

## ***Interactive comment on “A New End-to-End Workflow for the Community Earth System Model (version 2.0) for CMIP6” by Sheri Mickelson et al.***

### **Anonymous Referee #1**

Received and published: 24 July 2020

This manuscript uses CMIP5 and 6 as a case study to show the improvements made to the workflow in order to support CMIP6 at NCAR/CESM. While this work is appreciated and does do a decent job of comparing the workflow from the CMIP5 era to the present, the information presented in the manuscript needs to be more efficiently conveyed so it's useful to the community as a whole. Information on CMIP5 workflow is not very elaborate to get an in-depth understanding and appreciation for the CMIP6 workflow efforts. Some parts of the manuscript can be much more than just an "internal documentation". Schematics like Figure 1 can be vastly improved. Information in the manuscript should serve as a motivation point for other labs to consider new workflow models. Several points are highlighted below. Most importantly, human time consumption could also be provided in this manuscript. Secondly, as we move towards

C1

a new computing era, the manuscript should also let the readers know what else is out there in order to develop a new workflow, inspired by this manuscript. A data agnostic model, processing workflow, cloud-optimized workflows, etc should be touched upon at least in the conclusions section.

Overall, thanks for the manuscript and congratulations on publishing CMIP6 data on the ESGF.

Page 1 Ln 3: Statements such as "Many centers were not prepared.." needs to be rephrased to indicate the unexpected increase in complexity of CMIP6. Unless the relevant facts are cited and the data is provided, comments about other centers does not seem appropriate here.

Page 1 Ln 7: It is nice that there is six times improvement. Please verify if the actual data volume is specified in the manuscript for CMIP5 and CMIP6.

Page 1:Ln 20: For a diverse audience to follow, please indicate what is meant by post-processed data when it's mentioned for the first time.

Page 1. Ln 24. What is the factual evidence to show CESM ran relatively quickly compared to the other climate models? Are you referring to models from NCAR or other modeling centers? Is this number from a CPMIP computational metrics calculation? If there is no factual evidence or appropriate citation, this statement does not seem to be appropriate.

Page 2: Ln 1: What are the software inefficiencies? Workflow development tools? Please expand on this. Software seems very generic.

Page 2: Ln 44, The line that begins with "For CMIP5..", please break this sentence into two or three and avoid using "it.." several times in the same sentence, for clarity.

Page 2: Ln 45: What is different in CMIP6 versus CMIP5 in terms if requiring expert knowledge to ensure data met the correct standards? It does require expert knowledge in order to verify the correctness of scientific model output regardless for the phase

C2

of the CMIPs. If this is incorrect, please clarify and rephrase the sentence to avoid confusion.

Page 2 Ln 46: How is the standardized data verified? What is the Quality Assurance process in the workflow? What is meant by "standardized data"? Please spell out the conventions to be adhered to.

Page 2, Ln 51: Where are the simulations run? Information regarding the computational environment is completely missing. What is the name of the compute system? Is/Was there a batch workflow, job scheduler etc, etc.

Page 3. Ln 67: "The publication of CMIP5 data contributions to the ESGF was also a bottle neck within the data workflow" - This line needs more clarity to indicate what exactly is referred to as the bottle neck. Is ESGF the bottleneck? The tone of the sentence could be more constructive if a community developed federated framework is being criticized. Just as an example, thought not complete. Though there were performance issues with respect to the data publication onto the Earth System Grid Federation for a few reasons, the performance increased phenomenally due to .....

Page 4, Ln 91 TYPO: task based parallelism

Page 4, Ln 102: Try and avoid starting the sentence with "Because" .

Page 4, Ln 110: How dependent is the diagnostics framework on the supported languages you've described? What is the potential to expand the supported languages to say Python, Ruby, etc ?

Page 6, Ln 113: There is a mention of "this work". Please clarify what work this entails as part of this sentence for clarification. There is also a TYPO in "Specifically ..". Consider changing resulting to "results in" as you see fit.

Page 6, Ln 125: Was this a total re-write of the post-processing framework since CMIP5? Page 7: Figure 5- Y axis units missing. Please check all figures as well. Figure 5 caption indicates "46 seconds". Is this in sync with what is shown in the actual

C3

figure?please verify.

Page 7 Ln 122. Figure 5 simply has info on pyAverager. If there is a single image that cross-compares the speedup time for the two tools,that'd be effective and in sync with the text in Ln 122.

Page 7, Ln 140. Paraphrase this sentence, especially "without any modifications". E.g. ..to allow seamless multi-model comparison based on uniform data standards to avoid less rewriting and error-prone transformations while doing so.. etc.

Page 7, Ln 142. "required" is a strong word here and not quite accurate. please paraphrase this. Not all modeling centers used CMOR even for CMIP5.

Page 8, Ln 154. How are changes in dreqPy incorporated in the workflow? Was there a fixed version? How were corrections in the requirements considered and incorporated?

Page 9 Ln 158. CMOR may also have a Python interface, please double check and then change this sentence as needed.

Page 10, Section 2.4. There is no mention of Data Quality Assurance which is extremely important though parts of the workflow may be automated. Please indicate the steps taken to quality control datasets. PrePARE could be accounted for metadata QA, but not quite for data and I am curious how that was incorporated. PrePARE also comes at a later stage when the heavy lifting of data processing and prep is almost complete. So, a bug revealed at a later stage may have its own cons. There was also a similar CMOR checker in some form available for CMIP5, though for CMIP6 it was more robust.

Page 10, Ln 175: Not sure if it's a typo - under-development versus under-developed. This sentence needs to revised either way. A constructive tone would be great.

Page 10, Ln 181: Please cite CDNOT paper that was recently submitted if you haven't already. Ruth et al. 2020

C4

Page 10, Ln 183: Consider changing "harden" to strengthen or similar.

Page 10, Section 2.4. Versioning and ESDOC are two important components in data publication and the ESGF. These are not touched upon and would add immense value to the manuscript to include the process for these.

Page 10. A schematic for Cylc, e.g. Cylc dependency graph/dashboard would add value to the manuscript. Page 10. Ln 201, What is the internal DB implementation? How easy or difficult was it to get started with Cylc and is Cylc also used in other domains? Does the user have the ability to monitor the processes via Cylc and re-submit a job if needed?

Page 11 Ln 207: Sample configuration files from Cylc would be helpful. A section to explain how reproducibility is achieved in the workflow with a schematic and a case study– would be helpful.

Page 11, Ln 212: What are the setup steps? Is there an example of a definition file in the github repository references?

Page 11, Ln 218. How is troubleshooting and monitoring happening with Cylc and your workflow? Who manages that?

Page 11. Ln 225-227. This is nice. Was data publishing part of the automated pipeline? Please explain. In a fully automated workflow, what were the testing strategies, version control mechanisms, provenance capture mechanisms, etc? There is little mention about a couple of things, but more the better to make the manuscript stronger and reachable.

Page 11, Section 4, Ln 229: Again, please be constructive. Be specific as to what experiments you're referring to, what model, what modeling centre.

Page 11, Ln 235: How was information automatically harvested from the CESM experiment?

C5

Page 12, Ln 239: Please expand on what configurations and timing files mean here.

Page 12: Ln 241: What is the "code" that's referred to here? Please elaborate

Page 12: Ln 244: Are there sample analysis figures that could be provided? Are there any collaborative work on the diagnostic package that needs to be acknowledged? Is there scope for collaborative efforts since some of the diagnostic packages can be helpful to the community as a whole.

Page 12: Nice- HTML docs for viewing results.

Page 12, Ln 250 This seems to speak about monitoring capabilities, although the information provided here is not very useful for readers to learn from this work.

Page 12, Section 5. Line 260. By traditional tools, are serial tools referred to ? Since this seems to be the major difference in your workflow paradigm since CMIP5?

Line 263: Are there citations to the datasets referred to here? Any information on how CMIP6 citations were processed for CESM.

Conclusion should have future work as well, because what is considered traditional today will not hold good for the years to come. Lessons learned from CMIP6 exercise needs to be magnified in order to move towards cloud-optimized workflows and flexible APIs. The manuscript should give some food for thought to the readers. Examples to show if (if not) CMIP may be the only style of experiments that the workflow processes should be clarified.

---

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2020-133>, 2020.

C6