Comment on gmd-2020-418
Roland Séférian

This work provides a relevant and timely analysis of CMIP5 and CMIP6 models with respect to the representation of circulation in the northern hemisphere. On top this work goes well beyond a routine assessment of global model performance because some of those global models will be used to drive lateral boundary conditions of regional models or to derive climatic impact-drivers at regional/local scale.

While it also shows the general improvement from CMIP5 to CMIP6 in this aspect, this work shine light on some deficiency within the current generation of models.

With the objective the author tries to map model performance on two axes: the complexity and the resolution, both of which are difficult to separate.

In consequence, the high-level picture of the analysis emerging for this work strongly tights to the Table 1 — where we spotted some errors.

For instance, it is indicated that CNRM-CM6-1 and CNRM-CM6-1-HR included online chemistry onboard whereas the description of these model configurations in Voldoire et al. (2019) doesn't support this feature.

Same goes, for IPSL models and for GFDL-CM4 which are characterized as ‘ESMs’ in Table 1 while they do not fit the current understanding of what is an Earth system models (see Jones (2019)). As shown in Séférian et al. (2020), GFDL-CM4 indeed included marine biogeochemistry but only in a stylized manner (reduced complexity marine biogeochemical models). In consequence, there are no biophysical feedbacks represented in GFDL-CM4 whereas it does in GFDL-ESM4.

Regarding the axis of the resolution, providing the nominal resolution would help to compare model between each other. The nominal resolution has been reported by the modelling groups to CMIP6 for each component/realm.
Apart from these remarks/on Table 1, we would like to provide a couple of suggestions that could be useful for this work.

As this work focus on the performance over the historical period, it might be relevant to provide some information on how the model has been tuned/calibrated. At least to know if this set of metrics has been used as a target to prepare the model for CMIP5 and for CMIP6. Such questions tend to emerge now in the literature (see Spafford and MacDougall, in review ni GMDD) because of their implication on routine performance benchmark.

On the other hand, the paper is not clear on the treatment of the model realization. As shown in Olonscheck et al (2020), large ensemble of realization may improve the comparison with the observation. Considering the magnitude of the internal variability of the atmospheric circulation feature, considering additional information on available model member might help. With that said, comparing model with different ensemble size might complicate the picture but discussing the impact of the member on the overall model performance and ranking would be a very valuable outcome of the paper.

We hope that the author will find these comments and suggestions useful/relevant.

References:


