RESPONSE TO REVIEWER RC1

Anonymous Referee #1
Received and published: 13 January 2020

Dear Referee #1

Thank you for taking the time to read through our manuscript and provide comments and suggestions. Below we highlight our responses.

Overview:
The manuscript provides an update to the Obs4MIPs effort detailing progress in recent years and demonstrating the project's role in the most recent generation of the Coupled Model Intercomparison Project (CMIP6). The key points of this manuscript are:
1.) Obs4MIPs is undergoing expansion from 80 datasets to possibly close to 200,
2.) improved characterization of data quality, coverage, and suitability for datasets that are part of the project; and 3.) enhanced support for supplemental information and code to aid the end user. Notably, there is a fundamental shift in the type and quantity of datasets included in the effort that favors a more diverse collection of datasets that sometimes overlap each other. This more inclusive approach, however, is in potential conflict with prioritizing "ease of use" to the end user. Overall, this is an important manuscript and project that bridges the climate observational and modeling communities. I would recommend publication pending revisions that address the concerns
Below.

Major comments:

1. On Page 5, Lines 16-18, one of the stated goals of the project is that Obs4MIPs is to assist a growing community of scientists "without an expert's understanding of either the observations being employed or the climate models themselves." There are several notable advances in this effort to help a user better understand the observational dataset, such as the dataset indicators and more supplemental information. How does Obs4MIPs help users understand the climate models themselves? This point wasn't clear to me. Also, the stated goal of helping the end user is in potential conflict with other aims of the project. (See Major Comment #2.) Perhaps there are some use cases or metrics that could be provided as evidence that the Obs4MIPs effort is having in the community (publication counts referencing data, etc.)

A primary objective and advantage of obs4MIPs is to make observational products more readily available for users to compare with (CMIP) model output. Obs4MIPs doesn't seek to improve observed datasets, but rather better organize/format/disseminate the data that has been found to be useful for comparison to climate model output. Because it is the same data (whether or not it has been processed for obs4MIPs), it does not help the user "understand the climate models" any better than if the data been acquired from a source other than obs4MIPs. There is a very large body of published literature aimed at utilizing observations to assess the performance of climate models, and the answer to how this helps users understand the climate models really depends on the scientific question/process being addressed.

Frequently, model evaluation studies are able to describe the consistency between models and observations (i.e., identify errors), only rarely are they able to identify the root causes of those errors. In this paper we give two examples (Figs. 3 and 4) that typify how obs4MIPs (or any other data) is used to compare with model output.

2. The discussion on data redundancy, present on Page 18 Lines 22-27, deserves much more attention. The shift from a high degree of initial oversight toward a more inclusive approach to accepting datasets marks a fundamental philosophical shift in the project. For example, there are now multiple references to different SST products (Page 7 Line 22, Page 9 Line 24, Page 12 Line 18). I personally agree that a diversity of datasets is ultimately a virtue, but it is potentially in conflict with the goal of assisting the "non-expert" user. How can a non-expert choose which product is best for given application when multiple datasets of the same product exist? Are dataset indicators Enough?

We are glad that the reviewer agrees with the approach to diversify product availability. An important question has been raised here - "what is the best dataset to use for model evaluation?". As the reviewer has alluded to, one of the main objectives for developing the Dataset Indicators was to help guide the non-expert in their choice(s) of observation dataset(s) to use. These objectives and the sentiments below are discussed in Section 4c.

Many discussions among data experts and model analysts (including a targeted workshop; Ferraro et al., 2015) have led to the obs4MIPs dataset indicators which provide useful and unique information. How scientists address the issue of multiple datasets is application dependent. In practice, we typically do not have sufficient error-characterization information to objectively rank the suitability of different datasets for model evaluation. However, in absence of this, analysts now routinely use multiple datasets in their research to evaluate if their underlying conclusions might be sensitive to which reference dataset they use. This does not explicitly get at the issue of possible source dependency between different data products, however, it is often the best that can be done. In summary, obs4MIPs is now advancing to enable users to include multiple observation products (if available) in their research rather than attempting to identify which (single) product should be used. The Dataset Indicators was a key development to facilitate the virtue and limit the shortcomings of including multiple observation products.

3. The relationship with ana4MIPs and other similar efforts could be spelled out earlier in the manuscript. It is not until the last section that the reader discovers that the "ana4MIPs" effort is static and the introduction of CREATE-IP on Page 10 Line 11 is lacking context. In the community, Obs4MIPs and ana4MIPs are mentioned in the same breath. Important distinctions between the two are necessary.

Thank you for pointing this out.
We now mention the analogous effort of CREATE-IP (initially conceived as ana4MIPs) in the Introduction portion of the manuscript.

4. Can more be said about the different spatial coverages of the Obs4MIPs data? How are gridded and point measurements handled differently by Obs4MIPs? Are datasets derived from floats, such as ARGO, compatible with the Obs4MIPs? There are references in the last paragraph of the manuscript, but these issues need to be addressed

Earlier.

These are good questions, thank you for them.   We have added a paragraph at the end of Section 2 to address them.

Minor Comments:

Page 4, lines 17-19: These are necessary capabilities, but it is not a complete list.
AGREED, modified wording.

Page 5, line 12: Consider a sentence or two expanding about how observations help identify and correct model shortcomings. Observations allow validation of the models climatological mean state, annual cycle, and variability across timescales helping to assess model fidelity among many different Earth System Processes.
DONE

Page 7, lines 16-18: What was the criteria used to determine what datasets were "appropriated for climate model evaluations?" Was this based on demand from the modeling community, external users, assessment by the dataset creators, or some Combination?
This has been clarified in the text, noting that the selection of the initial datasets was based on dataset maturity and long-standing community use for model evaluation.

Page 8, line 15: "apprised" is an odd word choice. This implies that the Obs4MIP effort was taking place without a dialogue with the modeling community.
FIXED - changed to engaged

Page 9, lines 13-17: Rather than a standalone paragraph, consider joining these lines with the previous paragraph.
DONE

Page 12, lines 7-8: This sentence is awkward. How do the improvements in CMIP output specification impact the Obs4MIPs holdings?
FIXED
Section 4b: As data standards/specifications evolve, what happens to existing datasets? Are they left as-is or converted?

This is a good question. We have clarified that no significant changes to the obs4MIPs data specifications are expected until a next generation of CMIP is designed, probably not for at least 5 years. Nevertheless, we are still dealing with existing datasets, and their possible updates. New search facets are not difficult to deal with but altered ones are. After considerable experimentation, we have landed on a compromise as seen during a search with "CMIP5-era" and "CMIP6-era" specific facets. This enables us to deal with the changes/improvements to the data conventions.

Page 15, Lines 21-22: Can more be said about the indicator assignment process? Does it rely mainly on self-reporting? How is the review conducted?

The Dataset Indicators is new and yet to be fully exercised beyond the initial settings by the obs4MIPs Task team, and as
yet there has not been a case needing revision or adjudication with the dataset provider. The relevant paragraph updated to
reflecting the current plans and experience with the Dataset Indicators.

Page 19, Lines 7-10: How will the assignment of data DOIs work in practice? Would
an end user be citing both a data DOI and a scientific paper describing the dataset?
As noted, this problem extends to CMIP6 data and beyond, but it poses an interesting
issue for the peer-review publication process.

This reviewer seems to be well aware of data challenges associated with CMIP6 which in many respects parallel those of obs4MIPs. His/her familiarity with the topic has led to constructive comments that have helped guide the improvement of our manuscript - thank you!   With regards to DOI's, as noted in the manuscript there are important issues to solve.   For now, DOI's are being coined in an ad-hoc way, i.e., they are not applied systematically. The reviewer raises a good point about how best to handle citations of data in scientific papers. An additional possibility being considered would be to enable DOI for the "tech notes".   Leveraging an on-line only data journal is one possibility but more work is required.