

Review of: The Fractional Energy Balance Equation for Climate projections through 2100

1 Summary of the paper

The paper presents climate projections with a zero-dimensional energy balance equation and compares the resulting time series, equilibrium climate sensitivity, and transient climate response to CMIP5 and CMIP6 projections. The model parameters for the projections are determined using a Bayesian approach and historical observations. The projections are lower than the CMIP5/6 ensemble means, but mostly within their confidence intervals.

2 Major comments

With its focus on how a simple modeling approach can reproduce features from simulations by GCMs, the manuscript is tackling an interesting topic. However, in its present form major revisions are required to bring this manuscript into a form that is suitable for publication.

In particular:

- 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.
- 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.
- The reader does not get an idea of what the paper adds to Hébert 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.

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

3 Minor comments

Abstract

- L12 "We found that aerosols..." should mention that here aerosols refers to the CMIP aerosol forcings
- L22 It is unclear how FEBE complements (in addition to support) GCMs based on what is said.
- The abstract summarized well what is done in the paper, but fails to address limitations and the gap in literature it aims to fill.

Introduction

- L35 reference for the projection uncertainty in response to CO₂ doubling missing
- L43 should be motivated why it is desirable to construct models beyond linearity and stochasticity
- 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.
- L118 Please mention why causality is not respected otherwise.
- 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.
- L130 a short explanation of what fractional Relaxation noise is would be helpful

Methods and Material

- The implementation of the model is not described at all.

- It should be mentioned explicitly that the model produces annual data (same for the temporal resolution of the forcing).
- 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 Hébert 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".
- L145 It is unclear why 277ppm is the reference.
- L177 A short elaboration as to why a linear scaling factor is appropriate to describe the uncertainty would be nice.
- Section 2.2.3 For RCO scenarios it is stated that an aerosol forcing smaller than -1Wm^{-2} is implausible, but not why this is not also the case for the SSP forcing.
- 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.
- 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.
- L205 The point of comparison in "comparable effect" is missing.
- 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)
- L218 Would be helpful if the authors would expand on C_1 being the codimension of the mean and what the implications of this are.

- 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.
- L230-234 This sentence is quite complicated, especially with the choice of punctuation. Consider splitting it. Should be "addresses" instead of "address".
- 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.
- L254 The internal stochastic forcing was not actually introduced in the formulas before so the "Recall that" seems misleading.
- L264 Please expand on why low frequencies only weakly influence the likelihood function.
- L288-291 Please expand on how this five-dimensional parameter space was used (presumably to draw sets of parameters for an ensemble of simulations).

Results

- 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).
- Sec 3.1.1 The prior for H is not mentioned.
- L337 Here, the theoretical value is given as 0.32, but in L220 it was given as ≈ 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.
- The size of the ensembles is not mentioned, nor how large they had to be such that the probability functions converge.
- L347 The notations G_1 and G_2 were not used before. Please make this consistent throughout the paper.
- 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.

- 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.
- L355-356 Having the discussion of the projections before they have been shown might not be the best choice.
- end of Sec.3.2: Unclear whether G_1 or G_2 was used for the projections.
- 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.
- 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.
- L382 The median for λ in CMIP6 is not given.
- 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.
- 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.
- 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.

Projections

- 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).

- L419-422 It is not discussed whether FEBE tracking the hiatus is a sign of FEBE's skill or a results 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.
- L421 There is no lower plot in fig. 9.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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 1.5° threshold is exceeded.
- 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.

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

Conclusions

- L482-487: This seems more like a motivation suitable for the introduction than a conclusion.
- 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.
- 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.
- L569 It is unclear how FEBE can be used to understand the generational differences between CMIP models better.

Code and data availability

- Neither the model code nor the simulations with the FEBE are included in a code/data availability section and made available to the reader.

4 Technical corrections

General comments

- citation style hinders readability (brackets even when the citation is part of the sentence and double brackets when listing several references)
- 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 semicolon usage.
- 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
- the way "FEBE" is used in the text is inconsistent, sometimes it is "the FEBE", at other times just "FEBE" without an article
- inconsistent usage of "figure X" or "fig. X" in the text
- There are quite a lot of unnecessarily run long sentences that hamper readability.
- In general, the plots could be cleaner, i.e. for multi-panel plots the plot sizes as well as font sizes differ.

- 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.

Abstract

- L20 MME used without introducing the abbreviation
- L21 should be "the FEBE projections were...", also comma missing after the introductory phrase

Introduction

- 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 reformulated for better readability, though)
- L63 MME as an abbreviation is only introduced in L94
- L83 should be Hébert et al.'s truncated power law or at least it has not been introduced so far as having been developed by only one person
- L89 article missing in "In the methods and materials section"

Methods and Material

- L138 should have a reference to where that discussion happens
- L165 "While RCP8.5 and SSP585..." subject of the sentence is plural, rest of sentence is singular
- L171 "and" between "Kyoto protocol" and "ozone depleting substances" would make this sentence easier to read
- L197 should be "two 11-year solar cycles"

Results

- L327 the last part of the sentence lacks a verb
- L229 should be "find in/for both cases" or similar

- L348 no apostrophe for " its' "
- L349 "where" or similar instead of "whereas"
- L362 "CO₂ levels were" or "CO₂ was increased"

Projections

- L415 rogue bracket at the end of the sentence
- L416 either "Between 1915-1960" or "In the 1915-1960 period" or similar
- 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.
- 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)
- L448 forcings should be plural in "the other forcing are practically"
- L454 doubled "the"
- L461 drop "by" in "are by nearly 65% warmer"

Conclusions

- L547 "to" missing in "purely due differences"
- L556 I would suggest replacing "be passed" with "happen" in the context of this sentence

References

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

References

Hébert, R., Lovejoy, S., and Tremblay, B.: An observation-based scaling model for climate sensitivity estimates and global projections to 2100, *Climate Dynamics*, <https://doi.org/10.1007/s00382-020-05521-x>, URL <https://doi.org/10.1007/s00382-020-05521-x>, 2020.

Schmidt, G. A., Jungclaus, J. H., Ammann, C. M., Bard, E., Braconnot, P., Crowley, T. J., Delaygue, G., Joos, F., Krivova, N. A., Muscheler, R., Otto-Bliesner, B. L., Pongratz, J., Shindell, D. T., Solanki, S. K., Steinhilber, F., and Vieira, L. E. A.: Climate forcing reconstructions for use in PMIP simulations of the last millennium (v1.0), *Geosci. Model Dev.*, 4, 33–45, <https://doi.org/10.5194/gmd-4-33-2011>, 2011.