



Climate Modelling User Group

Note on Requirements on CCI datasets for inclusion in Obs4MIPs

Centres providing input: MOHC, MPI-M, ECMWF, MétéoFrance

Version nr.	Date	Status
0.1	30 April 14	Initial Draft by RS for comment/additions
0.2	05 May 14	Added details based on obs4MIP workshop by AL
0.3	06 May 14	Added Table in Annex from PPM



METEO FRANCE
Toujours un temps d'avance



Max-Planck-Institut
für Meteorologie



Note on Requirements on CCI datasets for inclusion in Obs4MIPs

This document gives the requirements that CCI datasets need to comply to in order to be included in the Obs4MIPs database which is used by the CMIP modelling community for comparing satellite observations with climate model predictions. A list of requirements for the data producers is given below which CCI teams will need to follow. All the information is provided on the Obs4MIPs site at: https://www.earthsystemcog.org/projects/obs4mips/how_to_contribute. A paper describing the Obs4MIPs objectives is available on-line (Teixeira *et. al.* 2014). A table giving the requirements for putting the data is given in Annex A. It should be noted however that proposal are underway to update these requirements for CMIP6 and so this document is subject to change when the new proposals are ratified by the WCRP Data Advisory Council Obs4MIPs task team.

Some comments specific to the ESA CCI data are included in italics below each item.

1. Criteria for Datasets to be included in Obs4MIPs

Observational datasets for Obs4MIPs must fulfil the following criteria:

- Has clear traceability from level 1 measured variables to retrieved variables in level 3 or 4 dataset
- be based on data that has a history of peer reviewed publications,
- reside in a long term and maintained archive,
- span a time period long enough to be of use for model comparison (3 years is a useful minimum although in some circumstances shorter data records may be considered),
- match a model variable in the CMIP5 protocol
- include an estimate of the uncertainty for each variable verified by validation of the retrieved variables

Most CCI datasets which are relevant to Obs4MIPs should conform to the above although it will take time for a history of peer reviewed publications to be available.

2. Input Dataset Gridding

The datasets for consideration for Obs4MIPs should be Level 3 (single sensor) or level 4 (multiple sensors) datasets which have been transformed on to a 1 degree grid square through averaging and/or interpolation and then averaged over 1 month. A gridding on a higher resolution (e.g. 0.25 degree) is in general possible and should be considered. For CMIP6, it is likely that also dataset with higher temporal resolution (e.g. daily) will be desired. For each grid square the fields should be complete (i.e. no data voids) and consideration should be given to ensuring the variables are still conserved in the regridding. It is assumed only observational data (i.e. no model analyses) are included. The associated uncertainties also need to be provided on the same grid and care has to be taken to derive these from the level 2 single field of view observations. Note that datasets within obs4MIP as of May 2014 contain only information about the variability at each grid cells (stdv., number of samples), but not the real observation uncertainties at the grid scale. Any biases in the original observations will propagate through to the gridded data but random errors will be reduced, hence averaging of uncertainties may not be appropriate. More details on the requirements are given here:

https://www.earthsystemcog.org/site_media/projects/obs4mips/obs4MIPsDatasetRequirements_v1.2.pdf

For some of the ECV's, value added L4 datasets could be useful. An example would be for instance land cover fraction information on a 0.25 degree grid, instead of the high resolution land cover dataset.

There are several issues to bear in mind when regridding data from level 2 to level 3 especially when dealing with uncertainties. These include maintaining consistency between variables after the regridding, consistently dealing with coastal areas correctly and how to fill data voids. It would be a



good idea for CCI teams to share experiences on their regridding methodology to avoid and/or document biases resulting from temporal and spatial aggregation.

3. Data Format

The file must be written in NetCDF version 3 and must follow the standard NetCDF Climate and Forecast (CF) Metadata convention <http://cf-pcmdi.llnl.gov/>. In the future, it is likely that netCDF4 format will become the standard for obs4MIP.

The output file must pass a CF compliance check. You can find a checker at:

<http://puma.nerc.ac.uk/cgi-bin/cf-checker.pl>

Choose the latest CF version when submitting the file for checking.

Each output file must contain a time series of ONLY ONE physical variable (e.g sea surface temperature, specific humidity). If the entire time series can be stored in less than 2GB, it must be stored in a SINGLE file. If it requires more than 2GB, it should be split into the minimum number of files required, with the size of each file being less than 2GB. Each file should contain a contiguous time series of complete data grid blocks. Each file must contain all of the required metadata applicable to the data subset contained in the file. Some software is provided at:

<http://cmip-pcmdi.llnl.gov/cmip5/obs4cmip5.html> to write datasets in the compliant format.

Each physical variable and coordinate variable must use the specified output/coordinate variable name given in the CMIP5 Requested Output list (standard_output.xls) or should be easily comparable to some derived values from these variables. Note however, that for CMIP6 a new output table will be defined and it is open to the CCI teams to suggest also further model output fields which are best comparable to CCI product definitions. For example, the latitude output name must be “lat”, and the air temperature output variable name must be “ta”.

Sharing of experience within the CCI on writing the compliant format datasets from the CCI data would be worthwhile. Feedback on any problems should be given to the Obs4MIPs team.

4. Documentation

A short technical note (5 pages max) must be provided with the dataset that conforms to the obs4MIPs technical note template:

https://earthsystemcog.org/site_media/projects/obs4mips/Obs4MIPsTechnicalNoteGuidancev3.pdf

It should be written bearing in mind the reader will not be familiar with satellite datasets. One important point to bear in mind is that there may be other datasets of the same variable available on the Obs4MIPs site and so the note should make it clear what are the advantages of the CCI datasets with respect to previous datasets already available through Obs4MIPs.

These technical notes are valuable in their own right to promote the CCI datasets and should be part of the phase 2 deliverables. CCI teams should make them available on their web sites.

5. Representation of uncertainties

The CCI teams have developed different approaches to represent product uncertainties. In the current obs4MIP datasets, the product uncertainties is just represented by the variance and number of samples for each grid cells.

For CMIP6, the usage of observational ensembles was discussed as an easy way to represent uncertainty information for the user.

6. Process for submission of datasets

There is a proposal form for dataset owners to complete here:

https://earthsystemcog.org/site_media/projects/obs4mips/obs4MIPs.DataSet.Form.v0.1.pdf

and submit by email to the Obs4MIPs team: <https://www.earthsystemcog.org/projects/obs4mips/contactus/>

The optimal approach to integrate datasets into obs4MIP is to go via a contact of the Earth System grid node provider (e.g. DKRZ, BADC)



The CCI project should keep a record of which datasets have been submitted to Obs4MIPs and which are available there.

7. References

REF-01 Teixeira, J., D. Waliser, R. Ferraro, P. Gleckler, T. Lee and G. Potter, 2014: Satellite Observations for CMIP5: The Genesis of Obs4MIPs. Bull. Amer. Meteor. Soc., early online release <http://dx.doi.org/10.1175/BAMS-D-12-00204.1>

URL-01 https://www.earthsystemcog.org/projects/obs4mips/how_to_contribute

URL-02 https://www.earthsystemcog.org/site_media/projects/obs4mips/obs4MIPsDatasetRequirements_v1.2.pdf

URL-03 https://www.earthsystemcog.org/site_media/projects/obs4mips/obs4MIPsGlobalAttributesRequirements_v1.1.pdf

URL-04 <http://cmip-pcmdi.llnl.gov/cmip5/obs4cmip5.html>

URL-05 https://www.earthsystemcog.org/site_media/projects/obs4mips/Obs4MIPsTechnicalNoteGuidancev3.pdf

URL-06 https://www.earthsystemcog.org/site_media/projects/obs4mips/obs4MIPs.DataSet.Form.v0.1.pdf

URL-07 <http://puma.nerc.ac.uk/cgi-bin/cf-checker.pl>



Annex-A Table of Requirements for submitting CCI datasets to Obs4MIPS

Req #	Ref	Requirements
RQ-01	URL-02	<i>Variable</i> must be on a <i>lat/long grid</i> currently 1deg but 0.25deg is possible for CMIP6 and averaged over a <i>fixed time period</i> currently 1 month but daily will be possible for CMIP6.
RQ-02	URL-01	The datasets for consideration for Obs4MIPs should be <i>Level 3 (single sensor) or level 4 (multiple sensors)</i> datasets which have been transformed on a regular grid square through averaging and/or interpolation and then averaged over a defined time period. For each grid square the fields should be complete (i.e. no data voids) and consideration should be given to ensuring the variables are still conserved in the regridding. It is assumed only observational data (i.e. no model analyses) are included.
RQ-03	URL-01	Has clear <i>traceability</i> from level 1 measured variables to retrieved variables in level 3 or 4 dataset
RQ-04	URL-01	<i>Uncertainties</i> also need to be provided on the same grid and care has to be taken to derive these from the level 2 single field of view observations. Any biases in the original observations will propagate through to the gridded data but random errors will be reduced, hence averaging of uncertainties may not be appropriate. The values should also be validated in some way. Note however, that the current obs4MIP datasets contain basically the variance and number of samples as “uncertainty” information, which is different from a real error estimate.
RQ-05	URL-03	Match a model variable which is in the CMIP5 protocol is desired. As an alternative, the observed variable should be easily derivable from CMIP output. Note also that it is possible to request additional model output fields for CMIP6 which are better comparable to observations.
RQ-06	URL-01	Span a time period long enough to be of use for model comparison (3 years is a useful minimum although in some circumstances shorter data records may be considered)
RQ-07	URL-01	Be based on data that has a history of <i>peer reviewed</i> publications,
RQ-08	URL-01	Reside in a long term and maintained archive.
RQ-09	URL-05	Documentation. A short technical note (5 pages max) must be provided with the dataset. that conforms to the template in the ref. It should be written bearing in mind the reader will not be familiar with satellite datasets. One important point to bear in mind is that there may be other datasets of the <u>same</u> variable available on the Obs4MIPs site and so the note should make it clear what are the advantages of the CCI datasets with respect to previous datasets already available through Obs4MIPs
RQ-10	URL-02 URL-04	The output file must be written in <i>NetCDF version 3</i> , but in the future, it is likely that NetCDF4 format will become the standard. The data output must follow the standard NetCDF Climate and Forecast (CF) Metadata convention. See CF compliance checker at
RQ-11	URL-02	Each output file must contain a time series of only one physical variable (e.g sea surface temperature, specific humidity)
RQ-12	URL-02	If the entire time series can be stored in less than 2GB, it must be stored in a single file. If it requires more than 2GB, it should be split into the minimum number of files required, with the size of each file being less than 2GB. Each file should contain a contiguous time series of complete data grid blocks. Each file must contain all of the required metadata applicable to the data subset contained in the file.
RQ-13	URL-03	Each physical variable and coordinate variable must use the specified



		output/coordinate variable name given in the CMIP5 Requested Output list (standard_output.xls). For example, the latitude output name must be “lat”, and the air temperature output variable name must be “ta”.
--	--	---

End of Report