Review of "Assessing the performance of climate change simulation results from BESM-OA2.5 in comparison to a CMIP5 model ensemble" by V.B. Capristrano et al. (revised manuscript)

Overall assessment and recommendation

I regret to conclude that this paper has not been sufficiently improved by the revision process to be acceptable. While I appreciate that the authors have tried to bring the characteristics of BESM-OA2.5 more in focus of their presentation (rather than discussing the general performance of CMIP5 models), the results is still a clumsy and partly dis-organized concatenation of results and result comparisons that do not lead to a clear assessment of the suitability of BESM for specific purposes. New text often has been insufficiently harmonized with the previous text, making reading through the manuscript still an extremely arduous task.

As I stated in my original review, my impression is that BESM is a reasonable model that could be useful for specific applications at least. Hence, I am reluctant to reject this paper once and for all. The authors should be allowed to make one more attempt to create a straightforward paper with a coherent message. To this end (as I have proposed before) the focus of future use of BESM should be made clear, considering the merits and shortcomings of this model. The authors should intensify their attempts to interlink the evidence arising from individual parameter evaluation. This already has been tried in a number of cases, but it too often results in circular reasoning, not approaching the roots of characteristic BESM features. Finally, I emphasize that just executing through my list of technical and language suggestions alone will not do! The author team apparently does not include an English native speaker, hence assistance in producing a proper English text ought to be given by either the editorial office or from some other consultant. Otherwise, I fear that I will be reluctant to read through this paper once again.

General remarks

1) Section 2.1 still contains elements of a comparison between BESM-OA2.5 and BESM-OA2.3 (e.g., p. 3, l.32) though a dedicated section (2.2) is supposed to cover such differences.
2) It is on occasions still hard to reproduce what has actually been done and why (e.g., p. 6, l.25).
3) No reason is given on p. 8, 2nd paragraph, why only 11 rather than 15 CMIP5 models are included here. Or are sometimes 11, sometime 15, models used, as could by read out of p. 8, l. 6?
4) Occasionally, I still miss a comment on the specific performance of BESM, even if it's well consistent with the CMIP5 ensemble (e.g. Figure 3).
5) Page 9, 1st paragraph: This has been reformulated, but is now even more confusing than before. Please reconsider, what is the intended message

here, with focus on BESM. Then stick to specific reasoning to underpin that message.

6) Page 9, 2nd paragraph: Here, too, the line of reasoning remains badly organized: What is the message: Does BESM simulate a stronger Arctic amplification than the CMIP ensemble (suggested by the more negative Planck feedback)? This could simply explain more snow/ice melting. Evidently the lapse rate feedback in BESM is exceptionally positive at Arctic latitudes, pointing at a enhanced vertical gradient in the temperature response. Can this be discussed in the context of the Veiga et al. paper (atmospheric temperature response)?
7) The last paragraph of section 4.2, with much newly introduced text, is very hard to understand both concerning the weak use of English language and a confusing inherent logic. I have read through this paragraph three times, but then gave up, being unable to reconcile the statements in the text with what the figures display.
8) Scatter Diagrams in Figures 8 and 9: Do you conclude anything from the apparent correlation between precipitation in piControl and abrupt4xCO2 on one side, and missing correlation for respective surface temperature levels on the other side? Does this have implications for the BESM model performance.
9) In the last paragraph of the conclusions an outlook to what is planned with BESM-OA2.5 (future research focus) is still lacking. However, this would be the logical outcome of the assessment of its merits and shortcomings, which I assume is what the present paper has been written for.

<u>Language and Technical Remarks</u>

p. 1, l. 7 (Abstract): " … the CMIP5 ensemble mean value …"

p. 1, l. 8 (Abstract): " … BESM simulation show zonally average feedbacks, estimated from radiative kernels, that lie within the ensemble standard deviation …"

p. 1, l. 11 (Abstract): "… BESM also features a strong positive …"

p. 1, l. 12 (Abstract): As this sentence mentions a merit of BESM, while the preceding sentence comments on a disagreement with CMIP, "moreover" makes quite an unlucky connection. By the way, "consistent" with what?

p. 2, l. 7: "… results in a temperature rise …"

p. 3, l. 7: "… models, also discussing peculiarities in the BESM climate response."

p. 3, l. 16: "… same as used by Veiga …"

p. 3, l. 17: "… model, with its dynamical core being based on …"

p. 3, l. 22: "… of physical parameterizations between BAM (as used in this paper) and BAM NWP …"

p. 3, l. 24: "… 28 layers, unevenly spaced, in the …"

p. 3, l. 29: "… is able to capture …"

p. 3, l. 30: "… with a double ITCZ …"

p. 3, l. 31: "improvement", despite of the "substantial biases" addressed in the preceding sentence?

p. 3, l. 33: "… decadal climate variability patterns." This is meant, isn't it?

p. 4, l. 4: I understand that AMOC is a circulation structure rather than a parameter. So, what "value" a you referring to? If required, please give an absolute or relative difference of the parameter you have in mind.

p. 4, l. 9: "… are determined, which are important …". Anyway, the content of this sentence to me resembles what is given below (p. 4, l. 14), with the sentences in between (starting with "The total energy balance …") causing an awkward logical break.

p. 4, l 19: This sentence again repeats what is given in p. 4, l. 9 …

p. 5, l. 6: "… which means a spin-up of 150 years." Does this mean that the 150 yrs of abrupt4xCO2 are regarded as a spin-up here (due to their non-equilibrium character)? Or are 150 yrs of abrupt4xCO2 swapped as a spin-up, and *another* 150 yrs evaluated as some kind of quasi-equilibrium? Please, clarify.

p. 5, l. 6: "… commonly employed … for climate change assessment"; please, be careful to distinguish between "climate change assessment" and "climate sensitivity assessment"! In my view, "climate change" in the CMIP context is rather assessed through historical simulations and future scenario simulations.

p. 5, l. 12: In this paragraph the "forth and back" jumping in addressing the merits of the regression and kernel method is somewhat confusing but could be easily avoided.

p. 5, l 27, 28: There's still something wrong with the sentences here. Suggestion: "As G can be approximated by backward regression towards $\Delta T_{as}=0$, ECS can be estimated as $ECS=-G/\lambda$."

p. 5, l 30: "… it is common to divide the result derived from 4xCO2 simulations by 2 (Andrews …"

p. 6, l. 9: "… is used next, in order to partition the …", as "next to" is confusing. By the way, "separate" or "split" may be preferred to "partition".

p. 6, l. 14: "integraly" -> "fully" (or "necessarily")

p. 6, l. 16: "This, however, assumes that …"

p. 6, equation 3: T_{as} is the near surface temperature (p. 5), but what is then T_s? I tried to clarify this by looking into Vial et al. (2013), without success. Please, be precise in citing, or explaining what you have done, and why.

p. 6, l. 25: Confusing: As q is in the data base, why should it be approximated based on the assumption of constant relative humidity? To my knowledge, this is not common in feedback analysis. Is it possible that you are mis-interpreting the cited references here?

p. 7, l. 5: "… changes are not accounted …"

p. 7, equation 5: It is not immediately obvious, what the indices "a" and "k" mean.

p. 8, l. 4: "… were assessed as it was performed …"; I do not understand this sentence. Are the data not from the ESGF data base (p. 5, l. 10) ??

p. 8, l. 6: "… e inmcm4 ", did you intend "… and inmcm4"?

p. 8, l. 13: witch -> which

p. 8, l. 29: Please, explain how Figure 4 is related to Figure 3. Is it simply an average over the latitudinal profile of Fig. 3? Your discussion of the Planck feedback is casting doubts concerning this: If it's constant by about -4 $Wm^{-2}K^{-1}$ (l.29) with mostly negative deviations at polar latitudes, how can this result in a global mean of -3.6 $Wm^{-2}K^{-1}$ (l. 26)? Please, cross-check the numbers.

p. 8, l. 32: "stronger vertically homogeneous warming". This is a strange reasoning, as the Planck feedback is essentially the surface warming, constantly extrapolated upward through the depth of the troposphere. Can the message of this sentence be reconciled with Figure 8?

p. 9, l. 20: The partly revised text in this paragraph (see also major comments) contains some sensible elements, but is also moving in circles, explaining stronger sea-ice melting with stronger surface warming and vice versa. More reorganisation of the text is necessary.

p. 9, l. 22: "Those negative values …", it is unclear which values are addressed.

p. 9, l. 30: "The highest positive values …", I would expect that backscattering increases if ice turns into water, driving the shortwave cloud feedback to *more negative* values. However, your later discussion (Figure 7, see also below) seems to suggest that the longwave cloud feedback is the dominant component.

p. 9, l. 31: I feel that the following text (until "… outlier for the cloud feedbacks.") is mainly repetitive.

p. 10, l. 4: "λ_a, λ_{ac}", are you referring to an analytical framework that is given in Cess et al. (1989)? Otherwise the reader is rather left in the dark here.

p. 11, l. 4: "models with … apparently do not show …"; please also replace "present" by "show" on many occasions thereafter.

p. 11, l. 24: "…quadrupling of atmospheric CO_2 with the piControl pre-industrial CO_2 concentrations …": meaning what? The two first sentences of this paragraph appear to transport the same statement.

p. 11, l. 28: "… precipitation increase is not governed …"

p. 11, l. 31: Does this have in any way implications for the use of these somewhat "outlying" models?

p. 12, l. 13: "…regions with the strongest increase of westerly winds at all levels indicate a southward jet displacement …"

p. 12, l.18: Is "omega" something different from "vertical velocity"? Anyway, "omega" isn't self-explaining, so please adjust the text.

p. 12, l. 31: "… radiative code transference …", do you mean "performance"? Is there any indication of that particular feature for BESM's radiative transfer model?

p. 12, l. 31: “… rapid adjustments …”; the rapid adjustment process is included in the CMIP5 model results as well, per construction. You apparently did not calculate the rapid adjustments for BESM, but do you have any indications that there might be a systematic bias with respect to CMIP (see Smith et al., 2018).

p. 13, l. 4: “Two regions indicate enhanced inter-model standard deviation for Planck, lapse-rate and albedo feedback”; also in the rest of this paragraph the use of English language is very weak, making the meaning nearly incomprehensible for me.

References (only if not already cited in the paper):

Smith, C.J. et al., 2018: Understanding rapid adjustments to diverse forcing agents, Geoph. Res. Lett. 45, 12023-12031.