Review of “Overview of climate change in the BESM-OA2.5 climate model” by V.B. Capristrano et al.

Overall assessment and recommendation

This paper compares a selected set of results from a reference (PI-Control) and a climate change simulation (abrupt4xCO2) from the newly presented BESM-OA2.5 model with respective results from other models as derived through analysis of the CMIP5 data base. Generally, BESM-OA2.5 appears to perform reasonably in the CMIP5 context, so the paper could provide a good reference for further dedicated research with this model. However, the paper has a number of severe structural deficits that *need to be overcome by a thorough revision*. Beyond, there are quite a number of baffling statements (or unlucky formulations, to put it more mildly) suggesting that, beyond re-structuring, also some kind of re-thinking may be necessary, in order to yield a more insightful presentation of BESM-OA2.5’s merits and shortcomings.

General remarks

As already observed by another reviewer, the paper is severely out of balance in that it dwells too much on discussing (and interpreting) CMIP5 model results, while entering too less into the potential origin of BESM-OA2.5 peculiarities. To deepen existing CMIP5 results, in case of need, for an optimal assessment of BESM-OA2.5 is not necessarily beyond the scope of GMD (as suggested by referee #1), but the focus needs to be on the BESM results an their proper appraisal.

In the current text I find the statements in the last paragraph (p. 12, l. 11ff.) rather strange. The main objective of BESM is not supposed to “show climate sensitivity and thermodynamical responses similar to … CMIP5” but rather “to study the climate system [with a model able] to reproduce changes that are physically understood”. Besides, that the latter objective should be pursued by any climate model activity, what does this mean for the present paper and its priorities? Is BESM-OA2.5 in fact planned to be applied for dedicated research questions? Is BESM rather intended to be employed with higher, process resolving resolution? Or is it to be developed towards an Earth System Model with high comprehensiveness? Or is it to be optimized as a testbed for different physical (e.g. cloud) parameterizations? The authors were well-advised to decide for their (future) scientific focus first, and then select the proper diagnostics for their CMIP5 comparison accordingly. Else, the reader will remain wondering why some parameter or process is evaluated here, while another one is not. Moreover, the current use and interpretation of diagnostics is made rather schematically, reflecting too little on the origin of specific BESM features.

Finally, I notice that some basic climatological features of BESM-OA2.5 have already been documented in a current GMD manuscript (Veiga et al., GMDD 2018), of which I am not a reviewer. Comments to Veiga et al. look promising, but I strongly recommend to the editors that the present work should only be accepted together with its companion paper.

Other major issues

The authors use (or rather combine) two ways of calculating radiative feedbacks, viz. the regression method from Gregory et al. (2004) and the individual feedback calculation method from radiative kernels. This is quite recommendable, on principle. However, the methods are not equivalent as the phrase "seemingly redundant" (p. 4; l. 27) is suggesting. The finer points of the methodical difference are not addressed properly in the paper. The kernel methods includes, if no specific measures are taken (as in Vial et al., 2013, or in Chung and Soden, 2015), rapid adjustments directly induced from the CO_2 forcing.

More severe, the regression method implies that the radiative feedbacks are consistent with the actual radiative transfer module used in the climate model, while this is not true for the kernel method, if another than the radiative kernel from the actual climate model is used (as is the case here). The authors are apparently aware of this fact (p. 5, l. 25, p. 7, l. 16), but repeatedly fail to appreciate it when interpreting results.

In the same context the authors might also consider to refer to Forster et al. (2016) and Smith et al. (2018) here (beyond Vial et al., 2013), with respect to the options of calculating and interpreting effective radiative forcings, radiative adjustments and feedbacks, and climate sensitivity parameters. Are there abrupt4xCO2 simulations with fixed SST from BESM-OA2.5 that could be included in the discussion? Or are those intended to be analyzed in further BESM studies?

Further, I have some concerns about deriving the ECS (which is for 2xCO2) from 4xCO2 simulations by using a factor 2 (p. 6, l. 22). Is this really a standard method? Then it's certainly at odds with available knowledge (e.g., Boer et al., 2003; Knutti and Rugenstein, 2015). However, the authors could argue that they used the same approximation, crude or not, for all evaluated models.

Even if the focus of the paper were re-directed towards the BESM performance, I still suggest a modified title, for example: "Assessing the performance of climate change simulation results from BESM-OA2.5 in comparison to a CMIP5 model ensemble".

Specific and Technical Remarks

p. 1, l. 8 (Abstract): For the following two sentences I would rather expect a general assessment of BESM rather than pure repetition of specific parameter results. While it is obviously true (and worth mentioning) that BESM-OA2.5 is not an outlier off the CMIP ensemble, its appraisal ought to be more process-directed.

p. 2, l. 1: "…, commonly referred to as", I think this is rather a simplification for less developed models, so "…, sometimes given as" may be preferable.

p. 2, l. 20: There is a formal contradiction here: "… is robust from ... models" does not fit with "… uncertainty is likely to arise from … inter-model spread", please reformulate.

p. 3, l. 3: "Differences …", this sentence may be omitted as it is essentially repeated at p. 3, l. 3.

p. 3 l. 16: "… uses BAM … with simpler and computationally cheaper parameterizations"; Does this mean that BESM-OA2.3 uses the original BAM? Why has this been changed and could there be consequences of the simplification for the response behavior of the model as addressed in the present paper?

p. 3, l. 20: From the preceding text, it is puzzling that the simplified model should have a better representation of the ToA radiative budget. I assume, however, that this is a result of more careful parameter tuning (but this is not mentioned). Like referee#1, I also wonder whether this relatively large ToA radiative balance bias leads to a considerable present-day surface temperature bias. Does the coupled atmosphere-ocean model use a flux correction?

p. 3, l. 22: "surface layer"; I assume you mean the "planetary boundary layer", don't you? Or does tis refer to pure diagnostics, as suggested by the following sentence.

p. 4, l. 4: "general mean present-day climate state"

p. 4, l. 7: "BESM-OA2.5 also is capable …"; this sentence is rather vague, are you talking about ocean variability here? Or does this include the leading modes of long-term atmospheric variability like NAO, PNA etc. ?

p. 4, l. 11: "overturning"

p. 4, l. 12: "slightly"

p. 4, l. 14: You might wish to address the matter of storm track variability here, but only if this is supposed to be a field of BESM application in the future. And if it has been actually studied, of course.

p. 4, l. 30: "… the Gregory et al. (2004) method …"; from various reasons it may be preferable to introduce (and refer to) the respective method as "… the regression method …". Mainly, because using the terms "regression" and "radiative kernel" directly points to the methodical differences.

p. 5, l. 17: "… we extract the clear sky radiative flux components from the BESM and CMIP data bases in order to …"

p. 5, l. 25: (see major remarks above) – as the assumption is not necessarily true a remark should be made on the consequence for interpretation in case that there are substantial differences between the radiation modules.

p. 6, l. 4 (and l. 11): No information is given on how stratospheric temperature (and water vapour) changes are accounted for when calculating the feedback parameters. I recommend at least making a statement, if those contributions are included in the Planck feedback, or if they are shifted to the residuum *Re* (which I guess is, what you actually did). See also Rieger et al. (2017, their Fig. 5).

p. 6, l. 17: "… cloud feedback is approximated using …"; I'm aware that this is a standard method, so the authors are not responsible for the quality of this approximation.

p. 6, l. 22: I expect that the respective 30 year periods are not fully stationary as the deep ocean components of the various models have not reached equilibrium. If your analysis allows, please give some information on the remaining trend in the evaluated periods. Or have the data been de-trended before using them as an input to the radiative kernels?

p. 7, l. 5: "… the spatial inner product …"; the authors might like to introduce the term in this way, but I assume they compute what is elsewhere called the 'Pearson correlation coefficient', hence I recommend to use the latter term through the rest of the paper.

p. 7, l. 9: "These linear regressions …"; this sentence is hard to read and needs rewriting. With the current formulation, it is not possible to unravel for which purpose all-sky or clear-sky data haven been used.

p. 7, l. 11: The values given are at odds with what is written p. 5, l. 12, concerning *G*, λ, and *ECS*. Please, give an explanation (which is probably to be found in the fact that no actual equilibrium has actually been reached).

p. 7, l. 13: "… similar to those of Andrews et al. …"; in fact, the reader certainly expects no less than this, as those authors used CMIP5 data as well. Where does the difference come from? Interpolation as mentioned on p. 4, l. 23?

In the simulations with BESM, has there any form of "radiation double calling" been used to calculate radiative forcings or feedbacks? That could help to assess whether the radiation parameterization within BESM produces results (largely) consistent with the GFDL and NCAR kernels.

p. 7, l. 28: "Both radiative kernels are used …"

p. 8, l. 4: The following discussion (of Figure 4) is an example in a text flow that is largely out of scope with the paper's focus. Most of this is established knowledge from a multitude of previous papers. A clear change of perspective towards the specific features of BESM is advisable.

p. 8, l. 6: "The faster increase …", I assume you mean "stronger", don't you?

p. 8, l. 7: The two sentences discussing the possible cause-and-effect relation of water vapor and lapse-rate feedbacks is somewhat confusing. The general notion, I think, is the different degree of turbulent mixing in tropical, mid and polar latitudes. I recommend referring to, e.g., Po-Chedley et al. (2018), who draw a lucid and consistent picture of the latitudinal differences.

p. 8, l. 16: "… as noted in yellow and blue shaded areas in Figure 4", this hint would better be given when the discussion of Figure 4 starts (l. 4) or, alternatively" in the figure caption.

p. 8, l. 22: This paragraph is either too short (different cloud feedback results from different methods being a highly complex issue) or too long (as these general issues are not necessarily within the scope of the paper). Please focus on what could be a reason for the specific behavior of BESM in this particular case.

p. 8, l. 35: "This is due to …", a rather technical reasoning (which continues throughout this paragraph). The reader would rather be interested in the physical reason. I the cloud cover response over sea (60°S) and over sea ice (Arctic) less well simulated by BESM compared to land areas? Or could it by that there is a problem with the cloud phase feedback (e.g., Mitchell et al., 1989; Tan et al., 2016) in BESM? I would find it sufficient, if some ideas could be formulated, with hints to future research.

p. 8, l. 6: "stratocumulus region", this is presumably a different entity and not connected to the BESM peculiarities showing up in Figure 4.

p. 9, l. 10: The section 4.3 with its figures 6 and 7 (the scatter plots) is not very insightful to me. What are these correlation diagrams (especially Figure 6) supposed to teach the reader? Is this a standard diagnostic? Does the placement of BESM in the third quadrant reveal anything about this model in a physical sense? Please, give some reasoning for the figure's usefulness in the present paper. Interpretation of precipitation change patterns is more lucid; yet, it would be fine to know whether, e.g., the southward shift of the SPCZ in BESM does occur in other CMIP models too (even if not in the ensemble mean).

p. 9, l. 27: "… near the equator compared to the subtropics …"; ("as opposed to" suggests that the subtropics grow colder)

The statement beginning on p. 10, l. 21 "This increase …" sounds somewhat counter-intuitive and is, in my opinion, an oversimplification of what the cited papers actually say. Rather, the non-linear increase of water vapor available for condensation, as suggested by the Clausius-Clapeyron relation, is limited towards a more linear relation by tropospheric radiative cooling (Mitchell et al., 1987).

p. 10, l. 25: "ACCESS1-0 and HadGEM2-ES use …" up to the end of this paragraph: that may all be true, but the reader would rather be interested whether this implies anything for BESM.

p. 10, l. 32: "… (SLP) response pattern …"

p. 10, l. 30: This whole paragraph gives a lot of (by no means unfounded!) physical reasoning on tropospheric variability patterns, but in the end takes a simple similarity of the SLP mean response patterns from BESM and from the CMIP ensemble to indicate that BESM may well represent such variability patterns. This is a bold conclusion, which in my view would need backing from actual variability pattern analysis. Is such analysis planned?

p. 11, l. 23: "It is shown …", this a very odd 'conclusion', as this statement is common knowledge motivating any research on global warming, and it is certainly not "… shown in this study". Even "… confirmed by this study" would be a summary much too weak for motivating publication of this paper. Please, find a more specific main conclusion that is directed towards the BESM performance.

p. 11, l. 31: Here, some information about the BESM radiation module and its evaluation would emphasize that the radiative feedbacks calculated from BESM output within the CMIP range indeed indicate a good representation of such feedbacks inside that model (see major issues).

p. 12, l. 5: You might delete "However,"; I see no contradiction of this sentence with the preceding one.

p. 12, l. 12: "… is not the aim for the BESM development", this whole paragraph is a very puzzling wrap-up of your paper (see general remarks).

Figure 3, Figure 8: Please ensure that this figure will appear larger in the eventual paper, otherwise it will be hard to decipher.

Caption of Figure 6: "Shaded areas"; this return in several other figure captions, too. You mean the *white* areas, don't you?

<u>References (only if not already cited in the paper):</u>

Boer, G, Yu , B., 2003: Climate sensitivity and climate state, Clim. Dyn. 21, 167-176.

Chung E.S., Soden, B., 2015: An assessment of direct radiative forcing, radiative adjustments, and radiative Ffeedbacks in coupled Ocean-Atmosphere models, J. Clim. 28, 4152-4170.

Forster, P.M., et al., 2016: Recommendations for diagnosing effective radiative forcing from climate models for CMIP6, J. Geophys. Res. 121, 12460-12475.

Knutti, R., Rugenstein, M., 2015: Feedbacks, climate sensitivity and the limits of linear models, Philos. Tr. Roy. Soc. A 373, 20150146.

Mitchell, J.F.B., Wilson, C.A., Cunnington, W.M., 1987: On CO2 climate sensitivity and model dependence of results, Q. J. Roy. Metorol. Soc. 113, 293-322.

Mitchell, J.F.B., Senior, C., Ingram, W., 1989: CO_2 and climate – a missing feedback, Nature 341, 132-134.

Po-Chedley et al., 2018: Sources of intermodal spread in the lapse-rate and water vapor feedbacks, J. Climate 31, 3187-3206.

Rieger, V., Dietmüller, S., Ponater, M., 2017: Can feedback analysis be used to uncover the physical origin of climate sensitivity and efficacy differences? Clim. Dyn. 49, 2831-2844.

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

Tan, I., Storelvmo, T., Zelinka, M., 2016: Observational constraints on mixed-phase clouds imply higher climate sensitivity, Science 352, 224-227.