Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-1-RC1, 2020

Geoscientific
Model Development
Discussions
Open Access
EGU

Interactive comment on "Evaluating Simulated Climate Patterns from theCMIP Archives Using Satellite and Reanalysis Datasets" *by* John T. Fasullo

Anonymous Referee #1

Received and published: 1 April 2020

The manuscript describes an objective approach to evaluate biases in climate model simulations, providing scores based on pattern correlation between key model fields and the most up-to-date observational datasets. Variables are selected on the basis of the most relevant open issues raised on model performances, and are gathered in three realms: the energy budget realm, the water cycle realm, the dynamical realm. Overall scores are obtained, combining weighted scores from different variables, and different timescales are taken into account. The improvement (or lack of) across different generation of the CMIP experiments is also assessed.

Overall, I think that the paper contains some interesting and useful comparisons, and,

as far as I am aware of, it is the first time that such diverse metrics are gathered, in order to assess biases in coupled model simulations in a synthetic and comprehensive way. Extending the analysis to newly available CMIP6 datasets is also a valuable point.

What I found lacking is a bit of context about other model diagnostics and a discussion of the physical relevance of biases. I also have a few remarks about the completeness in describing the methodology. I provide some suggestions on how the paper could be improved in the specific comments below. My general opinion is that the manuscript could be published, subject to minor revisions, as I detail in the following.

———-

Specific comments

l. 52-58: I found that this paragraph, focusing on model diagnostics as a research community service, lacks a bit of context in terms of background on how diagnostics of model performances have been developed in the context of the IPCC and the PCMDI. I also think that this section might benefit of a survey of known sources of biases in models, e.g. the parametrisations, the unresolved scales, the choice of the grids, the numerical scheme. In this respect, the author might mention some of the diagnostics and metrics that have been most recently designed to address some of the specific issues that are considered here, as for example Greve et al. 2018, for the water cycle, precipitation and its regional downscaling, or Lembo et al. 2019, for radiative budgets and transports.

l. 65-67: When data records are not available, I think that it is also important to weigh models beforehand, when the multi-model inter-comparison is performed (e.g. Knutti et al. 2017). These approach has been successfully applied to regional downscaling of global climate model projections (e.g Lorenz et al. 2018), proving that metrics are more relevant to the end user of the model exercise, if models are appropriately weighted. I wonder if it would be possible to adopt a similar approach, with relatively small effort, to the analysis here presented.

l. 68: I think that the appropriate reference for this is Hourdin et al. 2017. Schmidt et al. 2017 refer to a subset of US models from those analysed in Hourdin et al. 2017.

ll. 112-113: I agree that from an observational-based point of view the net surface fluxes are the most challenging, especially if dealing with satellite measurements and inverse techniques. On the other hand, from a model perspective, surface fluxes are the result of several parametrisations and are thus straightforwardly provided, while the retrieval of the vertical integral of atmospheric energy divergence is made difficult by the vertical discretisation and numerical sources of mass imbalance, requiring offline corrections. I think this would be worth mentioning here.

l. 135: why is the 500hPa eddy geopotential height preferred to the 500hPa geopotential height, which is usually made available as an output of a climate model (e.g. in the ESGF repositories for CMIP datasets)? Sect. 2.2: this is the only sentence in the manuscript where the methodology is mentioned. Even though the usage of pattern correlation is a quite usual practice for performance scores, it would be good to have a more detailed description of the method, at least of how the averages are weighted. In general, for sake of clarity, I would suggest to rearrange this first part of the manuscript in order to include a Data and Methods section.

Another suggestion is that the author mentions other possible ways to attribute a performance score to models based on its consistency with observational-based measurements. One can refer, for instance, among others, to the Wasserstein distance, as in Breverman et al. 2017, but there are many other examples...

l. 152: I wonder if there is a non-empirical explanation for the choice of weighting the ENSO timescale less than the annual and seasonal timescales in CESM1-LE.

ll. 155-156: this seems to me a pretty strong assumption, because I see no particular reason why the impact on the overall score from internal variability in other models shall be comparable to the one found in CESM1-LE.

ll. 166-168: a way to test the assumption mentioned in my previous comment could be to focus on a few CMIP models providing a reasonably large ensemble against the CESM1-LE. Would that be feasible?

Sect. 4.0: at this point, the author starts to describe the main results of the analysis. I am a bit puzzled, though, by the fact that no convincing discussion has been provided on the choice of the variables. While for the energy budget and water cycle realm it is clear to me that the author follows from the expert consensus outlined in Burrows et al. 2018, the variables for the dynamical regime seem to me not supported by sufficient argumentation. For instance, why is the eddy geopotential height preferred to the potential vorticity in the free troposphere? If the idea is to meet the experts' needs for key metrics, why not additionally considering the zonal mean wind or the potential vorticity at specific isobaric levels? These variables are fundamental for studies of the atmospheric dynamics, even though they have not been addressed in the paper by Burrows et al. 2018 or, if they are considered, they do not reach a (very) high consensus about their relevance.

ll. 266-268: stated like this, it seems to me more suggesting that only the central tercile is actually closing up to observations across the CMIP generations. . .

ll. 317-319: I wonder if the author might want to comment on why this is the case, and whether this could be really considered as an improvement in the overall performance of the multi-model mean.

l. 325: are these metrics telling something relevant about the behavior of subset of CMIP6 models with high sensitivity. Can something be said about it?

Figure 1: please add in the captions what the blue, red and black meridional sections displayed next to each map describe.

Technical corrections

l. 36: replace “increasing” by “increasingly”.

l. 217: replace “import” with “important”.

l. 223: Replace “Select” with “Selected”.

ll. 239-241 (and elsewhere): I think that it is sufficient to describe the layout of similar figures only once, when introducing Figure 6 and its panels. Considering removing the introductory sentence in this paragraph and in the successive ones.

References

Braverman, A., Chatterjee, S., Heyman, M., and Cressie, N.: Probabilistic evaluation of competing climate models, Adv. Stat. Clim. Meteorol. Oceanogr., 3, 93–105, 2017

Greve, P., Gudmundsson, L., and Seneviratne, S. I. Regional scaling of annual mean precipitation and water availability with global temperature change, Earth Syst. Dynam., 9, 227–240, 2018

Hourdin F, Mauritsen T, Gettelman A, et al. The Art and Science of Climate Model Tuning. Bull Am Meteorol Soc 98:589–602, 2017

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., and Eyring, V. A climate model projection weighting scheme accounting for performance and interdependence, Geophys. Res. Lett., 44, 1909–1918, 2017

Lembo, V., Lunkeit, F., and Lucarini, V.: TheDiaTo (v1.0) – a new diagnostic tool for water, energy and entropy budgets in climate models, Geosci. Model Dev., 12, 3805–3834, 2019

Lorenz, R., Herger, N., Sedláček, J., Eyring, V., Fischer, E. M., & Knutti, R. Prospects and caveats of weighting climate models for summer maximum temperature projections over North America. Journal of Geophysical Research: Atmospheres, 123, 4509–

4526, 2018

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-1, 2020.