

Geosci. Model Dev. Discuss., author comment AC2
<https://doi.org/10.5194/gmd-2021-91-AC2>, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Reply on RC3

Ingo Bethke et al.

Author 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-AC2>, 2021

We thank the reviewer for positive and constructive feedback. Below, we will address the comments one by one.

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(hind)-ACC(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.

Thanks for the information and suggestions.

We used ACC differences for mainly two reasons: to be able to directly compare with similar previous studies (e.g., Yeager et al., 2018), and to be able to evaluate skill benefits consistently with respect to different benchmarks (analysis, persistence, free historical). The use of ACC differences further allowed us to adopt the local and field significance testing from Yeager et al. (2018), whereas we would not know how to appropriately deal with uncertainty related to regressing out the forced signal. For these reasons we prefer not to replace our ACC differences plots. Instead we prefer to opt for presenting additional figures using the Smith et al. method in the Supplementary Information and briefly mention differences (if any) in the main manuscript. We prefer not to present any results with the new method in the main manuscript, as we are limited concerning the number of figures we can add. We will add discussion on using the Smith et al. method (i.e., regressing out the forced model signal prior to skill calculation) as an alternative, more robust approach to computing ACC differences.

Yeager, S. G., Danabasoglu, G., Rosenbloom, N. A., Strand, W., Bates, S. C., Meehl, G. A., Karspeck, A. R., Lindsay, K., Long, M. C., Teng, H., & Lovenduski, N. S. (2018). Predicting Near-Term Changes in the Earth System: A Large Ensemble of Initialized Decadal Prediction Simulations Using the Community Earth System Model, *Bulletin of the American Meteorological Society*, 99(9), 1867-1886.

I. 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?

Thanks for the suggestion. We will rephrase "non-assimilation experiments" to "NorESM1 simulations".

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

Thanks for the suggestion. We will add the below citations.

Årthun, M., E. W. Kolstad, T. Eldevik, and N. S. Keenlyside (2018), Time Scales and Sources of European Temperature Variability, *Geophys Res Lett*, 45 (0), doi:10.1002/2018GL077401.

Athanasiadis, P. J., S. Yeager, Y.-O. Kwon, A. Bellucci, D. W. Smith, and S. Tibaldi (2020), Decadal predictability of North Atlantic blocking and the NAO, *npj Climate and Atmospheric Science*, 3 (1), 20, 10.1038/s41612-020-0120-6.

Sutton, R. T., , and D. L. R. Hodson, 2005: Atlantic Ocean forcing of North American and European summer climate. *Science*, 309 , 115–118.

Omrani, N. E., N. S. Keenlyside, J. Bader, and E. Manzini (2014), Stratosphere key for wintertime atmospheric response to warm Atlantic decadal conditions, *Climate Dynamics*, 42 (3-4), 649-663, 10.1007/s00382-013-1860-3.

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.

Thanks for the suggestion. We will add citations as indicated below.

Current climate prediction systems are thought to not fully realise the predictive potential on multi-year times scales, although the practical limits of predictability themselves and their regional variations are poorly known (Branstator et al., 2012; Sanchez-Gomez et al., 2015; Smith et al., 2020).

The skill of climate prediction depends on the initialisation of internal climate variability state, the representation of the dynamics and processes that lead to predictability, and the representation of the climate responses to external forcings (Branstator and Teng, 2010; Latif and Keenlyside, 2011; Bellucci et al., 2015; Yeager and Robson, 2017).

Dynamical climate prediction systems typically use Earth system models (initially developed to provide uninitialised long-term climate projections) for representing the dynamics and the responses to external forcings (Meehl et al., 2009; Meehl et al., 2013).

Importantly, the dynamical prediction systems add initialisation capability to the ESMs, adopting a wide range of initialisation strategies (see Section 2.2.1) (Meehl et al., 2021).

A better understanding of the three aspects – initialisation, model dynamics, forcing response – is fundamental for better exploiting the climate predictive potential and improving estimates of climate predictability (Keenlyside and Ba, 2010; Cassou et al., 2018; Verfaillie et al., 2021).

The existing climate prediction systems undersample effects of model and initialisation uncertainty and are not necessarily well suited to address questions related to changes in the observing system. The benefits from using advanced data assimilation for

initialisation, especially in an ocean density coordinate framework, are not well explored.

Bellucci, A., et al. (2015), Advancements in decadal climate predictability: The role of nonoceanic drivers, *Rev Geophys*, 53, 165–202, 10.1002/2014RG000473.

Branstator, G., and H. Y. Teng (2010), Two Limits of Initial-Value Decadal Predictability in a CGCM, *J Climate*, 23(23), 6292–6311

Branstator, G., H. Y. Teng, G. A. Meehl, M. Kimoto, J. R. Knight, M. Latif, and A. Rosati (2012), Systematic Estimates of Initial-Value Decadal Predictability for Six AOGCMs, *J Climate*, 25(6), 1827–1846

Cassou, C., Y. Kushnir, E. Hawkins, A. Pirani, F. Kucharski, I.-S. Kang, and N. Caltabiano (2018), Decadal Climate Variability and Predictability: Challenges and Opportunities, *B Am Meteorol Soc*, 99(3), 479–490, 10.1175/BAMS-D-16-0286.1.

Keenlyside, N. S., and J. Ba (2010), Prospects for decadal climate prediction, *Wiley Interdisciplinary Reviews: Climate Change*, 1(5), 627–635, 10.1002/wcc.69.

Latif, M., and N. S. Keenlyside (2011), A perspective on decadal climate variability and predictability, *Deep Sea Research Part II: Topical Studies in Oceanography*, 58, 1880–1894

Meehl, G. A., et al. (2009), Decadal Prediction Can It Be Skillful?, *B Am Meteorol Soc*, 90(10), 1467–1486, <https://doi.org/10.1175/2009BAMS2778.1>.

Meehl, G. A., et al. (2013), Decadal Climate Prediction: An Update from the Trenches, *B Am Meteorol Soc*, 10.1175/bams-d-12-00241.1.

Meehl, G. A., et al. (2021), Initialized Earth System prediction from subseasonal to decadal timescales, *Nature Reviews Earth & Environment*, 2(5), 340–357, 10.1038/s43017-021-00155-x.

Sanchez-Gomez, E., C. Cassou, Y. Ruprich-Robert, E. Fernandez, and L. Terray (2015), Drift dynamics in a coupled model initialized for decadal forecasts, *Climate Dynamics*, 1–22, 10.1007/s00382-015-2678-y.

Smith, D. M., et al. (2020), North Atlantic climate far more predictable than models imply, *Nature*, 583(7818), 796–800, 10.1038/s41586-020-2525-0.

Verfaillie, D., F. J. Doblas-Reyes, M. G. Donat, N. Pérez-Zanón, B. Solaraju-Murali, V. Torralba, and S. Wild (2021), How Reliable Are Decadal Climate Predictions of Near-Surface Air Temperature?, *J Climate*, 34(2), 697–713, 10.1175/JCLI-D-20-0138.1.

Yeager, S. G., and J. I. Robson (2017), Recent Progress in Understanding and Predicting Atlantic Decadal Climate Variability, *Current Climate Change Reports*, 3(2), 112–127, 10.1007/s40641-017-0064-z.

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

Thanks for the suggestion. We will rephrase it to "two sets of DCPD coupled reanalysis simulations, and two sets of initialised DCPD hindcast simulations that obtain their initial conditions from the two reanalysis sets."

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

Because we realized that the original formulation "physical model" was inaccurate, we will rephrase the formulation to "The Earth system model used in NorCPM1".

NorESM1 does feature ocean biogeochemistry (which albeit does not influence the physical climate) as well as atmospheric chemistry. Also, some parameterizations of NorESM1 are based on empirical statistical relations rather than on first physical principles. Nevertheless they are intended to model physical processes. The EnKF does exploit statistical relations between physical variables, but relies on the ESM to obtain these relations.

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

We used two periods primarily because OISST data were not available prior September 1981. We will add this information and modify the sentence to "but over the period 1982-2010 when assimilating OISSTV2 data (i.e., beyond 2010) because OISSTV2 was not available before this period."

In hindsight, it would have been more consistent if we had extended the OISST back in time with HadISST2 data (like we did for *assim-i2*) and then used the period 1980-2010. But we checked and found the 1982-2010 OISST-based climatology and the 1980-2010 HadISST2/OISST-based climatology (1980-1981 HadISST, 1982-2010 OISST) are practically indistinguishable.

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

This is correct.

In RC1, the reviewer noted that characterising *assim-i2* as SCDA can be misleading because the term is typically used with respect to ocean and atmosphere and involves either atmospheric state update or use of atmospheric observations.

We will therefore rephrase the text in question to "In *assim-i2*, we allow the oceanic observations to update the ocean and the sea ice components. In this case the system is a strongly coupled DA system (SCDA) with respect to the ocean and sea ice components, where the oceanic observations influence the sea ice component of the system both at the DA step and during the model integration. To avoid confusion with atmosphere-ocean SCDA (e.g., Penny et al., 2019), we will refer to the *assim-i2* system approach as SCDA-OSI. The SCDA-OSI assures a more consistent initialisation across the ocean and sea ice components and exploits the longer temporal coverage of oceanic observations relative to sea ice observations (see also Appendix A)."

Penny, S. G., Bach, E., Bhargava, K., Chang, C.-C., Da, C., Sun, L., and Yoshida, T.: Strongly Coupled Data Assimilation in Multiscale Media: Experiments Using a Quasi-Geostrophic Coupled Model, *Journal of Advances in Modeling Earth Systems*, 11, 1803–1829, <https://doi.org/10.1029/2019MS001652>, <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001652>, 2019.

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.

We will add some text on potential ACC shortcomings related to spurious skill and

amplitude errors based on the reasoning below. We will also consider adding a MSSS map figure to the Supplementary Information of the manuscript to illustrate overall similarity between the ACC and MSSS measures (see Supplement to this reply).

While we would reserve the term "spurious trends" for trends that e.g. arise from observational errors (probably not what the reviewer is hinting at), we generally agree with the reviewer's statement. ACC scores of multi-year averages are subject to spurious correlations because the climate signal of the real world can covary with the numerical predictions just by chance during the limited evaluation window (which can be interpreted as sampling error due to undersampling of climate variability). In case of such random-covariability, the amplitudes of the signals tend to differ since there is no apparent reason they should be the same. The MSSS would penalize for the amplitude error while the ACC does not, making the MSSS more robust to effects of spurious correlations. Our interpretation of Smith et al. (2019) is, however, that they consider MSSS as overly pessimistic because of penalization of amplitude errors that potentially can be corrected posteriori and they therefore favor the use of ACC.

Initially, we prepared Figures for both ACC and MSSS. While the MSSS scores results were generally lower than the ACC scores (because of penalization of amplitude errors in the former), they were qualitatively consistent with the ACC scores and did not change our general skill assessment. We dropped the MSSS because the key literature we compare to favors ACC and also we realized that presenting both ACC and MSSS and discussing subtle differences between them would make the paper too long.

Figure 2 Are these global metrics?

The metrics shown in Figure 2 are global, considering both vertical and horizontal dimensions. They are computed as detailed in Equations 2–7 in Section 3.1.1.

We will change the Figure caption to "Global assimilation statistics (see Section 3.1.1. for definitions)." to make this clearer. We will also correct the numbering of the Equations from 2–7 to 1–6.

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

Thanks. We will fix it in the revised version.

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

Thanks for the suggestion. We will make clear this paragraph continues to discuss the impact of the assimilation on the mean state and will rephrase the text to "Some impact of DA on the mean state of *assim-i1* is also seen...".

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.

Thanks for noting this. We will remove the sentence "The improvements for S300 are smaller ...". We made a mistake in our first version of Figure 7, resulting in smaller ACCs for S300. Later, we noticed and rectified this but forgot to correct the manuscript text.

Also, we will try to refer more explicitly to sub-panels in the revised version of the manuscript.

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.

Thanks for the suggestion. We will make the labels/acronyms used on the figures consistent with those used in the manuscript text (we will also move the labels out of the panels as suggested by RC1). We will have to check whether the figure labels are still readable or become too small when changing to the long acronyms. Otherwise, we will use short acronyms also in the manuscript text. In addition, we will add a table that summarizes essential information of the different experiments.

I. 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.

The text in question was not intended as discussion but rather to provide some minimum motivation and background for the readers less familiar with the North Atlantic subpolar gyre concept. We therefore prefer to keep the text but will add the citation to Koul et al. (2020) for further reading.

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

Thanks for pointing this out.

We have computed a p-value of 0.085 using block-bootstrapping—with block lengths of 5 years—and re-sampling of members in the computation of the ensemble mean (see Appendix B for details on p-value computation).

Following Yeager et al. (2018), we have used a 10 % significance level throughout the manuscript and it would be inconsistent if we switched to 5 % significance level for this test. Results presented in Appendix D in Figure D2 are also suggestive for a small, but systematic influence from external forcing on simulated NINO34 SSTs, with positive correlation values for all calendar months.

We will moderate the statement to "The ensemble mean of *historical* has a smaller amplitude and is only marginally correlated with the observed index ($r=0.2$, $p=0.085$, $\alpha=0.1$), suggesting a potential small contribution from external forcing."

I. 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?).

Thanks for the suggestion. We will cite Medhaug et al. (2017).

I. 696-698 I assume the authors used ACC2 to calculate explained variance? This should be made explicit.**

Thanks for the suggestion. We will rephrase the sentence to "One should note that a change in correlation from 0.6 to 0.9 equates to more than doubling in explained variance from 36 % to 81 % (estimated by the square of the correlation)."

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?

We cannot say with certainty how our findings square with Borchert et al. (2021), who found an increasing contribution from external forcings to SPNA skill in CMIP6 relative to CMIP5. It would be interesting to know the isolated effect of including NorCPM1 CMIP6 DCP output in their multi-model analysis. It is possible that NorCPM1 represents a model outlier.

Our Figure 13c shows that the EN4-based observational estimate of upper SPNA temperature is negatively correlated with the ensemble mean of NorCPM1's historical experiment over the period 1950-present. This is contrasting the general positive SPNA skill contribution from external forcing that Borchert et al. found in multi-model ensembles (e.g., their Fig. 1). While less clear than for upper ocean temperature, SST is also negatively correlated over SPNA (Figure 14c) with a correlation value below -0.4 (Figure 14e). Moreover, Figure 12 suggests that the ACCs for SST in the SPNA region of the historical experiment are predominantly negative because (ACC_hindcast-i1 - ACC_historical) is larger than ACC_hindcast-i1 in most of the region. Notwithstanding the above, it is possible (but we have not checked it) that NorESM1's historical experiment using CMIP5 forcings features even more negative ACCs for SST in the SPNA region than NorCPM1's historical experiment with CMIP6 forcings, somewhat consistent with Borchert et al. (2021).

In the revised manuscript, we will point out the apparent discrepancy to the findings of Borchert et al. (2021). Following the sentence with "poorly performing simulated forced trend" (line 743) we will add "This result stands in contrast to multi-model findings (that include NorCPM1) suggesting a positive contribution of the forced signal to SPNA temperature skill over a comparable period (Borchert et al., 2021)."

As detailed in Appendix C (lines 1311-1325), we suspect a problem with CMIP6 land use change specification, leading to an unrealistic historical cooling trend over North America in NorCPM1. Via downstream effects, the continental cooling (which we suspect is an artifact) may contribute to the SPNA cooling trend shown after 1980 (e.g., Fig. 13c), exacerbating the discrepancy between the observed and simulated SPNA temperature evolutions. How rectifying NorCPM1's CMIP6 land use change specification and also how replacing NorESM1 with NorESM2 may impact SPNA prediction skill (both initialised and uninitialised) in NorCPM is subject to future investigation.

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.

As the manuscript text states, we adopted the definition of "potential predictability" from Yeager et al. (2018), a key reference for our study.

We agree that our use of "potential predictability" differs from the one in some earlier studies that utilized a piControl simulation (e.g., Collins et al., 2006) and we will more clearly point this out in the revised version.

We will rephrase the text to "Following Yeager et al. (2018), we therefore also analysed

the model's ability to hindcast its own analysis over the period 1960–2018 (Fig. 16). We will refer to this as the potential* predictability, using the asterisk to indicate that it differs from more conventional potential predictability estimates based on self-prediction that typically utilize a pre-industrial control simulation (e.g., Collins et al., 2006)."

Collins, M., Botzet, M., Carril, A. F., Drange, H., Jouzeau, A., Latif, M., Masina, S., Otteraa, O. H., Pohlmann, H., Sorteberg, A., Sutton, R., & Terray, L. (2006). Interannual to Decadal Climate Predictability in the North Atlantic: A Multimodel-Ensemble Study, *Journal of Climate*, 19(7), 1195-1203.

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

Thanks. We will correct this.

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?

Thanks for the comment. We will try to add a few sentences for the wider readership based on the below. Please let us know if you had something different in mind.

We consider the mean skill of local (grid-cell scale) sea ice concentration as a simple integrated metric for the sea ice performance of our prediction system that complements the skill of the hemispheric mean. It illustrates, for example, that skill reemergence in the second year seen in the total sea ice area does not necessarily translate into statistically significant local prediction skill during the second year. Presumably, this is partly related to the lack of skill for the Arctic oscillation which primarily causes a regional re-distribution of sea ice with limited effect on the hemispheric mean, whereas the hemispheric mean is more controlled by the externally forced trend. Efforts that target sea ice prediction from a climate service perspective are advised to do more elaborated analysis (that were beyond the scope of our paper), like predicting the temporal evolutions of principal components or regional averages of sea ice concentration, to additionally consider the predictability of atmosphere and subsurface ocean variability, and to perform statistical/dynamical downscaling.

I 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.

Thanks for the suggestion. We will move this paragraph to the main manuscript.

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?

We are a bit unsure about this comment. Could the reviewer elaborate?

Figure C3c shows the climatological Atlantic meridional overturning circulation (AMOC) streamfunction in units Sv for NorCPM1. To our knowledge, this is the standard way for characterising the mean state of AMOC. Values in excess of 30 Sv clearly indicate the vigorous nature of the simulated AMOC ("vigorous" in the sense of strong, not necessarily variable). While this is somewhat different from the "time-mean AMOC strength evaluated at a fixed latitude", these two characterizations are closely related. We will rephrase the figure caption text from "Atlantic meridional overturning streamfunction" to "Atlantic

meridional overturning circulation (AMOC) streamfunction" to be more precise.

The overly strong AMOC has been a longstanding, well-documented problem of previous NorESM versions (e.g., Bentsen et al., 2013, Cheng et al., 2013) but has been mitigated through updates and re-tuning of the ocean code in more recent NorESM versions (Guo et al., 2019; Seland et al., 2020).

Bentsen, M., Bethke, I., Debernard, J. B., Iversen, T., Kirkevåg, A., Seland, Ø., Drange, H., Roelandt, C., Seierstad, I. A., Hoose, C., and Kristjánsson, J. E.: The Norwegian Earth System Model, NorESM1-M – Part 1: Description and basic evaluation of the physical climate, *Geosci. Model Dev.*, 6, 687-720, <https://doi.org/10.5194/gmd-6-687-2013>, 2013.

Cheng, W., Chiang, J. C. H., & Zhang, D. (2013). Atlantic Meridional Overturning Circulation (AMOC) in CMIP5 Models: RCP and Historical Simulations, *Journal of Climate*, 26(18), 7187-7197.

Guo, C., Bentsen, M., Bethke, I., Ilicak, M., Tjiputra, J., Toniazzo, T., Schwinger, J., and Otterå, O. H.: Description and evaluation of NorESM1-F: a fast version of the Norwegian Earth System Model (NorESM), *Geosci. Model Dev.*, 12, 343-362, <https://doi.org/10.5194/gmd-12-343-2019>, 2019.

Seland, Ø., Bentsen, M., Olivié, D., Toniazzo, T., Gjermundsen, A., Graff, L. S., Debernard, J. B., Gupta, A. K., He, Y.-C., Kirkevåg, A., Schwinger, J., Tjiputra, J., Aas, K. S., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Karset, I. H. H., Landgren, O., Liakka, J., Moseid, K. O., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iversen, T., and Schulz, M.: Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations, *Geosci. Model Dev.*, 13, 6165-6200, <https://doi.org/10.5194/gmd-13-6165-2020>, 2020.

Please also note the supplement to this comment:

<https://gmd.copernicus.org/preprints/gmd-2021-91/gmd-2021-91-AC2-supplement.pdf>