

## Answer from authors to Reviewer 1 comments

# The computational and energy cost of simulation and storage for climate science: lessons from CMIP6

by

Mario C. Acosta, Sergi Palomas, Stella Paronuzzi and Gladys Utrera

We sincerely appreciate the reviewer's insights, which have contributed to improving the clarity of certain sections of the manuscript, refining key concept definitions, and to better specify the context of the work. We are grateful for their valuable input, which has enhanced the overall quality of our work. Below, you will find our responses to the feedback provided in more detail:

### General comments

This is an interesting and perhaps unique accounting of a subset of the models that participated in CMIP6 with a focus on model characteristics related to computing and carbon footprint details. My major comment is that the authors need to be clear that the models described in the paper are, in fact, a subset of the total number of models that participated in CMIP6. This subset is apparently the group of models that participated in the IS-ENES3 project. This needs to be made clear in the abstract and elsewhere.

### Detailed comments:

List of authors: One of the coauthor's names is in the wrong order and misspelled: "Joussame Sylvie" should be "Sylvie Joussaume"

The error has been fixed in the manuscript, and we contacted the journal to update it as well.

Line 4: Here is an example of where the authors need to clarify the scope of the paper. I'd suggest the wording be changed to,

"This paper shows the main results obtained from the collection of performance metrics from 30 models that participated in the IS-ENES3 and represent a subset of the total of 124 CMIP6 models. The document provides..."

We have updated the introduction to present IS-ENES3 consortium and clarify that we only collected the metrics for a subset of CMIP6 experiments.

HOWEVER, it's unclear exactly how many models are actually involved. I got the number 30 from line 42, but there are 32 models listed in Table 2, and 33 listed in Table 3.

The mismatch between the tables has been fixed. There are a total of 33 different experiments. We've substituted the term "model" with "experiment", which is more accurate and does not lead to confusion, given that a coupled experiment simulates multiple models.

We have modified the abstract, introduction and conclusions to state more clearly the institutions and experiments involved in the collection, and how many of them appear in the paper.

Line 41-42: Once again, the authors need to be clear how their models relate to the larger CMIP6 set of models. I recommend the following wording: "In this paper, we present in Sec. 2 the collection of CPMIP metrics from the 30 [or 32, or 33] models that participated in the IS-NES3 project (Joussaume, 2010), out of the total of 124 CMIP6 models, and were used to simulate almost 500,000 years..."

We have checked and fixed mismatches previously appearing in the text, following the reviewer's advice.

Table 2 caption: Pursuant to the comments above, the authors need to be clear how the models they list in this table relate to the total number of CMIP6 models. This would avoid a scientist reading this paper and looking at this table and not seeing their model and wondering why they aren't in the table. I'd recommend the following wording: "List of institutions and the models that provided metrics from their CMIP6 models to IS-NES3, which represents a subset of the total of 124 models in CMIP6. Also listed are HPC platform, and resolution used for the ATM and OCN components." Now, proceed to comment below for a suggestion on the rest of the caption.

We updated the caption for Tab. 2. Now it reads:

List of Institutions and models that provided the metrics from their CMIP6 executions. Also listed are the HPC platform, and resolution used for the atmospheric and oceanic components}

Table 2: In Table 1 the authors define "resol" as number of gridpoints, a singularly unhelpful metric when comparing models. Fortunately, here in Table 2 they relate that metric to the more conventional lat-lon resolution. However, I think in the table caption they should note something like, "Note "resol" in Table 1 is number of gridpoints. Here we define "ATM resol" as the horizontal resolution of the atmospheric model component in degrees of latitude and longitude, and "OCN resol" as the horizontal resolution of the ocean model component in degrees of latitude and longitude".

This paper builds on top of the previous definition for the metrics in Tab. 1 (Balaji. et al 2017), and the Resolution metric was defined as the number of gridpoints. Nevertheless, we appreciate the feedback and we added a note for the reader when changing to degrees as in Tab. 2 caption, as we agree that it is more convenient for model comparisons. In the analysis section for the resolution, we also now use the degrees instead of the number of gridpoints. Note, however, that we the authors have not designed these metrics, and therefore we collected the resolution following its definition, and only added the information in degrees for better understanding by the community.

Line 97: Following the comment above, please add clarification here, something like, "...the categorization of low, medium and high resolutions in terms of latitude-longitude grid spacing. Thus, for the grouping..."

Line 99: And again here, I recommend clarifying the wording as follows: "...as low resolution with roughly 1 degree latitude-longitude grid spacing and up to ..."

This part of the manuscript has been modified for clarity. While we maintain the gridpoint value, as it is required by the metric definition, we show the degrees threshold used for the categorisation:

"Most configurations have been categorised as low resolution and use up to  $2.10E+07$  grid-points in total, or no less than 0.7 degrees latitude-longitude grid spacing for any of the components (see Figure 1 and Table 2). On the other hand, only those experiments with an OCN/ATM resolution under 0.5 degree are treated as medium-high resolution configurations (see Figure 2).

Line 115: Please define "complexity". Does this mean number of components (e.g. atmosphere, ocean, sea ice, land ice, biogeochemistry, ocean ecosystem, cloud-aerosol interaction, etc.), number of parameterizations per component, or what? This needs to be defined up front in order to make sense of the subsequent discussion.

We added an introductory paragraph at the beginning of the Complexity analysis section (3.2): The complexity of a coupled model, as defined in Table 1, is the number of prognostic variables among all components. Here, "prognostic" refers to variables that the model directly predicts, such as temperature, atmospheric humidity, salinity, etc. In other words, variables that can be obtained directly as outcomes of the model.

Furthermore, we have also improved the definitions of other metrics that were not properly introduced before like the Complexity (section 3.2) Coupling cost (section 3.5) and Memory bloat (section 3.7)