Review of „Climate Model Selection by Independence, Performance and Spread (ClimSIPS) for regional applications“ by Merrifield et al. (2023)

by Swen Brands

General comment: The authors provide a comprehensive description of the ClimSIPS tool for weighting Global Climate Models according to the three criteria mentioned in the title. The study provides a detailed introduction to the concepts and methods used in this research field and comes with a detailed bibliography, so that it is almost a review article. The manuscript is well written, timely and relevant and I recommend a minor version in which the following points should be addressed:

1. Since the model independence results obtained in your study are very similar to those obtained in Brands (2022), the authors might wish to cite this study in the revised manuscript. Particularly the „one family one vote standard“ (e.g. line 454) was also adopted in Brands (2022), where GCM clusters were built by combining the a priori criterion „use of the same AGCM family“ based on Brands et al. (2023) with the a posteriori criterion „error pattern correlation coefficient > +0.65“ (the Boé 2018 nomenclature is followed here). Albeit another predictor was used to measure a posteriori model dependence, the outcome is similar to yours (compare Figure 3 in Brands 2022 to your Figure 3). This shows that the results are robust to changes in the applied methodology.

2. Lines 53-56: *„modeling centers often contribute several versions of their base model under different names as well (Leduc et al., 2016); these variants differ by, for example, the spatial resolution of some model components or biogeochemical cycling, which may influence their simulated climate in ways that are difficult to anticipate.“*

Should start with an uppercase letter (“Modeling centers...”). Model versions from the same center often differ by the inclusion of entire numerical sub-models describing specific Earth system components in addition to the basic four components atmosphere, land-surface, ocean and sea-ice. In this context it is interesting to note that the names and versions of all sub-models representing up to 12 climate system components is listed in Brands et al. (2023) for 61 nominally distinct GCMs from CMIP5 and 6. In this extensive metadata archive, you can see how the distinct modeling groups have built their models in terms of included sub-models / Earth System components.

2. Lines 56-57: *„Adding further complexity, even uniquely named models from different modeling centers fall along a spectrum of uniqueness.“*

I do not fully understand what you mean with this sentence. Could you provide an example for *„uniquely named models from different modeling centers“* ?

3. Lines 58-61: I think the Brands et al. (2023) GCM metadata archive is relevant in this sentence as well and the authors might wish to refer to it. The archive could be alternatively cited in lines 78-80 and is useful for determining „a priori“ dependencies within in the CMIP ensemble, as defined by Boé (2018).

4. Lines 119-121. Meanwhile, the EURO-CORDEX model selection team has come to a final recommendation for the driving GCMs from CMIP6. Please see Sobolowski et al. (2023) for more details.

6. Lines 205-217: Please indicate the time aggregation of the GCM and reanalysis data you are using. Is the study based on monthly-mean data?

7. Lines 218-219 and elsewhere: Would make sense to use the terms „a priori“ and „a posteriori“ model dependence (Boé 2017) in this study?

8. Lines 230-238: The definition of the INV and SME groups is clear but more information is needed on how you define the FAM group. For example, ACCESS-ESM1-5 is here considered an „SME“ model, meaning that it „*[...] is represented by multiple members (e.g., initial condition ensembles, perturbed physics ensembles, combinations thereof) but is not determined to be part of a broader multi-model family.*“

However, a closer look at the „source“ attributes of the corresponding netCDF files from ESGF and at the reference articles (doi: 10.1071/ES19035, 10.5194/gmd-12-4999-2019, 0.5194/gmd-4-723-2011,0.5194/gmd-4-1051-2011) reveals that the entire ACCESS GCM family is based on versions of the atmospheric sub-models (or AGCMs) developed at the MetOffice-Hadley Centre. Namely, ACCESS-ESM1-5 makes use of the „HadGAM2“ AGCM that is also used by the HadGEM2-ES and HadGEM2-CC coupled model configurations. HadGAM2 was further developed into „MetUM-HadGEM3-GA7.1“, constituting the AGCM used in both Hadley Centre‘s and CSIRO‘s coupled model configurations used in CMIP6, e.g. HadGEM3-GC31-MM (doi: 10.1071/ES19040) and ACCESS-CM2 (doi: 10.1071/ES19040). Thus, it is reasonable to put the HadGEM and ACCESS coupled model configurations to the same family, as was done in Brands (2022), because they essentially share their atmospheric component. Following your nomenclature, this would mean assigning a „FAM“ to ACCESS-ESM1-5. Note that all the aforementioned model metadata is available at one glance from Brands et al. (2023).

9. Lines 240-247: Could you also shortly refer to the disadvantages of the a posteriori / output data – driven approach to measure GCM dependence ? Here, only the advantages are described so far.

10. Lines 259-261: Please add an equation to define inter-member GCM distance.

11. Lines 279-280: The observational density underlying theses gridded dataset is also reduced during the first half of the 20th century, particularly in the Southern Hemisphere.

12. Lines 285-287: „relative change with respect to a historical period“ is not considered „model performance“, as far as I know. Traditionally, the term „model performance“ refers to model error with respect to observations.

13. Line 307: „confined to subtropical regions“ > „confined to the tropics and subtropics“

14. Lines 323-324: The Brands et al. (2023) metadata archive comprising names and versions of the sub-models used in each GCM configuration helps to identify the „very similar but differently named models“ you refer to in this sentence.

15. Figure 3b) I can here see 3 independent clusters instead of the 2 indicated in the caption.

16. Between page 16 and 17 it seems that some running text is missing.

17. Lines 434-436: Similarities might be caused by the use of similar ocean models. Distinct versions of the same OGCM (NEMO) are used in the CNRM and IPSL GCMs (see Brands et al. 2023 for further details).

References

Boé, J. (2018). Interdependency in multimodel climate projections: Component replication and result similarity, *Geophysical Research Letters*, 45, 2771– 2779. https://doi.org/10.1002/2017GL076829

Brands, S. (2022). Common error patterns in the regional atmospheric circulation simulated by the CMIP multi-model ensemble. *Geophysical Research Letters*, 49, e2022GL101446. https://doi.org/10.1029/2022GL101446

Brands, S., Tatebe H., Danek, C., Fernández, J., Swart, N. C., Volodin, E., Kim, Y.H., Collier, M., Bi, D., Tongwen, W. (2023). SwenBrands/gcm-metadata-for-cmip: First standalone version of GCM metadata archive "get_historical_metadata.py" (v1.1). Zenodo. https://doi.org/10.5281/zenodo.7813495

Sobolowski, S. et al. (2023). EURO-CORDEX CMIP6 GCM Selection & Ensemble Design: Best Practices and Recommendations. Zenodo. https://doi.org/10.5281/zenodo.7673400