

Geosci. Model Dev. Discuss., referee comment RC3
<https://doi.org/10.5194/gmd-2021-91-RC3>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-91

Anonymous Referee #2

Referee comment on "NorCPM1 and its contribution to CMIP6 DCP" by Ingo Bethke et al.,
Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-91-RC3>, 2021

This manuscript describes the NorESM1-based decadal climate prediction system NorCPM1, which contributes to the CMIP6 DCP set of decadal prediction systems. The different kinds of model simulations, assimilation simulations and decadal hindcasts, are presented and discussed extensively. Then, the skill of these simulations with respect to observations is quantified and discussed. The authors find that the Ensemble Kalman Filter used in data assimilation for these simulations is fit-for-purpose, resulting in skilful prediction of important climatic indices up to 10 years in advance. However, like other papers on these type of CMIP6 simulations, the authors also find a substantial amount of decadal prediction skill to be related to the forced response instead of hindcast initialisation. In the end, the authors demonstrate skill beyond the forced response for some variables, regions and certain lead times.

I find this manuscript comprehensive, well-structured, and well-written. Despite its length, it is relatively easy to follow and nice to read. After all, the long nature of this model description paper and its four appendices benefits the clarity of the manuscript in my opinion. The authors succeed in honestly and fully present and explain the main features as well as drawbacks of their model simulations, and place them in observational context. As a result, this is a very complete account of not only the NorCPM1 simulations, that could almost serve as a review paper about data assimilation and decadal prediction in general for anyone that is willing to put in the time to read the entire thing.

I thoroughly enjoyed the journey through this paper and therefore recommend publication in *Geoscientific Model Development* after some (mostly minor) comments, which I list below, have been considered by the authors.

Comments

I only have one suggestion to the authors that warrants mentioning outside the "specific comments" section. The comparison of skill between the hindcasts and/or assimilation simulations and the historical runs to assess the contribution of the forced response to skill currently relies on skill differences (i.e. $ACC(\text{hind})-ACC(\text{hist})$ and so on). Recently, Smith et al. (2019) demonstrated that this practice does not fully capture the benefit of initialisation due to non-linear interactions in the system, and propose the use of *residual ACC*, that subtracts the forced signal (hist) from the hindcast signal prior to skill calculation. In their work, Smith et al. show that the use of residuals more accurately describes what the authors set out to assess here. I suggest the authors either present some results using residuals to illustrate this alternative metric, or at least discuss the residual as an alternative (and potentially superior) approach.

Specific Comments

II. 24-25 I do not think it is clear at this point what "non-assimilation experiments" are. Maybe just call them "NorESM1 simulations" for clarity?

II. 54-55 This statement could benefit from one or two publications to back up the claim.

II. 63-74 Similarly, this entire paragraph can in my opinion not go without citations. Please add some published work to back up the statements made here.

II. 80-81 Take it or leave it: to me, the order historical->assimilation->hindcasts would make more sense.

II. 131, 134 Naïve question: is NorCPM1 a purely "physical model"?

II. 302-304 Why the two different baseline periods? This might be obvious, but I am struggling to understand the motivation for this choice.

I. 316 I suppose "The approach" here references the SCDA approach used in assim -i2?

II. 414-416 ACC is known to be sensitive to spurious trends as a skill metric (e.g. Smith et al., 2019). RMSE-based metrics such as MSSS might be more realistic when it comes to skill assessment. While I see why the authors choose ACC for this study, I think a short sentence on potential ACC shortcomings might be in order here or in the discussion.

Figure 2 Are these global metrics?

I. 465 There is a "." Missing after "observations".

I. 507 Is this referencing "changes" due to data assimilation? I suggest being explicit about that.

II. 532-533 I am struggling to see the smaller improvement for S300 compared to T300 reflected in figure 7. I suppose the authors reference figures 7f and 7g here, comparing the assimilation simulations' improvement relative to historical simulations for T300 and S300? When I visually compare 7f to 7g, I am struggling to make out a clear "winner". If anything, it appears to me that S300 shows slightly higher values. Could you please comment on that? In general, the text could at times benefit from more clear reference of figure sub-panels when making statements in the text.

Figure 7 The labelling of assim-i1 and assim-i2 as ANA1 and ANA2 in the figures is not optimal, as it is inconsistent with the labelling in the text. I suggest changing the labels in the figures for consistency. This is an issue in all figures that show skill comparison including assimilation simulations.

II. 559 ff. This discussion is reminiscent of work done by Koul et al. (2020) on which SPG index represents which underlying physical processes. I think this part could be shortened by referring to the above mentioned paper.

II. 608-610 I think it is important to point out this "small contribution" is insignificant (is it?).

II. 676-678 At least one citation should be given for the statement on internal vs external causes of the global warming hiatus (e.g. Medhaug et al. 2017?).

II. 696-698 I assume the authors used ACC^{*2} to calculate explained variance? This should be made explicit.

II. 741-754 In a recent paper, Borchert et al. (2021) showed increased contribution of forcing to decadal SPNA SST prediction skill in CMIP6 compared to CMIP5, using a multi-model ensemble including NorCPM1. How do the authors square their findings presented here with what Borchert et al. (2021) found?

I. 825 To me, the phrase "potential predictability" refers to the skill of simulations initialised from a piControl simulation with respect to the same piControl simulation. What the authors demonstrate here (skill of hindcasts initialised from a reanalysis-type simulation with respect to a reanalysis that did not directly assimilate biogeochemistry, but produces biogeochemistry that is consistent with observed physical climate) is to me more than that. The authors might want to re-consider their phrasing here so as to avoid underselling their findings.

I. 888 It should be "in the Pacific sector".

Figs. 25 & 26 The comparison of mean skill and skill of the mean signal is interesting. However, the authors only mention Fig. 26 briefly in one paragraph. Would 1-2 more sentences on this topic be interesting to a wider readership?

II. 1253 ff I particularly like this (brief) mention of observational error, particularly in light of the small but important differences between assim-i1 and assim-i2. This might be a candidate for the main text, as this info is currently a little hidden.

I. 1281 As far as I can make out, the mean state of AMOC is not shown in Figure C3c, nor is the vigorous nature of simulated AMOC. This might be a phrasing issue, but I was looking for this information in the figure. Rephrase to avoid this in the future?

Borchert, L. F., Menary, M. B., Swingedouw, D., et al. (2021) Improved decadal predictions of North Atlantic subpolar gyre SST in CMIP6. *Geophysical Research Letters*, 48, e2020GL091307. <https://doi.org/10.1029/2020GL091307>

Koul, V., Tesdal, J.E., Bersch, M. et al. (2020) Unraveling the choice of the north Atlantic subpolar gyre index. *Sci Rep* **10**, 1005. <https://doi.org/10.1038/s41598-020-57790-5>

Medhaug, I., Stolpe, M., Fischer, E. *et al.* (2017) Reconciling controversies about the 'global warming hiatus'. *Nature* **545**, 41–47. <https://doi.org/10.1038/nature22315>

Smith, D.M., Eade, R., Scaife, A.A. *et al.* (2019) Robust skill of decadal climate predictions. *npj Clim Atmos Sci* **2**, 13. <https://doi.org/10.1038/s41612-019-0071-y>