters?). As usual, the first is difficult to answer, we have now replaced the original fig. 9 with figs. 9a, 9b which show a comparison of the ensemble averaged hindcast compared with past data. In both figures shown below (figs.

9a,9b) at annual resolution, the historical mean temperature (red) is within the 90% CI of the FEBE forced response (with internal variability added) 92% of the years using the RCP scenario, and 94% using the SSP scenario. At the monthly resolution shown in figs. 9c, 9d, the historical mean temperature (red) is within the 90% CI of the FEBE forced response (with internal variability added) 90% of the months using the RCP scenario or the SSP scenario. At the monthly resolution shown in figs. 9c, 9d, the historical mean temperature (red) is within the 90% CI of the FEBE forced response (with internal variability added) 90% of the months using the RCP scenario or the SSP scenario. The uncertainty is therefore compatible with the data and shows a high degree of internal consistency in the model and in the Bayesian parameter estimates. We also note in the revision that the end of war (1945) temperature spike which lays out of the FEBE 90% CI may be explained due to biases explained in Chan and Huybers (2021). Physically, the FEBE agrees with two dynamical conservation principles: scale invariance of the storage and energy conservation. The parametric uncertainty is the uncertainty that we can quantify, it arises essentially due to poor knowledge (uncertainty) in the past forcing. The FEBE has already been generalized to a regional model (Lovejoy 2021) and we are currently preparing regional projections using it. In the future, it will be extended to nonlinear feedbacks (e.g. temperature albedo feedbacks). We have added some material in the discussion and conclusions sections on this.

Lovejoy, S.: The Half-order Energy Balance Equation, Part 2: The inhomogeneous HEBE and 2D energy balance models, *Earth Syst. Dynam. Discuss.* <https://doi.org/10.5194/esd-2020-13>, 2021.

Chan, Duo, and Peter Huybers. "Correcting Observational Biases in Sea-Surface Temperature Observations Removes Anomalous Warmth during World War II". *Journal of Climate* (2021): 1-44. <https://doi.org/10.1175/JCLI-D-20-0907.1>.

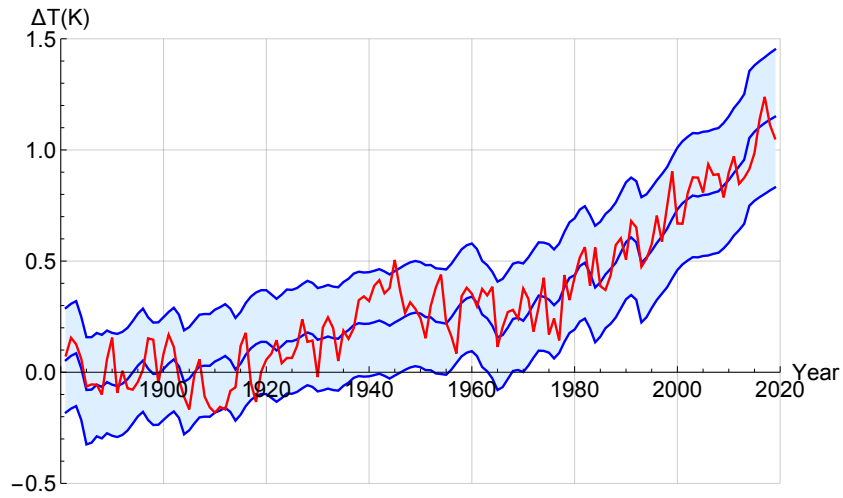


Fig. 9a: The historical reconstruction (forced temperature response and internal variability) of the FEBE, with parameters calibrated using F_{AerRCP} (blue) alongside mean of 5 observational temperature series (red) at monthly resolution; 90% CI (due to parametric uncertainty and internal variability) are indicated (shaded).

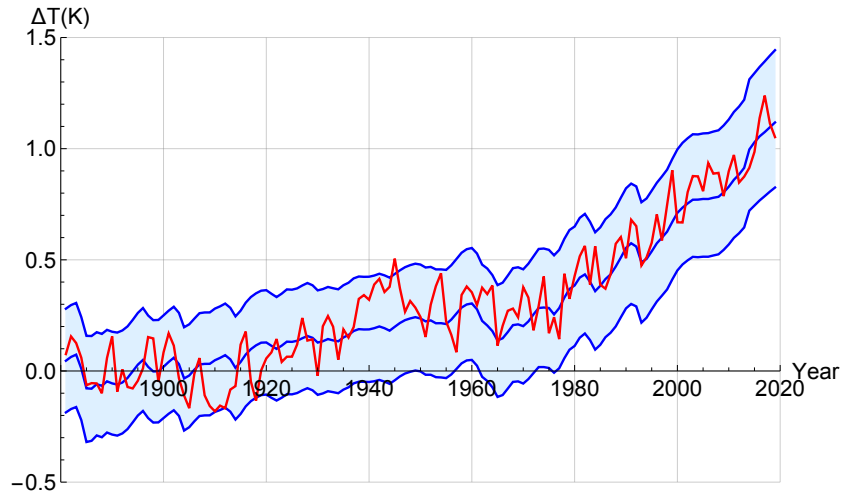


Fig. 9b: Same as fig. 9a except using F_{AerSSP} parameters and forcing.

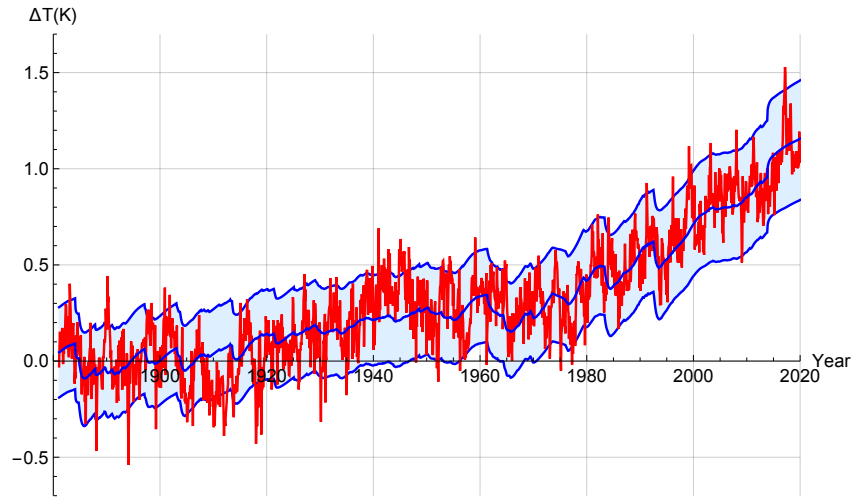


Fig. 9c: The historical reconstruction (forced temperature response and internal variability) of the FEBE, with parameters calibrated using F_{AerRCP} (blue) alongside mean of 5 observational temperature series (red) at monthly resolution; 90% CI (due to parametric uncertainty and internal variability) are indicated (shaded).

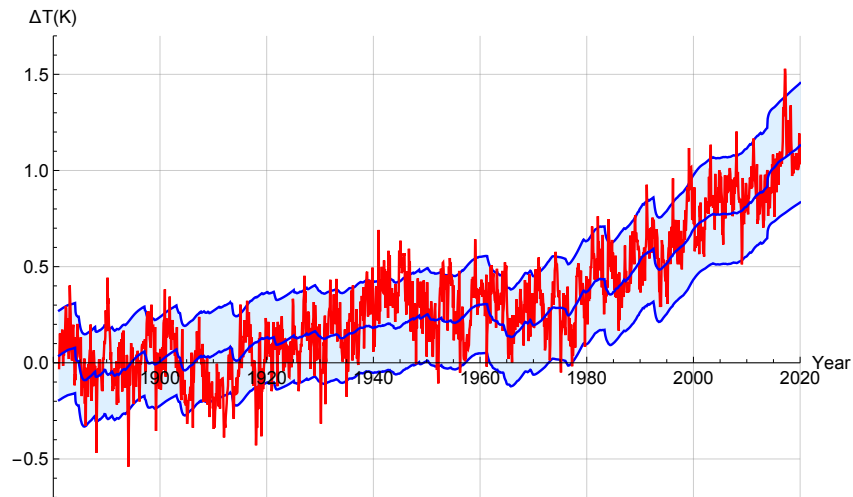


Fig. 9d: Same as fig. 9c except using F_{AerSSP} parameters and forcing.

The reader does not get an idea of what the paper adds to Hebert et al. (2020) aside from replacing the truncated power law in the model equation. Especially, with respect to the results a discussion of differences is needed.

Author: The main differences are at high frequencies as a consequence the FEBE models both the responses to the internal and external forcings and it can better take into account volcanic forcing. Whereas Hebert et. al. (2020) makes an ad hoc model for the internal variability, in the FEBE, the internal variability determines in the order h of the FEBE. A paragraph will be added exploring the differences between the model and results in Hebert et al. (2020).

The FEBE is able to estimate the forcing amplitude to reproduce the internal variability of the temperature (see Lovejoy et. al (2021)). This is discussed below and will be included in the text, highlighting the new aspect of the FEBE which allow us to physically model the internal variability compared to Hebert et al. (2020). A further discussion on the variance of $\gamma(t)$ along with estimates are in the response to referee 1, and will be included in the revision.

Lovejoy S, Procyk R, Hébert R, Rio Amador L. The fractional energy balance equation. QJR Meteorol Soc. 2021;1–25. <https://doi.org/10.1002/qj.4005>

The research is not sufficiently embedded in other literature on the topic of energy balance modeling in both the introduction and discussion.

Author: More background references and connections to standard literature will be included in the revision.

Trenberth, Kevin E., John T. Fasullo, and Jeffrey Kiehl. "Earth's Global Energy Budget". Bulletin of the American Meteorological Society 90.3 (2009): 311-324. <https://doi.org/10.1175/2008BAMS2634.1>

Rypdal, Martin, and Kristoffer Rypdal. "Long-Memory Effects in Linear Response Models of Earth's Temperature and Implications for Future Global Warming". Journal of Climate 27.14 (2014): 5240-5258. <https://doi.org/10.1175/JCLI-D-13-00296.1>

North, G. R., Kim, K. Y. (2017). Energy Balance Climate Models. John Wiley Sons.

Proistosescu, C., Donohoe, A., Armour, K. C., Roe, G. H., Stuecker, M. F., Bitz, C. M. (2018). Radiative feedbacks from stochastic variability in surface temperature and radiative imbalance. Geophysical Research Letters, 45, 5082–5094. <https://doi.org/10.1029/2018GL077678>

Ziegler, E. and Rehfeld, K.: TransEBM v. 1.0: Description, tuning, and

validation of a transient model of the Earth’s energy balance in two dimensions, Geosci. Model Dev. Discuss. <https://doi.org/10.5194/gmd-2020-237>, 2020.

L12 ”We found that aerosols...” should mention that here aerosols refers to the CMIP aerosol forcings

Author: Thank you, this will be clarified.

L22 It is unclear how FEBE complements (in addition to support) GCMs based on what is said.

Author: The FEBE projections to 2100 are entirely independent of the GCMs. Still, they are within the uncertainty bounds of the latter, effectively providing an independent confirmation of the GCM projections. This eliminates one of the key climate skeptic arguments: projections are not reliable since they are solely GCM-based.

It also has the advantage that the uncertainties are significantly reduced. This is important for policy (see the conclusions).

Another advantage is that it gives us information about the past forcings, especially the aerosols (also to a lesser extent about volcanic forcing). It supports numerous papers that find that the historic aerosol forcing in CMIP models was too strong (Stevens (2015), Zhou and Penner, (2017), Sato et al., (2018), Bellouin et al., (2020)). We could also diagnose what is wrong with models with respect to too strong aerosol and volcanic forcings by calibrating the FEBE on the historical temperature reconstruction of the model.

Stevens, B. ” Rethinking the Lower Bound on Aerosol Radiative Forcing”. Journal of Climate 28.12 (2015): 4794-4819. <https://doi.org/10.1175/JCLI-D-14-00656.1>

Zhou, C. and Penner, J. E.: Why do general circulation models overestimate the aerosol cloud lifetime effect? A case study comparing CAM5 and a CRM, Atmospheric Chemistry and Physics, 17, 21–29, <https://doi.org/10.5194/acp-17-21-2017>, <https://www.atmos-chem-phys.net/17/21/2017/>, 2017.

Sato, Y., Goto, D., Michibata, T., Suzuki, K., Takemura, T., Tomita, H., and Nakajima, T.: Aerosol effects on cloud water amounts were successfully simulated by a global cloud-system resolving model, Nature Communications, 9, 985, <https://doi.org/10.1038/s41467-018-03379-6>, <https://doi.org/10.1038/s41467-018-03379-6>, 2018

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., Boucher, O., Carslaw, K. S., Christensen, M., Daniau, A.- L., Dufresne, J.-L., Feingold, G., Fiedler, S., Forster, P., Gettelman, A., Haywood, J. M., Lohmann, U., Malavelle, F., Mauritsen, T., McCoy, D. T., Myhre, G., Mühlmenstädt,

J., Neubauer, D., Possner, A., Rugenstein, M., Sato, Y., Schulz, M., Schwartz, S. E., Sourdeval, O., Storelvmo, T., Toll, V., Winker, D., and Stevens, B.: Bounding Global Aerosol Radiative Forcing of Climate Change, *Reviews of Geophysics*, 58, e2019RG000660, <https://doi.org/10.1029/2019RG000660>, 2020.

The abstract summarized well what is done in the paper, but fails to address limitations and the gap in literature it aims to fill.

Author: Ok this will be revisited to better address the FEBE with respect to the literature and area of EBMs.

L35 reference for the projection uncertainty in response to CO2 doubling missing

Author: Thank you, the reference will now be included.

IPCC (2013), *Climate Change 2013: The Physical Science Basis (AR5)*, edited by T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, Cambridge University Press, Cambridge, U. K.

L43 should be motivated why it is desirable to construct models beyond linearity and stochasticity

Author: By themselves, these constraints are insufficient to specify any particular model, they remain too general. We have added a comment on this.

L69-72 These explanations require improvement: a) The EBE has not been explained enough for anyone who is not familiar with the model to fully comprehend this sentence. The reader has not been made aware that a storage term exists yet. b) Maybe include a short discussion of other for the storage term besides the one chosen.

Author: We have added a sentence explaining that the earth energy balance is not exact: at any instant there is a difference between incoming and outgoing radiation, this difference is stored (or comes out of storage). The usual storage is assumed to be in thermodynamic (molecular) degrees of freedom with one or two slabs or “boxes” instantaneously changing their temperatures to account for any radiative imbalances. (This is a property of global models such as the model discussed here: for regional models horizontal divergence of heat fluxes can also be important).

L118 Please mention why causality is not respected otherwise.

Author: The equation expresses causal antecedence: cause must precede effect. This is a necessary part of the causal principle, If this property is not held,

then the future forcings can affect the past temperatures (i.e. the convolution for any given time would integrate both past and future forcing).

L121 Please explain why the step response function is "more physical compared to the impulse response". It is not necessarily clear to the reader which response function will be used in which studies. This could be further elaborated in the text to clarify that issue.

Author: Dirac and Step response functions are mathematically equivalent with the latter corresponding to the more usual experiments (e.g. CO2 doubling). The step function response shows how the model approaches energy balance whereas the Dirac response shows the effect of a short perturbation. We will replace "more physical" with "easier to physically interpret" and make clear in the text that they are equivalent.

L130 a short explanation of what fractional Relaxation noise is would be helpful

Author: We will add a short explanation: it is the response of the FEBE to a white noise forcing.

The implementation of the model is not described at all.

Author: The model is calibrated over the historical temperature observations using a Bayesian scheme. This will be better introduced and explained throughout sec 2.3 in the revision.

It should be mentioned explicitly that the model produces annual data (same for the temporal resolution of the forcing).

Author: The model produces data at the same resolution as the forcing, be it annual or monthly (see figs. 9a, 9b for annual resolution, and figs. 9c, 9d for monthly resolution). This will be explicitly mentioned in the revision. In fact, monthly forcing and response was used to fully exploit the volcanic data and better constrain the model parameters. For clarity we showed annual resolution for all projection plots, this will be clearly indicated.

L135-138 Would help to start a new paragraph here to clarify that now the text refers back to equations 5 and 6 for the H=1 case and not to the ones discussed in the meantime. It would also be helpful to the reader if it was shown how the function from Hebert et al. With $H_f=0.5$ is the same as eq. 6 with $H=0.5$ and why this is the same as "that corresponding to the internal forcing".

Author: Thank you for mentioning this point. H_f is the asymptotic exponent determining the convergence of FEBE to a step forcing, mathematically it

is given by the asymptotic expansion of the step Green's function that shows that $H_f = -h$ (note that we changed the FEBE order "H" to "h" to avoid confusion with numerous exponents "H" used in similar circumstances). This will be further explained in the revision.

L145 It is unclear why 277ppm is the reference.

Author: It is a standard reference value, the reference value is the pre-industrial CO2 concentration taken from page 140. We added the reference to the text.

Forster, P., V. Ramaswamy, P. Artaxo, T. Berntsen, R. Betts, D.W. Fahey, J. Haywood, J. Lean, D.C. Lowe, G. Myhre, J. Nganga, R. Prinn, G. Raga, M. Schulz and R. Van Dorland, 2007: Changes in Atmospheric Constituents and in Radiative Forcing. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

L177 A short elaboration as to why a linear scaling factor is appropriate to describe the uncertainty would be nice.

Author: The primary uncertainty in the radiative forcing is the total uncertainty in anthropogenic aerosol forcing that arises from the uncertainty in aerosol radiative properties and cloud effects as well as in their concentrations over the industrial period (Penner et al. (2001), Ramaswamy et al. (2001)). We suppose aerosols are known only within a multiplicative-scale factor α , which is a unity-mean, normally distributed random variable. This scaling of the magnitude of aerosol forcing is an approach that has been adopted previously by Forest et al. (2002), Harvey and Kaufmann (2002), Forest et al. (2006), and Padilla et al. (2011).

J. E. Penner et al., in Climate Change 2001, The Scientific Basis, J. T. Houghton et al., Eds. (Cambridge Univ. Press, Cambridge, 2001), pp. 289-348.

V. Ramaswamy et al., in Climate Change 2001, The Scientific Basis, J. T. Houghton et al., Eds. (Cambridge Univ. Press, Cambridge, 2001), pp. 349-416.

Forest CE, Stone PH, Sokolov AP, Allen MR, Webster MD. Quantifying uncertainties in climate system properties with the use of recent climate observations. Science. 2002 Jan 4;295(5552):113-7. doi: 10.1126/science.1064419

Harvey, L. D. Danny, and Robert K. Kaufmann. Simultaneously Constraining Climate Sensitivity and Aerosol Radiative Forcing. Journal of Climate 15.20

(2002): 2837-2861. | [https://doi.org/10.1175/1520-0442\(2002\)015;2837:SCCSAA;2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015;2837:SCCSAA;2.0.CO;2).

Forest, C. E., Stone, P. H., and Sokolov, A. P. (2006), Estimated PDFs of climate system properties including natural and anthropogenic forcings, *Geophys. Res. Lett.*, 33, L01705, doi:10.1029/2005GL023977.

Padilla, Lauren E., Geoffrey K. Vallis, and Clarence W. Rowley. " Probabilistic Estimates of Transient Climate Sensitivity Subject to Uncertainty in Forcing and Natural Variability". *Journal of Climate* 24.21 (2011): 5521-5537. <https://doi.org/10.1175/2011JCLI3989.1>

Section 2.2.3 For RCP scenarios it is stated that an aerosol forcing smaller than 1Wm2 is implausible, but not why this is not also the case for the SSP forcing.

Author: The forcing value itself is implausible, of course this applies to any scenario. We do not say that it applies particularly to the RCP, we only have quoted the AR5 values.

Fig.1: a) lower plot has two sets of dashed data unlike the upper plot and unlike the plot description b) overlap between data and axis labels in lower plot could be prevented if the axis was at the bottom c) It is unclear why the forcing time series differ in length. d) The upper plot uses "Aer" and not "Ant" in the legend. e) Having the explanations for solid vs dashed vs dotted in the legend, too, would make the plots easier to understand. f) The plot sizes are different, would be nice if they were consistent.

Author: We have made the following changes to fig. 1, shown below.

- a) We have now included the SSP 245 scenario into our updated results. Both plots will have two sets of 3 different dashed lines representing RCP 2.6/SSP 126, RCP 4.5/SSP 245, and RCP 8.5/SSP 585.
- b) Ok, this will be improved.
- c) The series are equal of length - this is a graphical error and misunderstanding. The plots will be better aligned in the revision. All series run from 1765–2100, the historical (solid) section for the RCP scenarios (blue) goes to 2005, while for the SSP scenarios (black) it extends to 2015.
- d) The aerosol forcing of the top plot will be removed, so as to just compare the greenhouse gas forcing of the various scenarios.
- e) In the revised image we will have the solid lines be the historical forcing (historic forcings for each scenario group (RCP or SSP) are the same) and RCP 2.6/SSP 126, dashed will be for RCP 4.5/SSP 245, and dotted will be for RCP 8.5/SSP 585.
- f) Thank you for pointing this out, it will be fixed.

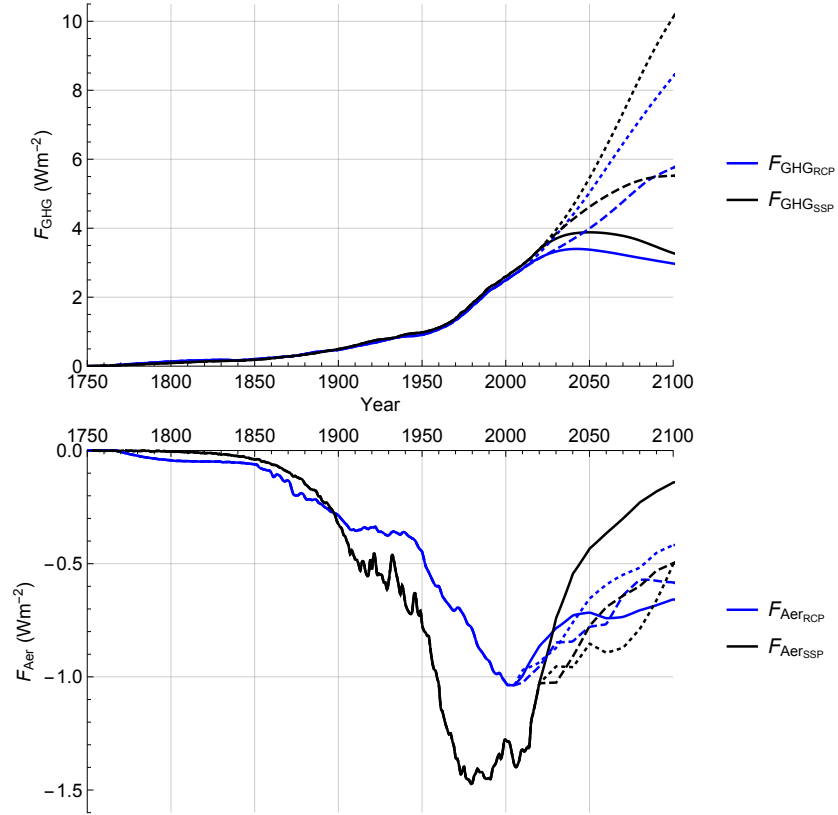


Fig. 1: (top) The total GHG forcing series, are shown over the historical period and projection period until 2100 for RCP 2.6/SSP 126 (solid), RCP 4.5/SSP 245 (dashed), and RCP 8.5/SSP 585 (dotted). (bottom) The anthropogenic aerosol forcing series used, F_{AerRCP} (blue) and F_{AerSSP} (black). Updated from (Hébert et al., 2020). All series run from 1765–2100, the historical (solid) section for the RCP scenarios (blue) goes to 2005, while for the SSP scenarios (black) it extends to 2015.

L202 It is unclear why 27 was chosen as the conversion factor. With regards to the literature this seems quite high and might also explain why the forcing was overpowered. AR5 recommends 25, Schmidt et al. (2011) 20 and for AR6 18 is suggested. Might be interesting to look at what the effective AOD would have been that the "damped" volcanic forcing the authors ended up using corresponds to.

Author: The -27 value was taken from Sato (2012), lower values could be used but this is something which the non-linear volcanic parameter ν takes into account. Performing a linear regression between the non-linear damped volcanic forcing using with best parameter estimate, $\nu = 0.28$ and the AOD we find a conversion factor ≈ -4 , thus even if lower values were used the forcing would still be overpowered.

Sato M (2012) Forcings in GISS climate model : stratospheric aerosol optical thickness. <https://doi.org/10.1038/nature04237>

L205 The point of comparison in "comparable effect" is missing.

Author: This sentence should be rewritten; the volcanic forcings are scaled down (typically using a linear factor by other authors) so to best fit the global temperature response in climate models as found by Lewis and Curry, 2015. See also:

Gregory, J.M., Andrews, T., Good, P. et al. Small global-mean cooling due to volcanic radiative forcing. *Clim Dyn* 47, 3979–3991 (2016). <https://doi.org/10.1007/s00382-016-3055-1>. which mention volcanism may be overpowered in a model simulation.

Tomassini L, Reichert P, Knutti R, Stocker TF, Borsuk ME (2007) Robust Bayesian uncertainty analysis of climate system properties using Markov chain Monte Carlo methods. *J Clim* 20:1239–1254

Ring MJ, Lindner D, Cross EF, Schlesinger ME (2012) Causes of the global warming observed since the 19th century. *Atmos Clim Sci* 2:401–415

Meinshausen M, Smith SJ, Calvin K, Daniel JS, Kainuma MLT, Lamarque J-F, Matsumoto K, Montzka SA, Raper SCB, Riahi K, Thomson A, Velders GJM, van Vuuren DPP (2011) The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim Change* 109(1–2):213–241

L208-210 An expansion of what it means that a quantification in a multi-fractal framework is possible would be helpful. The explanation of why it must be transformed remains unclear to me. (esp. relevant since a multifractality index is introduced in L218)

Author: This will be clarified in the revision, a brief explanation can be found in the response to the next comment.

L218 Would be helpful if the authors would expand on C1 being the codimension of the mean and what the implications of this are.

Author: Volcanic forcing is peculiar as it is strong and highly intermittent. The intermittency can be quantified by the parameter C_1 which corresponds to the fractal codimension (i.e. for a 1-D (time series), it is 1 minus the fractal dimension) characterizing the sparseness of volcanic “spikes” of mean amplitude. (see Lovejoy and Varotsos 2016).

The volcanic response appears to be non-linear as the intermittency (“spikiness”, sparseness of the spikes) parameter C_1 changes from about $C_{1,FV} \approx 0.16$ for the input volcanic forcing to $C_{1,T} \approx 0.03$ for the temperature response: the latter is therefore much less intermittent than the former although it is possible that the estimated C_1 changes slightly due to finite size effects and internal variability (Lovejoy and Varotsos 2016). When two processes have different intermittencies, they cannot be linearly related. However, by taking appropriate powers, they can be. Therefore to put the volcanic forcing into the linear forcing framework it must first be nonlinearly transformed to that it has the same intermittency as the temperature response.

Lovejoy S., Varotsos C. (2016) Scaling regimes and linear/nonlinear responses of last millennium climate to volcanic and solar forcings. *Earth Syst Dyn* 7:133–150. <https://doi.org/10.1002/wcc.397>

L222 and Eq. 13 This would serve as a good introduction to the whole description of the forcings instead of being hidden away at the end of the description of the volcanic forcing.

Author: Ok, thanks, will be moved to earlier on.

L230-234 This sentence is quite complicated, especially with the choice of punctuation. Consider splitting it. Should be “addresses” instead of “address”.

Author: Ok.

Sec. 2.3: A plot of the time series would be helpful at this point (historic datasets + CMIP5/6 MME). The discussion is needlessly difficult to follow without it.

Author: A figure will be introduced here of the historical datasets and CMIP5/6 MME.

L254 The internal stochastic forcing was not actually introduced in the formulas before so the "Recall that" seems misleading.

Author: Ok, thanks, this will be rewritten.

L264 Please expand on why low frequencies only weakly influence the likelihood function.

Author: The calibration of the FEBE is done on historical temperature observation residuals, calculated by removing the forced temperature response (anthropogenic and natural). The residuals contain mainly high frequency variability as most of the low frequency variability has been removed by the forced response (all forcings but volcanic contain primarily low frequency information).

L288-291 Please expand on how this five-dimensional parameter space was used (presumably to draw sets of parameters for an ensemble of simulations).

Author: The five-dimensional parameter space was indeed used to draw sets of parameters for an ensemble of simulations - this will be made explicit in the revision. It was the basis of the Monte Carlo technique used to determine the projection mean and uncertainty. Our goal was to estimate the ensemble average over an infinite ensemble, but Monte Carlo methods were used only as a convenient numerical technique.

Fig.3 (left) There is no discussion about why the PDFs for the NOAA dataset differ from all others (in particular GISTEMP, which uses the same base).

Author: The difference in PDFs for the NOAA and GISTEMP dataset is most prominent in the PDFs of the scaling exponent H and the volcanic intermittency correction exponent ν . The exact cause of the difference is not clear although it arises from the MLOST dataset's use of a complex frequency algorithm with low-frequency tuning. This low-frequency tuning along with the spatio-temporal smoothing applied in the MLOST dataset is likely the cause of a lower ν (i.e. a smoother volcanic forcing) along with a slightly higher H (i.e. a smoother temperature response).

Smith, Thomas M., Richard W. Reynolds, Thomas C. Peterson, and Jay Lawrimore. "Improvements to NOAA's Historical Merged Land–Ocean Surface Temperature Analysis (1880–2006)". *Journal of Climate* 21.10 (2008): 2283-2296. <https://doi.org/10.1175/2007JCLI2100.1>.

Sec 3.1.1 The prior for H is not mentioned.

Author: It was mentioned earlier in the text in Sec. 2.3; we will include the

prior here also.

L337 Here, the theoretical value is given as 0.32, but in L220 it was given as ≈ 0.3 .

Author: Thank you for pointing this out, we will make it consistent as $\nu \approx 0.3$.

L341 I would suggest starting a new section for the total forcing series here, instead of keeping the discussion of the full forcing series and Greens functions under the volcanic forcing header.

Author: Ok, good idea, thanks.

The size of the ensembles is not mentioned, nor how large they had to be such that the probability functions converge.

Author: For the FEBE, the spread of the forced projections is the spread of an infinite ensemble as it is only deterministic so the uncertainty comes from the parametric uncertainty. The probability distributions for the parameters are derived by calibrating FEBE using either the RCP or SSP historical forcing on the historical observed temperature series using a Bayesian scheme explained in Sec. 2.3.

L347 The notations G1 and G2 were not used before. Please make this consistent throughout the paper.

Author: Thank you for pointing this out, we will correct this so the notation used is consistently G_0, G_1, G_2 instead of G_δ, G_Θ .

L349-353 The wording "singular response" is unclear, as well as why it allows for the accurate reproduction of "the statistics of the internal variability" and makes FEBE "more sensitive to volcanic forcings. It seems more sensitive to everything up to ≈ 300 years. It is also not explained why this is a better representation of the historical datasets. It is unclear why 25 years in particular are mentioned when the gap remains similarly large beyond that.

Author: For $h < 1$, the impulse (Dirac) response Green's function is mathematically singular at the origin: for small t it behaves as t^{h-1} . This makes it particularly sensitive to high frequency forcing and this is mostly from volcanoes.

L354 Please give a reference for this as the standard ramp time. I am unsure whether the implied meaning of multi-year as anything starting from 2 was intended here.

Author: The transient climate response (TCR) has traditionally been defined as the change in global mean temperature at the time of a doubling of atmospheric CO2 concentration increasing at a rate of 1% per year, which takes 69 years, (approximately) 70 years (Collins et al. (2013)). The implied meaning of multi-year is the scale where high frequency variability unimportant - this will be rewritten in the revision.

Collins M, Knutti R, Arblaster J, Dufresne J-L, Fichet T, Friedlingstein P, Gao X, Gutowski WJ, Johns T, Krinner G, Shongwe M, Tebaldi C, Weaver AJ, Wehner M (2013) Long-term climate change: projections, commitments and irreversibility. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge

L355-356 Having the discussion of the projections before they have been shown might not be the best choice.

Author: The plot position will be moved.

end of Sec.3.2: Unclear whether G1 or G2 was used for the projections.

Author: Either can be used with respect to the linear framework shown in eq. 3 with G_0 - the impulse response function. In fact G_1 was used since it is smoother (at small time scales, it varies as t^h i.e. it is not singular at $t = 0$), and this makes it numerically attractive. This will be explicitly clarified in the revision in an appendix.

Recall eq. 3, putting $s = 1$:

$$T(t) = \int_0^t G_0(t-u)F(u)du,$$

at small t , $G_0(t) \approx t^{h-1}$, i.e. it is singular. To avoid the corresponding numerical issues, we instead first calculate the integral of T:

$$\zeta(t) = \int_0^t T(v)dv.$$

This is given by:

$$\zeta(t) = \int_0^t \left(\int_0^v G_0(v-u)F(u)du \right) dv = \int_0^t G_1(t-v)F(v)dv$$

Where we have used:

$$G_1(t) = \int_0^t G_0(v)dv.$$

At small t , $G_1(t) \sim t^h$ which is regular since $h > 0$. Thus using $\zeta(t)$ we determine $T(t)$ using finite differences. To calculate the forced response for the projections we use an efficient numerical convolution algorithm.

Fig. 6: a) The data in the inset (or the axes) does not seem to fit the data shown in the full plot. This is particularly notable with the IPCC model. The x-axis of the shown inset and the inset area marked in the full plot do not agree. b) (top) Even in the inset, small times are barely visible. This is made worse by the fact that the border of the box does not align with the y-axis. The labels for the insets are too small, those for the lower plot are barely legible. c) G_2 is not discussed in the text, please change this.

Author:

- a) The inset area marked on the full plot will be changed to agree with the x-axis range of the inset.
- b) Ok, this will be fixed.
- c) $G_2(t)$ will be included in a discussion in Sec. 2.1.

There is a lot of emphasis put on the results with FEBE being less uncertain (i.e. L372) than for CMIP, but this does not necessarily mean that they are a better representation of climate. In addition, I would like to know how much this is due to the CMIP ensemble consisting of different models versus the FEBE ensemble being produced with one model.

Author: In the case of GCMs, the uncertainty is termed "structural". The multi-model ensemble is treated as a single "super" process. Each GCM is considered to be a different stochastic realization of this super process. It may indeed be difficult to justify this assumption, but that is what the community has used. Questions surrounding the validity of this assumption help motivate and justify the development of other approaches such as the approach discussed here. In comparison, for the FEBE it is the more conventional parametric uncertainty, that is, uncertainty about the values that should be assigned to a climate model's parameters.

In both cases, the projections were deterministic but with uncertainty limits due to their respective model uncertainties and yielded an estimate of the forced component with qualitatively different uncertainty bounds.

L382 The median for λ in CMIP6 is not given.

Author: The median for λ (s) is $0.92 K(Wm^{-2})$ in CMIP6 - this will be updated in the revision, also, we have changed the notation, using the more standard symbol “ s ” for the climate sensitivity.

L393-399 Unclear how much of the higher RWF in FEBE is due to energy balance models usually responding faster than GCMs. A discussion on how reliable the FEBE estimate is in comparison to the CMIP one is missing.

Author: The Realized Warming Fraction (RWF) will be further explored in the following revision. The rapidity depends on the response function G , and model parameters τ, H . The reliability of the FEBE is discussed in the response to referee 1.

Sec. 3.3.1 and 3.3.2 Might be interesting to discuss why both ECS and TCR best estimates are smaller for the SSP than the RCP scenarios with FEBE, but it is the opposite in CMIP.

Author: The ECS and TCR estimates using the SSP scenarios with the FEBE are lower than those using RCP due to the overly strong aerosols over the historical period in the SSPs which require a lower aerosol linear factor along with lower ECS to best match the historical temperature record. The difference between the shape of the RCP and SSP aerosol forcing can also account for this. We will expand upon this in the revision.

Fig. 8: The figure would be easier to read if the single plots were aligned with each other according to their axes and were shown with the same font size.

Author: Ok.

L417 The CMIP5 MME is mostly warmer for the given period, but not always (compare i.e. the early 1940s where observed temperatures are higher than the mean). Why the upper limit of 1960 was chosen is unclear, since the CMIP simulations remain warmer most of the time afterwards (and in particular, after 2000).

Author: This statement will be changed to entail the whole historic temperature sets in the revision.

L419-422 It is not discussed whether FEBE tracking the hiatus is a sign of FEBE’s skill or a result of the fact that the parameters were chosen based on the same observations that include this hiatus. Since EBMs usually have trouble with this hiatus, it would be interesting to

discuss whether there exist other conceptual models that reproduce the hiatus or whether this is a standout feature of FEBE.

Author: One of the reasons why the FEBE tracks the hiatus well is due to the fact that the parameters are calibrated over the historical temperature observations. We show that the FEBE tracks the entire period since 1880 well, not only the hiatus. We clarify this in the revision.

L421 There is no lower plot in fig. 9.

Author: The main plot (rather than the inset) was meant, this will be corrected.

Fig. 9: The observations show strong interannual variability, a discussion on whether a temporally higher resolved FEBE could capture such short-term variability (or even whether a higher resolution in time is feasible with this approach). Equally interesting would be to discuss whether other response functions could be designed to capture variability on arbitrary timescales. A bit unclear whether FEBE is useful for discussions of variability at all or should be applied only to discussions of the mean.

Author: In Lovejoy et. al. (2021) and Lovejoy (2021), the FEBE is shown to plausibly reproduce the annual cycle at monthly resolution, in particular to explain the lag between the temperature maximum and the maximum in the radiative forcing. The small scale FEBE limit is not known, although it is likely to be 1 month. Justification comes from the success of the high frequency FEBE limit that successfully forecasts monthly and seasonal temperatures (see Del Rio Amador et al. (2019)). The variance of $\gamma(t)$ is the amplitude of the internal forcing assumed to be a Gaussian white noise is estimated in the response to referee 1 using Lovejoy et al. (2021). The following plot (fig. 1) shows the mean observational temperatures series along with one realization of the FEBE (with optimal parameters based on F_{AerRCP}) using both the internal and external forced variability. This work needs more work to be fully convincing, but it is promising at the moment.

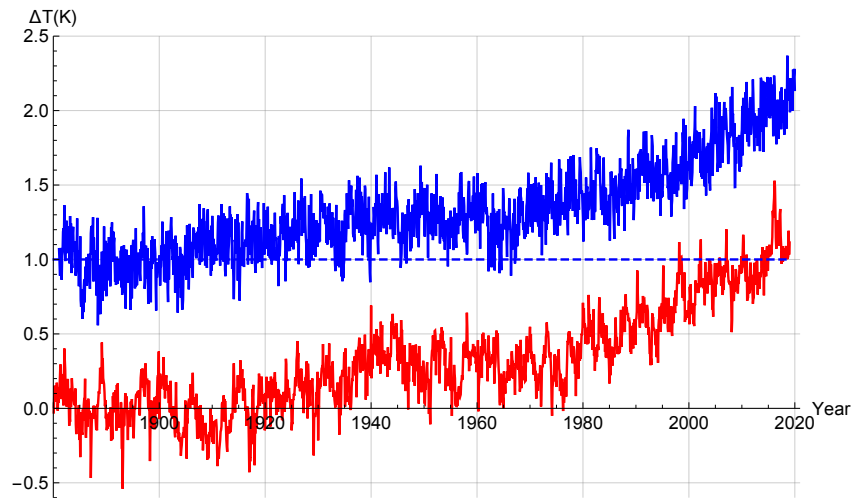


Fig. 1: The mean observational temperature series (red) shown alongside a single realization of the FEBE with internal and externally forced variability (blue, shifted up 1 for clarity.)

Del Rio Amador, L., and Lovejoy, S., Predicting the global temperature with the Stochastic Seasonal to Interannual Prediction System (StocSIPS) Clim. Dyn. doi: [org/10.1007/s00382-019-04791-4](https://doi.org/10.1007/s00382-019-04791-4), 2019.

Lovejoy, S., Procyk, R., Hébert, R. and del Rio Amador, L. (2021), The Fractional Energy Balance Equation. Q J R Meteorol Soc. <https://doi.org/10.1002/qj.4005>

Lovejoy, S.: The Half-order Energy Balance Equation, Part 1: The homogeneous HEBE and long memories, Earth Syst. Dynam. Discuss., <https://doi.org/10.5194/esd-2020-12>, 2021.

L461-462 According to the section on aerosol forcings (fig. 1), there is a future aerosol forcing in the RCP and SSP scenarios. It is unclear whether here the removal of the volcanic forcing is meant. Since the forcings in FEBE and SSP should be consistent (with the exception to the dampening parameters for the aerosols mentioned in the first part of the sentence) unclear why this would produce the temperature difference.

Author: Thank you for pointing out how this line is unclear - the higher ECS found in CMIP6 models is related to compensating for the too strong aerosols in the historical part, thus when the concentration of aerosols begins to drop to zero (as in the SSP scenarios) the CMIP6 models heat much quicker than the FEBE. This will be better clarified in the revision.

The effect of the lack of volcanic forcing in the future projections is not discussed, nor whether (and if so, how) the authors suggest its inclusion in future studies.

Author: Future volcanism was not prescribed by the CMIP experiments so it was set to null in the future. Since we may be confident that volcanoes will continue to erupt, a mean rate of volcanic cooling could be prescribed set to the time average of past volcanism assuming that future volcanism is on average the same as the past. It could be interesting to include stochastic volcanic forcing in the future for ensemble realizations of projections, in a future paper.

Fig. 10: The sine-like oscillations in the future projections by FEBE (esp. RCP2.6 and RCP4.5) should be mentioned in the text. I assume they reflect the solar forcing cycle, but remain uncertain whether the different magnitude of this cycle in the scenarios is due to a difference in overall magnitudes of forcings between scenarios.

Author: The sine-like oscillations are from the solar forcing cycle which prescribed the reproduction of the last 11-year cycle for the future - their magnitude is the same for each scenario.

Fig. 11: It is not discussed why these FEBE projections are notably smoother than the ones in fig. 9 and whether this is only a reflection of the difference in forcing.

Author: These projections are smoother due to the scale shown in the plots - this will be addressed in the revision.

A discussion about what could lead FEBE to overestimate future warming would be interesting and in particular whether it can be argued based on the projections in Fig. 10 11 that FEBE provides a baseline of minimal warming.

Author: The FEBE constrains the strength of historic aerosol cooling to best match the observed temperature record, this results in less aerosol cooling than that modelled in the CMIP experiments. So if a large future aerosol forcing is introduced the resulting warming may be underestimated by FEBE due to the reduced strength of the aerosol forcing in the model compared to the CMIP models.

In general if the FEBE underestimates the memory of the system to radiative forcing, then it may underestimate the warming, but would at least be a robust lower bound. The converse could be said about overestimating the memory, and thus overestimating the warming as the forcing decreases, but this looks less likely.

L478-479 It seems to me that this phrasing is slightly misleading. It reads as if extremely likely refers to the 15 years. However, it seems like 15 years later is just the point when it becomes extremely likely that the 1.5o threshold is exceeded.

Author: Thank you for noticing this - it is indeed poor phrasing, it will be fixed in the revision.

Fig.12 bottom: It is not mentioned that the solid RCP scenario is at 0. Would be good to change the plot so that this becomes visible. Would also help to mark the "extremely likely" threshold at 0.95 discussed in the text.

Author: It will be mentioned that the solid RCP scenario is at 0 and reflected in an improved plot.

Fig.13 top: It is not discussed why the probability for the SSP126 scenario decreases so strongly after 2070 in the FEBE ensemble.

Author: The probability for SSP126 scenario (blue, dotted) decreasing so strongly after 2070 was partially a visual artefact, this has been corrected in the updated fig.13. The probability still decreases strongly, but with a probability higher than before, this decrease is due to the combination of a lower FEBE ECS with regards to the CMIP6 MME and the strong decrease of CO_2 emissions in the scenario. We have also included the CMIP6 MME for scenario SSP245 in the revision (dashed lines).

L482-487: This seems more like a motivation suitable for the introduction than a conclusion.

Author: These lines are to reiterate why complementary models should be developed, and to explain the motivation for the FEBE.

It is not discussed for what kind of studies the authors suggest the usage of FEBE, nor what the gaps are in studies with more complex models that FEBE might fill.

Author: EBMs are useful as they are able to produce individual climate projections with reduced computational resources. The FEBE is based on the real climate not model climates, it is therefore complementary to the GCMs. An extended discussion on the benefits and uses of this will be included in the revision.

L568 The authors say that their goal is to improve future projections, but do not expand on how they think that will be possible with FEBE.

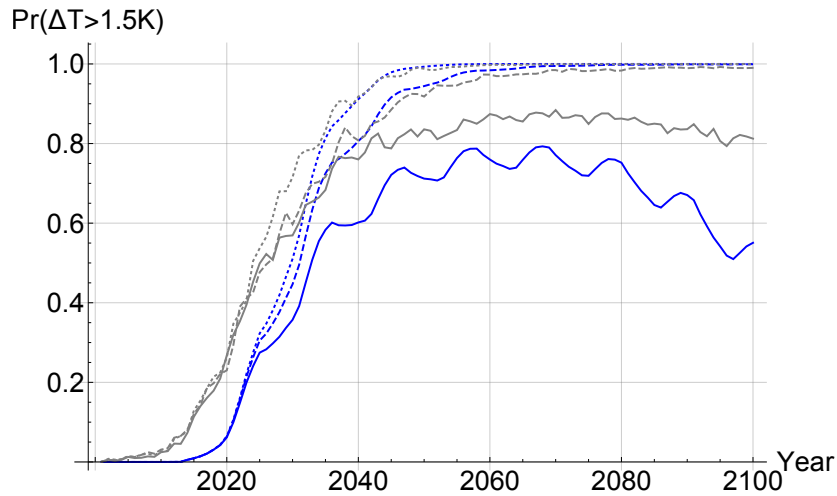


Fig. 13: The probability for the global mean surface temperature of exceeding a 1.5K threshold (top), and a 2K (bottom) are given as a function of years for the FEBE, using $F_{Aer_{RCP}}$ (blue) and for the CMIP6 MME (black). The three SSP scenarios are considered for each case: SSP 126 (solid), SSP 245 (dashed), and SSP 585 (circles).

Author: With tighter constraints on ECS and TCR from the FEBE we can better estimate future warming when bringing together multiple lines of evidence such as that done is Sherwood et al. (2020). The FEBE once expanded spatially provides a flexible framework which can be calibrated directly on observations, providing a direct representation of forcing to response relationships. We can also calibrate the FEBE on the historical runs of the CMIP models, updating our parameter estimates, allowing for GCM-FEBE hybrid projections.

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, 58, e2019RG000678. <https://doi.org/10.1029/2019RG000678>

L569 It is unclear how FEBE can be used to understand the generational differences between CMIP models better.

Author: We plan on calibrating the FEBE on the historical runs of the CMIP models in order to perform a feedback analysis to investigate the differences between how models treat their volcanic and aerosol forcings through the parameters ν and α in a future work.

Neither the model code nor the simulations with the FEBE are

included in a code/data availability section and made available to the reader.

Author: The forcing, and temperature data along with realizations of the FEBE will be included in an open source format, and the code of the model will be made available upon publication, but can be made directly available to the reviewer.

citation style hinders readability (brackets even when the citation is part of the sentence and double brackets when listing several references)

Author: Ok, this will be cleaned up.

The whole text needs to be checked for correct comma placement, especially with respect to introductory phrases and interrupters. The punctuation could also be improved with respect to colon and semi-colon usage.

Author: A thorough proof reading will be done to improve the grammar and readability.

in this state of the manuscript it did not seem to make sense to give detailed feedback on specific grammatical errors, so these are not included in this review

Author: The grammar will be improved.

the way "FEBE" is used in the text is inconsistent, sometimes it is "the FEBE", at other times just "FEBE" without an article

Author: "The FEBE" will be made consistent throughout the text.

inconsistent usage of "figure X" or "fig. X" in the text

Author: We will revise using ESD's publishing style:

- The abbreviation "Fig." should be used when it appears in running text and should be followed by a number unless it comes at the beginning of a sentence, e.g.: "The results are depicted in Fig. 5. Figure 9 reveals that..."

There are quite a lot of unnecessarily run long sentences that hamper readability.

Author: These will be shortened or split into two to improve readability.

In general, the plots could be cleaner, i.e. for multi-panel plots the plot sizes as well as font sizes differ.

Author: Thank you for pointing this out, they will be made cleaner.

The linkage between sentences, paragraphs and sections could be improved to make the reader's life easier. Also the internal structure of sections, paragraphs and even sentences could be more coherent. For example, it's not always easy to identify the topic and stress in a sentence. This could be facilitated by an improved structure of sentences. Sometimes, the connection between subject and verb gets lost due to a sentence structures that is too complicated. In general, reducing long sentences and technical language would improve the reader's understanding.

Author: Ok.

L20 MME used without introducing the abbreviation

Author: Multi-Model Ensemble will be used first before introducing the abbreviation.

L21 should be "the FEBE projections were...", also comma missing after the introductory phrase

Author: Ok, corrected.

L40 different dashes are used at the beginning and end of the insertion, "that" after the insertion should be discarded (whole sentence could benefit from being re- formulated for better readability, though)

Author: Ok.

L63 MME as an abbreviation is only introduced in L94

Author: Ok.

L83 should be Hebert et al.'s truncated power law or at least it has not been introduced so far as having been developed by only one person

Author: Ok.

L89 article missing in "In the methods and materials section"

Author: This will be corrected, thanks.

L138 should have a reference to where that discussion happens

Author: Ok.

L165 "While RCP8.5 and SSP585..." subject of the sentence is plural, rest of sentence is singular

Author: Ok.

L171 "and" between "Kyoto protocol" and "ozone depleting substances" would make this sentence easier to read

Author: Ok.

L197 should be "two 11-year solar cycles"

Author: Ok.

L327 the last part of the sentence lacks a verb

Author: Ok.

L229 should be "find in/for both cases" or similar

Author: Ok.

L348 no apostrophe for " its' "

Author: Ok.

L349 "where" or similar instead of "whereas"

Author: Ok.

L362 "CO2 levels were" or "CO2 was increased"

Author: Ok.

L415 rogue bracket at the end of the sentence

Author: Thank you for noticing this.

L416 either "Between 1915-1960" or "In the 1915-1960 period" or similar

Author: Ok.

Fig.9 There are white artefacts in the grid lines, box borders, and axes of the plot as well as in the upper limit of the CMIP5 MME. It is unclear why the x-axis of the inset unnecessarily includes negative values. The plot would appear cleaner if the borders of the inset were aligned with the axes. CMIP5 MME is in gray not black (same in later plots). Referring to the inset as the top is confusing. The left border of the inset in the whole plot is placed too late.

Author: Ok.

L431 scratch "In comparison" at the beginning of the sentence (at this point it is unclear with respect to what the comparison is supposed to be and it does not match the remainder of the sentence)

Author: Ok.

L448 forcings should be plural in "the other forcing are practically"

Author: Ok.

L454 doubled "the"

Author: Ok.

L461 drop "by" in "are by nearly 65% warmer"

Author: Ok.

L547 "to" missing in "purely due differences"

Author: Ok.

L556 I would suggest replacing "be passed" with "happen" in the context of this sentence

Author: Ok.

Held et al. citation not in correct place in alphabetical order

Author: This will be corrected, thanks.

Interactive comment on “The Fractional Energy
Balance Equation for Climate projections
through 2100” by Roman Procyk et. al.

Roman Procyk et. al.

The manuscript provides, at times, incomplete details on how the results were obtained and how the approach was designed, making reproducibility of the results difficult. I would suggest sufficient details be added. This would also help make potential errors in the approach more detectable.

Author: A revised discussion will be given to how the results were obtained in the revision. The forcing, and temperature data will be included in an open source format, and the code of the model will be made available in the future. Extra information about what is to be included was given in the response to reviewer 1 and 2.

I would also like to see some discussion on approximating the joint prior probability density functions with a multivariate Gaussian process, including what motivated this decision, how accurate this approximation is and if this could have any repercussions for the final results. How the covariance between the parameters used in the joint prior distribution is determined should also be included.

Author: The results shown in Figures 3,4 and 7 are the marginal distributions for each of our five parameters - their Gaussianity was checked and thus we were motivated to approximate the joint posterior probability density functions with a multivariate Gaussian for our future projections otherwise this process can be computationally expensive (prior experience of a coauthors). The way it is written in the paper, we understand the confusion - the prior probability functions are not approximated by a multivariate Gaussian but the posterior probability was with the correlations between all parameters taken into account, this will be clarified further in the revision.

Physical units and chemical formulas should not be typeset in italics.

Author: Ok, this will be corrected.

Ranges should be denoted using en dashes (–), instead of hyphens (-), e.g. 1998–2015.

Author: Ok, this will be corrected.

Top panel in figure 6: The horizontal axis of the smaller window ranges from 0 to 25 years, but the dotted lines suggest it is taken from the range of 0 to 50 years from the larger figure. Please correct/clarify.

Author: The inset is for 0 to 25 years, this will be cleaned up in a revised figure 6 so that the dotted lines reflect this.

Line 274: The likelihood function is said to be a posterior probability. This is incorrect, as the likelihood is a distinct probability distribution. Also in line 274: the terms “posterior” and “a priori” seem to have been switched.

Author: This is correct, a change will be made in the revision.

Being in the Bayesian modeling framework, where parameters are treated as stochastic variables instead of fixed and true values, I think it would be more appropriate to use the term “credible intervals” instead of “confidence intervals”. The former refers to an interval within which an unobserved (stochastic) parameter value falls with a particular probability, whilst the latter is an interval that we are, to a certain degree, confident include the true (deterministic) parameter value.

Author: Thank you for pointing out this useful distinction, it will be changed in the revision.