Climate Modelling User Group [CMUG]

Deliverable 5.7f

Obs4MIPs User requirements and gap analysis report

Centres providing input: Met Office, STFC-UKRI, University of Reading

Version nr.	Date	Status
0.1	19 November 2024	Interview analysis added
0.2	04 December 2024	Gap analysis added
0.3	06 December 2024	Uncertainty requirements added
1.0	09 December 2024	Submitted to ESA
1.1	10 December 2024	Updated and submitted to ESA
1.2	15 December 2024	Further updates and edits

Deliverable 5.7f

Obs4MIPs user requirements and gap analysis report

Contents

1. Purpose and scope

This report was commissioned by ESA CCI to inform the future evolution of Obs4MIPs (Teixeira et al., 2014). Obs4MIPs is a database containing observational Climate Data Record (CDR) data in a format designed to facilitate easy intercomparison with model output. This activity was initially driven by the Coupled Model Intercomparison Project (CMIP) as part of the World Climate Research Programme (WCRP) (Waliser et al., 2020).

This report consolidates the information gathered through interviews with 36 climate scientists carried out by CMUG during July-November 2024, it also includes a section on gap analysis from Alison Waterfall (CEDA) and recommendations on the treatment of uncertainty within obs4MIPS from Claire Bulgin (University of Reading).

2. Introduction

The Observations for Model Intercomparisons Projects (Obs4MIPs) is a database containing climate observations consistent with the model output from the World Climate Research Programme's (WCRP) Coupled Model Intercomparison Project (CMIP). The aim of obs4MIPs is to provide observational data to the climate science community, with a focus on the applications most relevant to the Coupled Model Intercomparison Project (CMIP), such as model evaluation and development (Gleckler et al., 2011; Teixeira et al., 2014; Ferraro et al., 2015; Waliser et al., 2020). The data are in standardised format and each dataset is accompanied by consistent documentation in the form of a 3-page Tech Note.

In recent years the process for updating datasets on obs4MIPs has become extremely slow and also it has become difficult for data providers to add new data. As a result, use by the modelling community has decreased. However, the original concept has significant value and this report provides a number of recommendations to reinvigorate obs4MIPS activity so that it can take its place in the ecosystem of CDRs as an important observation data base for climate data users.

The analysis of the interviews with climate scientists is summarised in Section 3, the gap analysis is presented in Section 4 and recommendations for the treatment of observation uncertainty are provided in Section 5. Meta data on the interviewees is provided in Annex 1 and a list of datasets included in obs4MIPS at the time of writing (November 2024) is given in Annex 2. Annex 3 lists the interview responses gathered on barriers to the use of obs4MIPs.

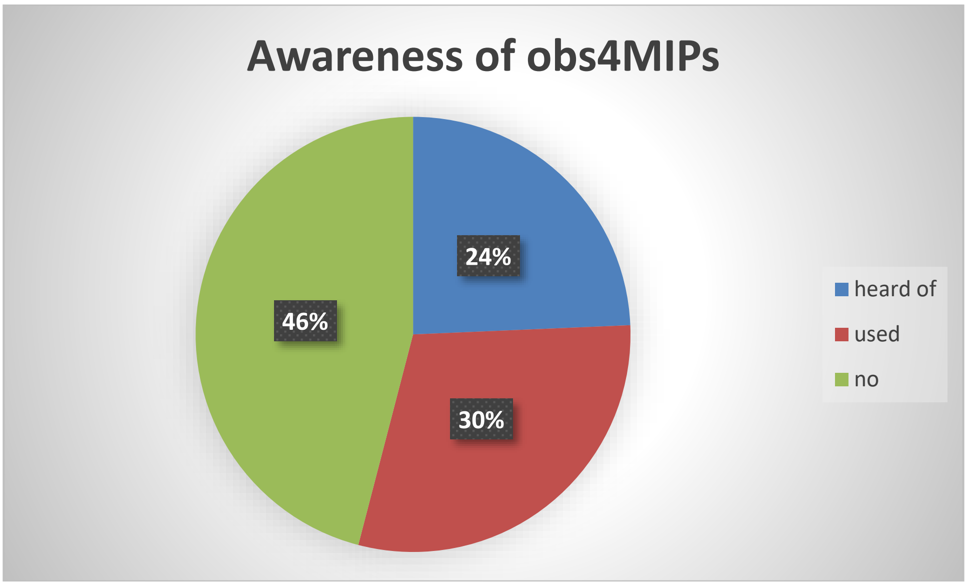
Those interviewed were from a range of 17 institutions with a wide range of applications for climate observations (listed in the Annex 1). A cross-section of geographical areas was sampled with Europe, South America, North America and Africa represented, but due to the Eurocentric nature of the interviewers and ESA CCI most interviewees are from Europe (89%), most of those from the UK (64%), and most of those from the UK Met Office (53%).

3. Interview feedback

3.1 Awareness

As summarised in Annex 1 the interview participants came from a wide range of climate disciplines however knowledge of and use of obs4MIPs was not widespread, of 36 interviewees 19 had heard of obs4MIPs and 12 of those had used it, which means 17 of the people interviewed had no knowledge of it before being interviewed.

A number of interviewees (9) used ESMValTool and were aware that this allowed easy access obs4MIPs data, the standardisation between the two was listed as a significant bonus for both.



Therefore, the first barrier to obs4MIPs is a simple issue of publicity. If more people are aware it raises the likelihood that it will be used. Recommendations by the CMIP IPO that obs4MIPs should be used for evaluation of climate models during CMIP7 will certainly help with publicity. And given the recent resurgence in development of obs4MIPs publicity will be needed

This publicity should be clear in stating what the scope of obs4MIPs is, some interview participants (4) mentioned that they thought only atmospheric data were included, leading them to dismiss it as a source of observational data. A clear statement of what obs4MIPs is trying to achieve is essential, the scope needs to be carefully defined and clearly communicated.

Recommendation 1: Obs4MIPs to be publicised more widely with a clear statement of scope especially where the variable format is defined by the project ie:

- **which variables are currently included and any restrictions for future variables to be added**
- **resolution (spatial and temporal)**
- **length of timeseries**
- **level of quality control carried out by obs4MIPs**

Recommendation 1.1: One way in which this could be facilitated would be a regular (quarterly) obs4MIPs newsletter which contains:

- **a list of all current datasets available**
- **any recent updates to these**
- **plans for the next period**

This newsletter could be subscribed to, and all major modelling centres could be contacted through CORDEX and CMIP IPOs to publicise when this is set up.

Recommendation 1.2: the top level landing page when searching on the internet for obs4MIPs should lead to a list of content, clearly outlined, subdivided by climate domain (atmosphere, cryosphere etc) and with links direct to the technical notes for each dataset

Recommendation 1.3: this top level landing page needs to be kept up to date frequently. CMUG has found through the interviews that out of date information is one of the main barriers to obs4MIPs use: either out of date descriptions or out of date datasets

3.2 Documentation and metadata

Suitable levels of documentation were important to all interviewees and most (19) found the level provided by the existing obs4MIPs tech notes to be suitable for a first introduction to the datasets. They indicated that they would use tech notes to inform their decision on whether or not to go ahead with investigating use of a dataset. Only one participant disliked the tech note format stating that it did not give enough information, they wanted details on all the data processing which has gone into each dataset and any quality analysis that has been carried out.

Making the tech notes available in the landing/contents page for obs4MIPs (see recommendation 1.2) was highly popular.

Information which should (continue to) be included in the technote:

- provision of a DOI by which to reference the dataset
- clear description of the variable and how it differs from the CMIP variable, where the same naming convention has been used. There was in general a dislike of having the same variable name in Obs4MIPs and CMIP if those variables represent slightly different quantities. This was seen as misleading.
- Top level information on consistency between datasets. When climate processes are being investigated it is very important that datasets of related variables are consistent, or that any inconsistencies are well documented.

Many interviewees (13) would like to see links within the tech notes to further information such as

- quick start guide
- Product user guide (PUG)
- Quality assessments
- Data processing details such as
    - retrieval algorithm
    - sensor information

- gridding processes
- gap filling approach
- homogenisation process
- any other assumptions made during the production of the dataset

- Detailed qualitative uncertainty information including the methods used to calculate and propagate the uncertainty during all parts of the processing
- Published papers on dataset production
- Published papers using the data
- Suitability of the data for different applications
- Limitations of the data (eg if not available when cloud is present or over high altitudes
- Comparison of the different datasets on obs4MIPs for the same variable with analysis of the quality of each and guidance on which is best to use for which application
- Information on how to regrid the dataset to a different resolution and how to sensibly propagate uncertainty information during regridding
- Guidance on which CMIP variables the dataset can sensibly be compared to

Live documents which are updated with latest analysis were popular. A number of good practice examples on data listing were cited by the interviewees such as the LST_cci table listing bugs and latest quality analysis in each of their datasets (Figure 1) and the Copernicus data table for CDS[1]. Again, linking to recommendation 1.3, these would need to be regularly kept up to date.

Use the drop down menus to filter by entry, e.g, to find all reports associated with a particular product, select that product from the drop down list in 'Product Affected'

Report	Date Post	Latest Upda	Product Affected	Versio	Report Ty	Details	Comments/Next Steps	Entered by
1	28-Sep-23	28-Sep-23	L3S-LST-IRMGP_-0.05deg	2.00	Bug report	All the MODIS and AVHRR data used in this dataset do not have consistent processing compared to the other data used in this product.	To be corrected in the Cycle 1.5 update, scheduled for completion by June 2024.	Lizzie Good
2	10-Dec-23		L3C-LST-SSMI13-0.125deg	4.11	Other	Inspection of the dataset is showing that the temporal correction (to adjust the LSTs for orbital drift) may not be working as well as the previous	An investigation on this issue is ongoig.	Carlos Jimenez
3	10-Dec-23		L3C-LST-SSMI17-0.125deg	4.11	Other	Inspection of the dataset is showing that the temporal correction (to adjust the LSTs for orbital drift) may not be working as well as the previous	An investigation on this issue is ongoig.	Carlos Jimenez
4	10-Dec-23		L3C-LST-AMSRE-0.125deg	4.11	Other	Inspection of the dataset is showing that the flag to detect convective clouds and the associated anomalously cold LSTs may miss some cases,	The flag derivation is being revisited; improvements are expected for the next release in	Carlos Jimenez
5	10-Dec-23		L3C-LST-AMSR2-0.125deg	4.11	Other	Inspection of the dataset is showing that the flag to detect convective clouds and the associated anomalously cold LSTs may miss some cases,	The flag derivation is being revisited; improvements are expected for the next release in	Carlos Jimenez
6	10-Dec-23		L3C-LST-SSMI13-0.125deg	4.11	Other	Inspection of the dataset is showing that the flag to detect convective clouds and the associated anomalously cold LSTs may miss some cases,	The flag derivation is being revisited; improvements are expected for the next release in	Carlos Jimenez
7	10-Dec-23		L3C-LST-SSMI17-0.125deg	4.11	Other	Inspection of the dataset is showing that the flag to detect convective clouds and the associated anomalously cold LSTs may miss some cases,	The flag derivation is being revisited; improvements are expected for the next release in	Carlos Jimenez
8	12-Nov-23	2-Jul-24	L3S-LST-IRMGP_-0.05deg	2.00	Bug report	There is a gridding issue with this dataset - data is observed to be 'smeared' over several gridboxes.	Project team are investigating - further details to follow.	Karen Veal
9	12-Dec-23	2-Jul-24	L3S-LST-IRMGP_-0.05deg	2.00	Other	There is no inter-calibration between sensors (except between MODIS Aqua/Terra). Dataset	Some intercalibration will be performed for future releases -	Sofia Ermida
10	10-Jan-24	2-Jul-24	L3C-LST-AVHRMA-0.01deg	1.10	New release	New version of Metop-A AVHRR (v1.10) realeased to fix the previous severe cloud contamination in	Now on CEDA	Darren Ghent
11	10-Jan-24	2-Jul-24	L3S-LST-IRCDR_-0.01deg	2.00	Bug report	Issue with some blocks of MODIS pixels that pesistently look 'odd' - e.g. there was a block over	Project team are investigating - further details to follow.	Darren Ghent

Figure 1: Screenshot of the LST_cci updates and issues database which clearly lists any quality issues and latest versions of the LST datasets

[1] https://cds.climate.copernicus.eu/datasets

Conversely, the ERA5 documentation was given as an example of poor practice by some interviewees (2) who stated that it was too complex with too many sub levels making finding the information you are looking for very difficult.

Use of metadata was a popular method to provide information on the datasets due to the common practice of downloading a file and renaming it. If the variable, version number and link to the documentation is in the metadata it protects against loss of vital information and the provenance of the datasets can be retained. However, broken links in the metadata are a risk and these would have to be guaranteed to exist for a reasonable length of time.

There was also a request for metadata to be completely standardised as this causes problems in scripts when there are variations. Also, that the QC flags should be consistent across obs4MIPs, a barrier to using these is having to understand a different complex set for each observation dataset used. If this information can be provided in a machine accessible format it will be most widely used across the community.

Information that was requested in the metadata (in order of popularity):

- Quantitative uncertainty information (see section x) (13)
- QC flags (consistent across obs4MIPs) (11)
- DOI (6)
- Links to documentation (4)
- Version number (3)
- Description of processing (2)
- Error analysis (2)
- Data masks for land, cloud and QC (2)
- Relevant shapefiles (1)
- Clear time step information (1)
- Thresholds outside of which data is not valid (1)

It should be noted that while many interviewees would like additional metadata, these need to remain CF compliant and in CMOR format for compatibility with ESMValTool. Given the large number of current users in this cohort, who access obs4MIPs through ESMValTool, no major change to the file format should be undertaken without consultation with them. In fact stricter compliance with the CF conventions was requested by the interviewees involved in ESMValTool development, and they recommended that this should be checked more closely before a dataset is accepted for inclusion in obs4MIPs (see section X on file format).

Related to all of the above is a really clear statement on

Recommendation 2: Improved documentation and metadata

Recommendation 2.1: Technical notes to continue in current format but to include more links to detailed documentation

Recommendation 2.2: Technical notes to be linked from top level landing page/table of contents for obs4MIPs (see recommendation 1.2)

Recommendation 2.3: Technical notes to be live documents updated as quality issues come to light, or perhaps with link to data quality tables for each version as exemplified by LST_CCI bugs and issues database.

Recommendation 2.4: Standardised metadata and QC flags.

Recommendation 2.5: CF compliance is important, but if more information can be added to the metadata within this format this would be welcome. Minimum top level metadata should include DOI and version number for dataset. Adding grid box by grid box metadata would also be welcome.

Recommendation 2.6:

Recommendation 2.6: Regular literature searches should be carried out by the observation providers or the obs4MIPs team and recent published peer reviewed papers which use the obs4MIPs datasets should be linked to the tech note to allow climate researchers to easily access information on how the data have been used.

Recommendation 2.7: A feedback form on the web page to allow users to raise issues that they encounter around access or data availability. This is only worthwhile if the feedback is read and acted on.

3.3 Access, format and updating

A wide range of preferred access methods were listed by the interview participants. These are shown in Table 1

Table 1: Preferred access method of participants

Access method	Number of interviewees' preference
command line/embedded in scripts eg wget	9
webpage download	7
Cloud based access and processing	3
API	2
CEDA/ESGF/JASMIN	3
Copernicus style interface	1
ftp	1
Database should be machine accessible	1

The method of access was not the major barrier for most scientists, rather the inconsistencies between datasets and the time needed to understand the idiosyncrasies of each one led to barriers, and the large volumes of data were problems, especially for those from smaller institutions or in the global south where connectivity is less reliable. Conversely, when the data are split into smaller chunks (eg, one file per month) this also causes downloading and time issues for those who need to complete dataset. Therefore, a flexible approach is recommended.

Licensing issues for use of data were mentioned by some interviewees (2), if obs4MIPs seeks to actively increase the number of available datasets this should be considered, perhaps to allow credit to be given to the observation provider or other licensing constraints.

Specifically the obs4MIPs grid labels were noted as being particularly hard to understand, a better explanation of this would be welcome.

NetCDF file format, using CMOR3 and CF compliant were in general popular across all participants, Table 2 shows this plus the other file formats mentioned in the interviews. In general the consistency provided by obs4MIPs was very much valued, although this must be policed strictly so that datasets are not accepted if they do not meet the requirements.

As mentioned above 9 of the interviewees use ESMValTool and once an obs4MIPs dataset has been incorporated into ESMValTool this is an extremely easy method to manipulate the data.

File format	Number of interviewees
Netcdf	22
HDF5	2
BUFR WMO format	1
zar	1
grib	2
CF compliant	3
xarray	1
Compatible with iris	1
obs4MIPs format should be consistent with CMIP6	1
Doesn't matter as long as there are data readers to read it	1
Must be consistent	2
Must be compatible with ESMValTool	1
CMOR used	3
ascii	1
Everything in same file (eg uncertainties)	1

Table 2: preferences for file formats among the interviewees

The obs4MIPs data specification document[2] is a thorough description of what is in each of the dataset files, however this is not easy to find, and so this should be linked from each tech note.

A significant barrier to use of obs4MIPs among the interviewees was out of date datasets. The process for updating to the latest version or adding another time period to the data needs to be improved, to allow this to be done very easily. The process by which CEDA updates

[2] https://zenodo.org/records/11500474

datasets regularly, could be used as a guide to best practice. A minimum of annual updates is required by most users interviewed.

As a result of this feedback CMUG recommends a number of tools are made available to smooth the accessing process and that annual updates are required across the obs4MIPs datasets.

Recommendation 3: data access and manipulation tools should be provided for the obs4MIPs datasets.

Recommendation 3.1: All data and metadata formats should be completely consistent across obsMIPs.

Recommendation 3.2: ODS2.5[2] should be linked from each tech note and the link should be updated when the document is updated.

Recommendation 3.3: All live datasets should be updated once per year (preferably at the same time) to include the latest time period. All datasets (static or live) should have the ability to be updated quickly if bugs or errors are fixed or other improvements made.

Recommendation 3.4: Closer working relationships between obs4MIPs and ESMValTool should be built – potentially each observation provider could be required to provide a small recipe within ESMValTool to manipulate their data correctly providing a starting point for those doing analysis

Recommendation 3.5: subsetting tools should be provided to allow smaller data volumes to be downloaded, this should allow subsetting by spatial extent or by time period

Recommendation 3.6: regridding tools should be provided to correctly map the data (and uncertainties, see Section 5) to different resolutions and grids (e.g. rotated pole, unstructured).

Recommendation 3.6: better explanation of grid labels

Recommendation 3.7: a cloud computing environment isuch as JASMIN would be most useful for those in the global south, to mitigate against the connectivity issues

3.4 Uncertainties

The appetite for detailed uncertainty information varied considerably across the interviewees (Table 3). The point was made quite a number of times that too often climate modelers use observations as truth and spend little time investigating the sources of uncertainty. There are a number of reasons for this:

- Traditional observation datasets did not provide uncertainty information
- Lack of time to investigate the widely varying information if many different observation datasets are being used
- Lack of understanding about the source of uncertainty in many observation techniques
- Lack of understanding in what to do with the uncertainty information that is provided

Therefore, obs4MIPs is perfectly placed to educate the climate modelling community, to standardise this as far as possible and to provide the information in many different ways to suit a variety of users. These could be:

- Top level qualitative uncertainty in documentation
- Clear listing of the sources of uncertainty
- Overall quality information for entire dataset
- Information on conditions which will increase the uncertainty (cloud, land, high altitude)
- Grid box by grid box uncertainty values (standard deviation) where all sources of uncertainty are combined to give an overall estimate

CMUG support the work of obs4MIPs Task Team 3 on exploring options to include ancillary information, including uncertainties. See Section 5 for more on uncertainties and the recommendations.

Table 3: Uncertainty preferences given by participants

uncertainty	
grid box level	11
All sources of uncertainty should be detailed	8
Only total uncertainty needed	7
Qualitative overview in documentation	6
Standard deviation	4
More education on uncertainty and how to use	4
Believable range of data	3
Consistent across datasets	3
Want a published, peer reviewed paper to refer to	3
Will use more than one dataset for each variable to give idea of range/want a number of realisations of each dataset	4
Infilling process will introduce uncertainties, need to know what these are	2
accuracy	2
confidence value/weight	2
Need variability within each grid box	2
Acceptable uncertainty range is less than 2C	1
Need information on completeness of data	1
Quality metrics should be supplied	1
Mean	1
Systematic biases quantified	1
Correlation length	1
Quantitative uncertainties needed on climate model resolution and for different time resolutions	1

3.5 Dataset improvement

The gap analysis in Section 4 outlines the current shortcomings of the provision by obs4MIPs for widespread use across the climate science community, the feedback from the interview participants, summarised in this section, supports the conclusions drawn there (page 22).

Areas where potential improvements were noted in the data provided by obs4MIPs:

- More ocean information
- Emissions observations
- Higher spatial and temporal resolution
- Multiple datasets of same variable

Resolution requirements ranged from every 10 minutes to 30 year climate means, temporally, and spatially from 1 km up to 1.5 degrees. See Tables 4a and 4b.

spatial resolution	Number requiring
50 km	1
25 km	2
12 km	2
5 km	1
3 km	1
4 km	1
2 km	1
1.3 km	1
1 km	5
300 m	2
200 m	1
a few degrees	1
1.5 degrees	1
1 degree	5
0.5 degrees	4
0.25 degrees	2
0.1 degree	3
0.05 degrees	2
high res	8
lower res - same as global models	2
flexible grid resolution	1
Doesn’t matter, we always have to regrid anyway	2

Table 4a: spatial resolution requirements

temporal resolution	Number requiring
Climatology	1
annual	1
seasonality resolved	1
monthly	11

daily	15
sub daily	4
diurnal cycle resolved	5
day + night	2
6 hourly	1
3 hourly	2
hourly	5
15 minutes	1
10 minutes	1
as high as possible	4
Flexible grid resolution	1

Table 4b: temporal resolution requirements

The specific regions which are a focus for the interviewees are listed in Table 5.

regions	**Number of interviewees**
city scale	2
Europe	2
N Europe	1
South America	1
Rift valley basin	1
mid latitudes	1
Germany	1
EURO Cordex	1
Arctic	1
UK	3
global	7
Brazil	1
Amazon	2
high latitudes	1
land only	1
India	1
Africa	1

Table 5: Focus regions

The specific observational variables used by the interviewees are listed in Table 6.

variables	**Number of interviewees**
temperature	13
cloud variables	10
precipitation	8
wind	7
water vapour/humidity	6
LST	6
land cover	6
Soil moisture	5

Fire (burn area/emissions)	5
GHG emissions	5
evapotranspiration	4
river discharge	4
LAI	4
SSH	4
AOD	4
NDVI	3
methane	3
water budget	3
sea ice	3
SST	3
surface streamflow	2
surface solar radiation	2
co2	2
wetland area	2
GPP	2
solar induced fluorescence	2
surface pressure	2
surface fluxes	2
atmospheric composition	2
snow	1
near surface processes	1
global average temperature	1
SW TOA flux	1
LWP	1
concentrations	1
land-atmosphere interactions	1
ocean carbon	1
hydrology	1
permafrost	1
deep ocean	1
sea currents	1
water storage	1
inundation fraction	1
canopy height	1
flood	1
ocean biogeochemistry	1

Table 6: variables used by interviewees

Recommendation 5: Obs4MIPs steering panel should identify key datasets and invite the producers to contribute.

See recommendation 2.6, regular action on feedback from users and potential users will allow obs4MIPs to remain relevant as modelling evolves.

3.6 Mismatch between model and observation variables

The difference in what is measured and provided in an observational dataset and what a model output produces was discussed by many interviewees. A clear understanding of each variable is vital and provision of a tool to translate from observation to model space and vice versa was seen as extremely beneficial. Most participants thought the development of such a tool (likely needed on a dataset by dataset basis) would have to be produced in a collaboration between observation providers and modellers in order for the complexities of each dataset to be properly understood.

Recommendation 6: obs4MIPs to encourage observation simulators to be developed through partnerships between observation providers and users and for obs4MIPs to facilitate the sharing of these to users who need them (through links from documentation, or hosting with other tools which have been recommended, recommendation 3).

4. Gap Analysis

In this section, the contents of the existing Obs4MIPs database have been analysed and gaps and issues identified.

4.1 Obs4MIPs dataset selection

The Obs4MIPs datasets are hosted within the Earth System Grid Federation, which provides federated access to climate modelling and observational datasets hosted across a world-wide distributed network of nodes. The Obs4MIPs datasets are published to several different ESGF data nodes and are then downloadable and accessible via any participating ESGF index node.

In recent years, the main publishers of Obs4MIPs datasets have been the Lawrence Livermore National Laboratory (LLNL) (https://aims2.llnl.gov/search) and the Centre for Environmental Data Analysis (CEDA) (https://esgf-ui.ceda.ac.uk/cog/search/obs4mips-ceda/) nodes. However, due to different search options used in the configuration of each index node, the two nodes are currently display differing numbers of Obs4MIPs datasets. This difference is purely in the search filters used, as searching each node using the other's default search, returns the same results.

This mismatch in configuration is something that should be resolved in the near future and is under discussion in the context of the Obs4MIPs Steering Group.

The differences currently found between the nodes (as of 3/12/2024) are summarised in Table 7.

Table 7: differences between obs4MIPs nodes

Index node	Search request used	No of Obs4MIPs datasets returned
LLNL	https://esgf-node.llnl.gov/esg-search/search?activity_id=obs4MIPs&offset=0&limit=150&type=Dataset&format=application%2Fsolr%2Bjson&latest=true&query=*&	69
CEDA	https://esgf.ceda.ac.uk/esg-search/search/?offset=0&limit=150&type=Dataset&replica=false&latest=true&project=obs4MIPs&format=application%2Fsolr%2Bjson	135
Modified CEDA (allowing replica = True)	https://esgf.ceda.ac.uk/esg-search/search/?offset=0&limit=150&type=Dataset&latest=true&project=obs4MIPs&format=application%2Fsolr%2Bjson	139

The difference between the nodes are for two reasons: a) LLNL is searching for datasets identified by project = obs4MIPs, which is a field that was only introduced in v2.1 of the Obs4MIPs data standards, so older datasets are not included, whereas CEDA is searching on activity_id which includes older datasets; b) CEDA is missing 4 datasets that are included in LLNL as they are only searching for datasets with replica = false set. Removing this search term ensures that all the LLNL datasets are also returned within the CEDA output.

Further discussion is required with LLNL and the Obs4MIPs steering group to determine exactly which datasets are considered to be current and therefore should be shown by default on the Obs4MIPs portals. For the purpose of this analysis we have used the superset of all the datasets, to ensure that all potential Obs4MIPs datasets are included. However, it should be noted that some of the older Obs4MIPs datasets, although searchable, are likely very out of date, and additionally some are on data nodes that are no longer accessible, so these would benefit from updated versions of the data products being produced in the latest format and republished to active nodes. For some of these, work is already ongoing.

The issues identified with the data nodes at the time of writing were:

- Data hosted on eridanus.eoc.dlr.de and esgdata.gfdl.noaa.gov returned errors when trying to access the data
- Data on aims3.llnl.gov and vesg.ipsl.upmc.fr returned an error for the given http download link, but this worked when changed to https.

4.2 Analysis of Datasets

A snapshot table of the current obs4mips datasets (as of 3/12/2024) is given in Annex 2: Obs4MIPs datasets.

From an analysis of those datasets, we have first identified some top level gaps and then made a closer comparison to requirements for CMIP.

1. There are currently only a small number of data products included on Obs4MIPs. These correspond to only 44 source ids (where the source id represents a set of data that has been similarly produced e.g. using the same instrument and algorithm by a given data provider). In many cases, the version ID is also included as part of the source id, so in reality there are even less than 44 separate data product families. These cover around 65 different variables, so a number of datasets are focused on similar areas.

2. Nearly all the observation datasets are currently satellite-derived. In-situ datasets are not represented at all. There is a single reanalysis dataset from ERA-Interim (obs4MIPs.ECMWF.ERA-interim.atmos.mon.v20160614), which is an old dataset that is now likely superseded by more recent ECMWF products.

3. All the datasets that have been published currently come from US or European satellites and data producers.

4. Many of the datasets were published to Obs4MIPs some time ago, and may therefore represent data for which there is an improved version available externally, or where the data has since been extended forward in time. Even where this is not the case, the resulting data follows different versions of the Obs4MIPs data format specifications, as shown in Figure 2. Here, the large number of unspecified datasets are likely to be datasets which are in a format predating ODS2.1; it would be beneficial to consider updating these datasets so that Obs4MIPs users can work with a consistent format.

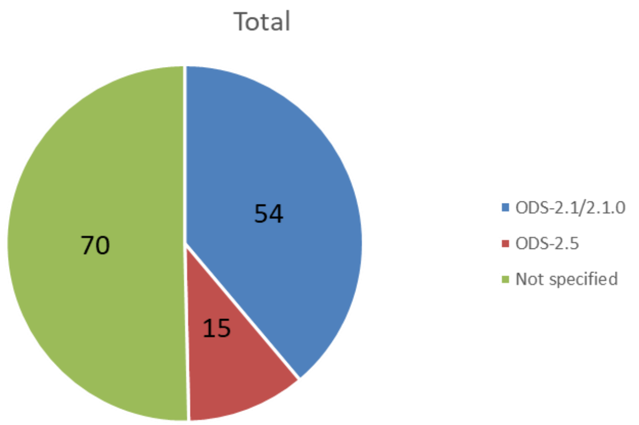


Figure 2: Number of Obs4MIPs datasets in each of the different versions of Obs4MIPs format. The data marked as not specified is likely to predate v2.1 of the Obs4MIPs Data Specifications (ODS-v2.1), where the format specification version was added as a field.

5. The datasets cover a range of spatial gridding as shown in 3. However, it can be seen that most datasets are labelled as unspecified, as earlier datasets did not include the field used to analyse the gridded resolution. These are however, likely to mostly be fairly low resolution grid (e.g. 1x1 degree) to match CMIP requirements.

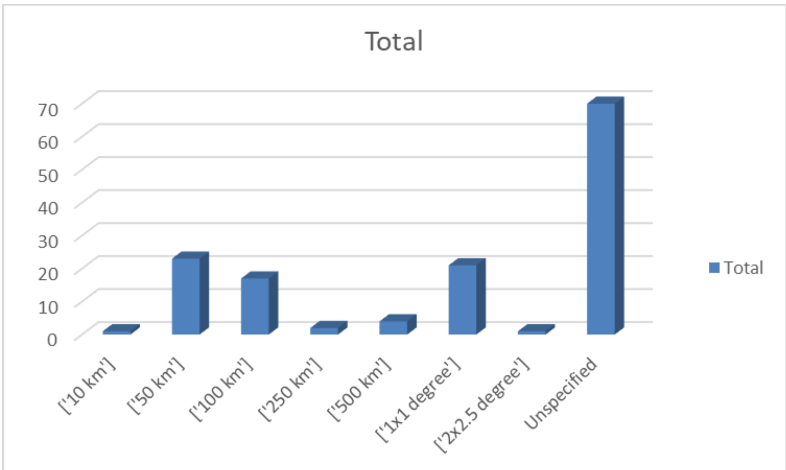


Figure 3: Spread of nominal spatial resolution of Obs4MIPs dataset

6. The majority of datasets in Obs4MIPs are monthly averaged datasets, with a small number of daily data (including daytime /nighttime separation), 3 hourly datasets, and monthly averages of hourly data. This can be seen in 4.

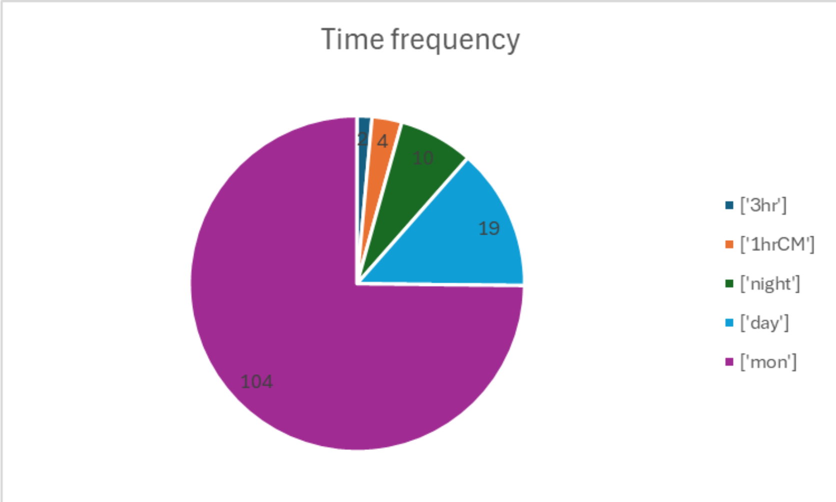


Figure 4: Temporal frequency of Obs4MIPs datasets

7. Within Obs4MIPs the datasets are classified into realms, using the equivalent CMIP-6 controlled vocabulary terms, with the full list comprising: Aerosol ("aerosol"); Atmosphere ("atmos"); Atmospheric Chemistry ("atmosChem"); Land Surface ("land"); Land Ice ("landIce"); Ocean ("ocean"); Ocean Biogeochemistry ("ocnBgchem"); Sea Ice ("seaIce")

   However, as can be seen from 5**Error! Reference source not found.Error! Reference source not found.** nearly all the current Obs4MIPs datasets correspond to the 'Atmosphere' realm, and some realms are not covered at all

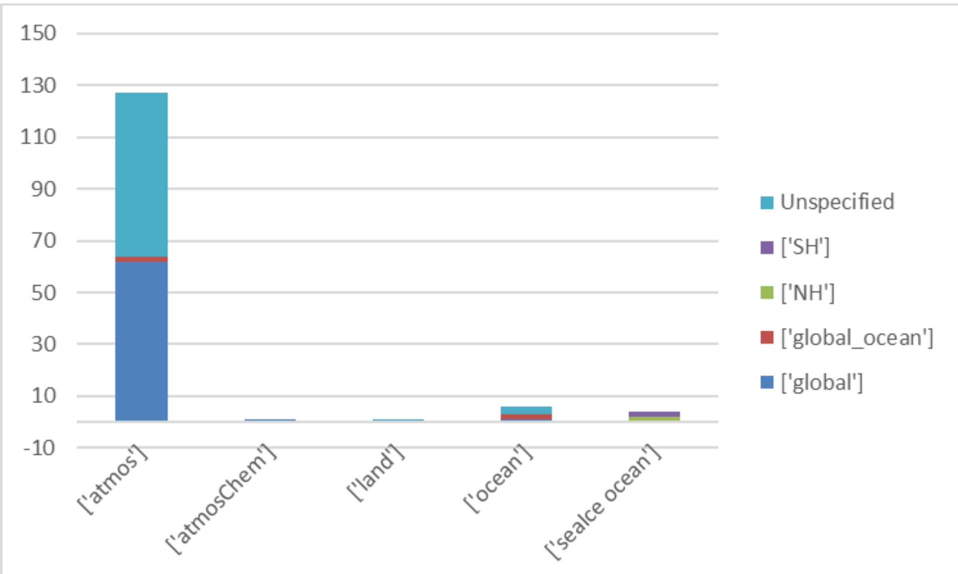


Figure 51: Obs4MIPs datasets by CMIP-6 realm. Bars are further subdivided by geographical coverage.

A very limited amount of data is available outside of the atmosphere domain. These non-atmospheric datasets are summarised in Table 8. As well as only covering a handful of non-atmospheric variables, these datasets were all published to obs4MIPs several years ago, and are likely to have been either superseded by later versions or extended forward in time. There is now a much wider range of satellite variables covering these domains from activities such as CCI as well as in-situ datasets, so it would be beneficial to increase the focus of Obs4MIPS beyond the atmospheric datasets.

Table 8: Current Obs4MIPs datasets which are not in the Atmosphere CMIP realm

Atmospheric Chemistry	obs4MIPs.NASA-GSFC.MODIS-1-0.mon.clt.gn	MODIS total cloud fraction, published to Obs4MIPs in 2019
Land	obs4MIPs.NASA-GSFC.MODIS.land.mon	Leaf Area Index from MODIS, published to Obs4MIPs in 2020
Ocean	obs4MIPs.UOE.ARC-SST-1-1.mon	Sea Surface Temperature from the ARC project, published to Obs4MIPs in 2016
	obs4MIPs.UReading.ESA-CCI-SST-v2-1.mon.tos.gn	SST from the ESA CCI project (version 2.1), which has heritage from the ARC project above and is a more recent dataset, published to Obs4MIPs in 2021. NB. There is now a newer version (v3) of SST available from CCI
	obs4MIPs.NCEI.OISST.tos.mon and obs4MIPs.NCEI.OISST.tos.day	Monthly and daily averaged Sea Surface Temperature produced by NCEO, published in 2017
	obs4MIPs.CNES.AVISO-1-0.mon.zos.gn	sea surface height about the geoid from the CNES AVISO service, published to Obs4MIPs in 2019
Sea Ice	obs4MIPs.NCEI.PMSIC.SH.sic.day, obs4MIPs.NCEI.PMSIC.NH.sic.day, obs4MIPs.NCEI.PMSIC.SH.sic.mon, obs4MIPs.NCEI.PMSIC.NH.sic.mon	Daily and monthly sea ice concentration data produced by NCEI. Published to Obs4MIPs in 2017

4.3 Comparison to CMIP

To date, the Obs4MIPs project has been focused on providing datasets that map to an CMIP-6 dataset, using analogous Obs4MIPs tables. The CMIP-6 data request comprises over 2000 variables, which are organised into MIP tables focusing on a given area (e.g. the Amon mip table refers to monthly atmospheric variables). The details of the variables in the CMIP-6 data request can be found at https://clipc-services.ceda.ac.uk/dreq/index.html. Currently Obs4MIPs provides only a small number of the variables used in CMIP-6, and as discussed in the section above, the vast majority of the Obs4MIPs datasets cover variables within the Amon MIP table, with only a small handful covering other realms or temporal averaging.

Work is now underway on CMIP-7, and the first version of the CMIP-7 data request has now been released (https://wcrp-cmip.org/cmip7-data-request-v1-0/). This is a current live process, and is currently limited to the AR7 Fast Track experiments. Although a more focused approach has been taken to the core CMIP variables needed, there are still over 1000 variables in the initial version of the CMIP-7 data request, so it is not feasible to do a comprehensive gap analysis compared to Obs4MIPs, as most CMIP-7 variables will not be included in Obs4MIPS. However, a comparison was made against the 132 baseline variables chosen as highest priority (https://egusphere.copernicus.org/preprints/2024/egusphere-2024-2363/egusphere-2024-2363.pdf). Note, the comparison has been done just on the claimed variable names in Obs4MIPs, so it is possible that some of the Obs4MIPs variables may differ slightly in definition. Additionally, the status of the Obs4MIPs datasets and whether it is in need of updating has not been considered here.

As before, it can be seen that there is good coverage of many of the baseline variables within the monthly averaged atmosphere table (AMON), but very little datasets covering the other realms or other time periods.

Table 9 CMIP-7 baseline variables (from the core variables in the CMIP-7 v1.0 data request) indicating which variables are at least partly represented in Obs4MIPs.

CMIP-7 compound name	Present in Obs4MIPs	CMIP-7 compound name	Present in Obs4MIPs	CMIP-7 compound name	Present in Obs4MIPs
3hr.huss		Amon.tasmin		Lmon.evspsblveg	
3hr.pr	X	Amon.tauu		Lmon.lai	X
3hr.tas		Amon.tauv		Lmon.mrfso	
3hrPt.uas		Amon.ts		Lmon.mrro	
3hrPt.vas		Amon.ua	X	Lmon.mrros	
6hrPlev.hurs		Amon.uas	X	Lmon.mrso	
6hrPlevPt.ta		Amon.va	X	Lmon.mrsos	
6hrPlevPt.ua		Amon.vas	X	Oday.sos	
6hrPlevPt.va		Amon.wap		Oday.tos	
Amon.cl	X	Amon.zg	X	Ofx.areacello	
Amon.cli	X	CFday.ps		Ofx.basin	
Amon.clivi	X	day.clt		Ofx.deptho	
Amon.clt	X	day.hur		Ofx.hfgeou	
Amon.clw	X	day.hurs		Ofx.masscello	
Amon.clwvi	X	day.hus		Ofx.sftof	
Amon.evspsbl		day.huss		Ofx.thkcello	
Amon.hfls		day.pr		Omon.bigthetao	
Amon.hfss		day.psl		Omon.hfds	
Amon.hur	X	day.rsds		Omon.masscello	

Amon.hurs		day.sfcWind		Omon.mlotst	
Amon.hus	X	day.ta		Omon.so	
Amon.huss		day.tas		Omon.sos	
Amon.pr	X	day.tasmax		Omon.tauuo	
Amon.prc		day.tasmin		Omon.tauvo	
Amon.prsn		day.ua		Omon.thetao	
Amon.prw	X	day.uas		Omon.thkcello	
Amon.ps		day.va		Omon.tos	X
Amon.psl		day.vas		Omon.umo	
Amon.rlds	X	day.wap		Omon.uo	
Amon.rldscs	X	E1hr.pr		Omon.vmo	
Amon.rlus	X	Eday.hus		Omon.vo	
Amon.rlut	X	Eday.ua		Omon.wmo	
Amon.rlutcs	X	Eday.va		Omon.wo	
Amon.rsds	X	Eday.zg		Omon.zos	X
Amon.rsdscs	X	Efx.slthick		Omon.zostoga	
Amon.rsdt	X	fx.areacella		SIday.siconc	
Amon.rsus	X	fx.mrsofc		SImon.siconc	
Amon.rsuscs	X	fx.orog		SImon.simass	
Amon.rsut	X	fx.rootd		SImon.sisnthick	
Amon.rsutcs	X	fx.sftgif		SImon.sitemptop	
Amon.sfcWind	X	fx.sftlf		SImon.sithick	
Amon.ta	X	LImon.snc		SImon.sitimefrac	
Amon.tas		LImon.snw		SImon.siu	
Amon.tasmax		Lmon.evspsblsoi		SImon.siv	

4.4 CCI ECV datasets

The Essential Climate Variable (ECV) datasets produced under the ESA Climate Change Initiative programme cover a much wider range of variables than are currently represented in the Obs4MIPs datasets, and therefore represent a good opportunity for broadening out the datasets into other CMIP realms.

Current CCI datasets that have been included in Obs4MIPs are:

- *Cloud data products from ATSR-2/AATSR, AVHRR-AM, and AVHRR-PM satellites* (source ids: ESACCI-CLOUD-ATSR2-AATSR-3-0, ESACCI-CLOUD-AVHRR-AM-3-0', ESACCI-CLOUD-AVHRR-PM-3-0'). Variables included are: CCI Cloud Area Fraction; Ice Water Path; CCI Total Cloud Fraction; CCI Liquid Cloud Area Fraction; CCI Total Liquid Cloud Area Fraction; Condensed Water Path; CCI Mean Cloud Top Pressure
- *Sea Surface Temperature* (source id: ESA-CCI-SST-v2-1). Now superseded by v3 CCI data, so could be updated.
- *Column averaged mole fraction of atmospheric methane* (source id: XCH4_CRDP3). Now superseded by more recent versions in C3S, so could be updated
- *Column averaged mole fraction of atmospheric carbon dioxide* (source id: XCO2_CRDP3). Now superseded by more recent versions in C3S, so could be updated.
- *Total Column Ozone* (source id: C3S-GTO-ECV-9-0)

- *Ambient Aerosol Optical Thickness at 550nm: (source id: obs4mips.SU.ATSR2-AATSR.od550aer.mon):* On a data node that is no longer available, so needs moving elsewhere.

As there are now more than 25 ECV focused CCI projects and other projects in the pipeline, then there is considerable scope to increase the range of datasets in Obs4MIPs. The inclusion of other datasets has in part been limited by whether or not a close match to a CMIP variable can be made with observational data. However, within the current phase of Obs4MIPs, the concept of an 'exploratory' dataset that does not need to be an exact match to a CMIP dataset, but which is still considered to be useful for model evaluation, is being introduced, and this will expand the number of CCI datasets that can be easily included within Obs4MIPs.

4.5 Gap Analysis summary

From the analysis above, a number of gaps in the current Obs4MIPs datasets have been identified. The main points are summarised below:

- Many of the Obs4MIPs datasets are relatively old, with the disadvantages that they may have been superseded by more recent versions, they do not cover the latest time periods, and they may not be in the latest Obs4MIPs format. For those datasets, there would be a benefit to providing updated data where relevant. (In some cases, this is already happening).
- Obs4MIPs is currently very focused on atmospheric products, with limited reach into other domains.
- In-situ datasets are not currently included in obs4MIPs, but their inclusion could fill some of the gaps.
- The provision of datasets has so far been focused on a small number of participants within the US and Europe, limiting the types of data that have been included in Obs4MIPS.
- Obs4MIPs datasets are mostly monthly averaged. In the future, it is likely that there will be an increased need in the climate modelling community for higher resolution datasets (both temporally and spatially) which should be reflected in the Obs4MIPs datasets.

5. Treatment of uncertainties

Since the conceptualization of Obs4MIPs in 2010, substantial developments have been made in the way that observation uncertainties are calculated and conveyed to the user, particularly in remote sensing datasets, which presently cannot be effectively included in the Obs4MIPs file specification. This report discusses the importance of these developments and possible routes to inclusion of more detailed uncertainty information in future iterations of Obs4MIPs data.

5.1. Introduction

Obs4MIPs is a data format designed to facilitate easy intercomparison between observational datasets and model output, initially driven by the Coupled Model Intercomparison Project (CMIP) as part of the World Climate Research Programme (WCRP) (Waliser et al, 2020). The latest Obs4MIPs data specifications (ODS2.5) were released in June 2024 and specify that each Obs4MIPs data file contains only one variable, with additional data fields providing the means to geolocate the data e.g. time, latitude and longitude (Gleckler et al, 2024). Each Obs4MIPs variable can have an accompanying technical note, in which information on uncertainties can be detailed. The technical note preparation guidelines indicate that uncertainty information can be provided as single values, potentially with an indication of spatial or temporal variation, with use of a supplementary figure or table as required[3]. In the Obs4MIPS data specification, only in the case of a variable being the average of several observations in there the potential to provide a per datum measure of uncertainty in the form of a standard error or standard deviation (this was done with SST CCI v2 data). This technical note format was deemed to meet the recommendation by the WCRP Data Advisory Council (WDAC) Task Team on Observations for Model Evaluation to 'develop a capability to accommodate reliable and defendable uncertainty measures' set out in 2014 ahead of preparations for CMIP6 (Waliser et al, 2020). No subsequent updates to this aspect of the Obs4MIPs specification have been made in preparation for CMIP7 (Glecker et al, 2024).

Since the recommendations were made by the WDAC Task Team, significant developments have been made in our understanding of error sources and the way in which uncertainties are calculated. This has arisen from a close collaboration between Earth Observation scientists and metrologists (metrology is the science of measurement). This has happened across a variety of projects but two of significant note were the Horizon 2020 FIDUCEO project (Mittaz et al, 2019) and the European Space Agency (ESA) Climate Change Initiative (CCI) (Merchant et al, 2017). The result has been the provision of per datum prognostic uncertainties for all data products for a number of Essential Climate Variables (ECVs). The most developed thinking to date has been for the Sea Surface Temperature (SST), Land Surface Temperature (LST) and Sea Level products [Bulgin et al, 2016a; Bulgin et al, 2016b; Ghent et al, 2019, Ablain et al, 2019], with other ECVs such as soil moisture taking a similar approach (Dorigo et al, 2023). Within the CCI programme, provision of uncertainty information, which is essential for using the data products, has become a central pillar of all funded ECV activities. The aim of this document is to discuss the importance and relevance of these developments to the provision of Obs4MIPs data.

5.2. Uncertainties in Sea and Land Surface Temperature Products

5.2.1 Specification of uncertainties in remote sensing products

[3] https://zenodo.org/records/14276263

All ESA CCI SST and LST products provide a per datum uncertainty for the retrieved surface temperature. Without this uncertainty, the confidence with which the product can be used in a decision-making context is reduced as the user would have no measure of confidence in the provided data. In addition to the total per datum uncertainty, the uncertainty budget is also further sub-divided into its constituent components; the sum of which in quadrature is equal to the total (Bulgin et al, 2016a, Bulgin et al, 2016b, Ghent et al, 2019).

The uncertainty budget is constructed by first identifying all possible sources of error (error effects) at each stage of the retrieval process; from the initial satellite observation to the retrieval itself and the re-gridding of data onto a regular lat-lon grid (Bulgin et al, 2016a, Bulgin et al, 2016b, Mittaz et al, 2019, Ghent et al, 2019) and then quantifying the uncertainty associated with each error effect, where possible. Error is defined as the difference between the measured value and the reference value (JCGM, 2012) and is typically unknown. If we were able to calculate the error on a given observation, this would then be corrected for. Instead, what we can estimate is the uncertainty, the degree to which a measurement is 'in doubt' according to the measurement process. This is often expressed as the dispersion of the possible error values attributable to the measurand (JCGM, 2012). Uncertainties add in quadrature, so once we have the uncertainty attributable to each error source, these can be summed to provide a total uncertainty. Each step in the retrieval process will require propagation of uncertainties through from the previous step, plus calculation of any new uncertainties introduced by the process employed at the given step (Bulgin et al, 2016a; Bulgin et al, 2016b). This definition and propagation of uncertainties follows the standard laws of uncertainty propagation as set out in the Guide to the expression of Uncertainty in Measurement (JCGM, 2008).

In the final products, the uncertainty breakdown is provided according to the correlation length scale of each error source e.g. the uncertainties for all error sources that were uncorrelated between pixels would be grouped together. The three types of uncertainty component provided are: 1) independent: this type of uncertainty is uncorrelated between a given observation and the neighbouring observations. 2) structured or locally correlated: this type of uncertainty is correlated with neighbouring observations over a correlation length scale attributed to the effects, in time and/or space. 3) common – this type of uncertainty is fully correlated between all observations in a given satellite mission (Bulgin et al, 2016a). The uncertainties are grouped in this way so that users who want to calculate further derived products, such as coarser re-gridding or regional means can propagate the best estimate of uncertainty from their input data as the method of propagation differs depending on the correlation length scale.

In SST and LST products, the independent component of the uncertainty is arising mainly from instrument noise and from sampling uncertainty when re-gridding observations from the satellite image grid to regular lat-lon outputs (Bulgin et al, 2016a, Bulgin et al, 2016b). By definition, these surface temperature products are 'gappy' as cloud obscures the Earth's surface and prevents temperature retrieval in its presence. Thus, the sampling uncertainty arises when a given region is only partially observed (Bulgin et al, 2016b). Both SST and LST have a structured uncertainty component that relates to the parameterisation of the atmosphere in the retrieval, through which the satellite is viewing the Earth's surface. For SST products, the correlation length scale of this component is considered to be of a synoptic scale (100 km, 1-day) (Bulgin et al, 2016a), whilst for LST much finer timescales are assumed (5 km and 5 minutes) (Ghent et al, 2019). LST also has a second structured component of uncertainty that relates to the surface specification. This is characterised predominantly on the basis of the assigned land cover classification for each observation (Ghent et al, 2019). Both products have a common uncertainty component that reflects calibration errors common to all observations made by a given satellite (Bulgin et al, 2016a, Ghent et al, 2019).

5.2.2 Relevance of uncertainties to Obs4MIPs data

Uncertainties are inherently essential to interpreting and using any data as they describe the degree of confidence in the measurement provided. In the context of Obs4MIPs, observation data are commonly used to evaluate model performance. Climate modellers are often well-aware of the short comings of the model they are working with and indeed regularly assess the model space through the use of ensembles. CMIP is set up to do exactly this, providing one ensemble member per file, which together can be analysed to understand the spread in the model predictions. Observations are no different, but they come in a single realisation, with a pre-calculated uncertainty, which is not well suited to the current Obs4MIPs framework. Historically observations have been considered to be the 'truth' or 'reference' for benchmarking climate models, and this may in part have influenced the original structure of the Obs4MIPs data. In reality, observations can have uncertainties that vary significantly in space and time and can be large (at least some of the time). Without a proper understanding of these uncertainties, then erroneous or weak conclusions can be drawn about the ability of a climate model to correctly represent the current observation space.

Providing uncertainty information with Obs4MIPs data becomes even more critical when one considers the nature of the comparisons being made. It is rare that a model output will be directly replicated by a satellite observation. Consider the case when a comparison between a daytime average LST is required between a model ensemble member and a satellite observation dataset. The model will output a globally complete LST field at a given time interval e.g. hourly. To calculate the daytime average, a mean value can be calculated for all daytime outputs. In the case of the satellite data, the closest representation to what a model is producing would be a geostationary satellite (looking at the same part of the globe, at 10-15 minute intervals). In this case, you could take the hourly data and average these data. However, the resulting values still wouldn't be exactly the same as 1) the sensor takes a while to complete the scan over the whole area (it isn't instantaneous) and 2) some observations in some time steps would be missing due to the presence of cloud. If, in a given location, a satellite observation was only available at 1200 and 1300 UTC (all other times were cloudy), this would give a very different average value to if the diurnal cycle were fully sampled. These differences would occur in addition to uncertainties that arise from the measurement process itself.

In the case of polar orbiting satellites, the comparison becomes even more difficult to align. The satellite will typically take around 90 minutes to complete one revolution of the Earth's surface, with 14-15 orbits in a given day. Despite being sun-synchronous, overpassing the equator at the same local time of day, not all observations at every part of the orbit will be made at the same local time of day. Gaps will occur in the data both due to cloud and spaces between orbits (very few satellites achieve global coverage in 24 hours) and there may not be any repeat measurements except in polar regions where orbits overlap. So, in reality a daytime 'average' LST is likely a single retrieval, where available, at varying (local) times of day, which is not immediately comparable to the model output. Corrections may be made to the observations to provide an estimate at a uniform time, but this correction will also have an associated uncertainty, which would be communicated by the data producer using the per datum uncertainty information.

Under these circumstances, the uncertainties in the observations are fundamental to making a meaningful comparison between the model and the observations. Often the first step required will be to coarsen the satellite observations to the resolution of the model, or perhaps further, to look over a specific region of interest. In calculating that new LST value, the uncertainties from the input need to be correctly propagated and a further sampling uncertainty calculated if those data are still 'gappy' in nature [Bulgin et al, 2024 (in preparation)]. This has just been recognised as a necessity within the Climate Modelling User Group (CMUG) and coding is underway to facilitate the correct propagation of uncertainties when re-gridding satellite observations within ESMValTool [CMUG D5.3v1, in review].

5.3. Recommendations for inclusion of uncertainty information in Obs4MIPs

It is recommended that satellite per datum uncertainty information (where available) are included in Obs4MIPs datasets (Merchant et al, 2017). The amount of information required is dependent on the dataset use (Table 1). In reality, the first case, where the observation dataset is simply used 'as provided' is likely uncommon. This is because observational datasets are typically of higher resolution than models and coarsening is generally required in order to compare like-for-like.

Observation Dataset Use	**Recommendation**
Dataset to be used 'as is' with no derived products calculated e.g. no coarsening or re-gridding of data.	Include the total uncertainty with the data.
Dataset to be used to generate derived products e.g. coarsening to model resolution, regional averaging.	Likely requirement to include full breakdown of uncertainty components but this depends on the correlation length scales of the components, their relative weighting and the resolution at which the re-gridding is applied. Total is optional where the individual components are provided as it can be calculated by summing the components in quadrature.

Table 10. Recommendations for inclusion of observation uncertainty information in Obs4MIPs format data.

5.4. Routes to inclusion of uncertainty information in Obs4MIPs

In order to include observational uncertainty information in Obs4MIPs datasets, one of two routes could be taken:

1. Supplying uncertainty information using additional Obs4MIPs format files. This would entail a user downloading one file containing the geophysical variable of interest and then a further 1-4+ files depending on whether they required access to the total uncertainty only, or the total uncertainty in addition to the breakdown of uncertainty components.
2. Updating the Obs4MIPs file specification to allow for the inclusion of uncertainty information, in addition to the geophysical variable. This could also continue to support standard error or standard deviation fields for backwards compatibility with existing standards.

Of the two suggested routes, the second is the recommended approach. This is because it provides all the information required by a user to correctly use the Obs4MIPs dataset, in the same place. It reduces the possibility that a data user would be either 1) unaware that the uncertainty information was provided in a separate file or 2) unable to successfully download and/or match the correct uncertainty information to the geophysical variable.

As both observations and models develop, we should prioritise maximising the use of the best possible data from each, and for observations this includes the uncertainty information. We must also recognise that there is a scale of product maturity with regard to the uncertainty information provided and that even at the most mature end of the scale, uncertainty information may be (knowingly) incomplete. As such, some flexibility is required in any update to the Obs4MIPs file specification.

The recommendation would be to have a series of optional additional ‘variables’ or ‘supplementary/ancillary’ fields (dependent on the preferred nomenclature) that enable producers of observational datasets to include the most appropriate uncertainty information for their product. A list of suggested optional fields is provided in Table 2. These would be per datum fields, constrained in time and space by the existing geolocation information provided in the file to define the geophysical variable. Where data producers have less detailed information e.g. not per datum, they could still use the existing route to communication of uncertainty via the accompanying technical note. In this case it would not be recommended to include an optional per datum field filled with a constant value as this would unnecessarily inflate the dataset size. Defining these fields as optional would enable data users to provide an appropriate subset of the fields depending on the maturity of their data product and the approach taken to quantify uncertainty.

Additional variable/auxiliary information	Description
total_uncertainty (this field would be mandatory)	This variable would contain the total per-datum uncertainty associated with the geophysical variable. If the independent, structured and common uncertainty components are also provided, this would be equal to the sum in quadrature of these components.
independent_uncertainty	This variable would contain the per-datum component of uncertainty that is uncorrelated between observations.

structured_uncertainty	This variable would contain the per-datum component of uncertainty that is structured and correlated over a defined space/time scale. This correlation length scale in space and time must be provided in the variable metadata.
common_uncertainty	This variable would contain the per-datum component of uncertainty that is common to all observations of a given type (specific instrument, resolution, variable type) in Obs4MIPs format
standard_deviation	Where an observation average is provided over a given space/time scale, the data producer may include the standard deviation of the data comprising the average in this field.
standard_error	Where an observation average is provided over a given space/time scale, the data producer may include the standard error of the data comprising the average in this field.

Table 11: List of potential new optional variables/ancillary information for Obs4MIPs format data. Total uncertainty would be mandatory and other fields optional.

Some flexibility would be required in the definition of the structured uncertainty component, as some variables, LST being a good example, contain more than one locally correlated uncertainty component (Ghent et al, 2019). The need for multiple structured uncertainty components arises when the correlation length scale of the uncertainty differs between the two(+). In the case of LST, one is correlated with atmospheric processes (5 minutes, 5 km), whilst the other is correlated with the surface biome definition.

It would also be recommended that the scope of the accompanying technical note would be expanded, to enable the data provider to specify some more information about the total uncertainty and breakdown of uncertainty components, where these are provided. This could include things such as the contributing error effects, or known omissions (known unknowns) as some uncertainty components are still very difficult to fully characterise. Metadata on the correlation time and length scale would be essential for all structured uncertainty components.

5.5. Enabling modellers to understand and use Obs4MIPs uncertainties

The main purpose of Obs4MIPs data is to allow intercomparison with model outputs, and this is predominantly done by modellers who want to evaluate model performance. As such, accessible information on how to use uncertainties is essential to facilitate the use of observational uncertainty information by modellers, who by definition are not experts in the observational dataset production. If this is not provided, then much effort may be expended in improving Obs4MIPs format datasets without tangible benefits to the ensuing science (if uncertainty information is ignored or used incorrectly).

One possibility is the provision of a ‘universal recipe’ explaining how to use observational uncertainty components is provided for modellers. This recipe should be applicable to any Obs4MIPs dataset containing uncertainty components, and this should be ensured by the constraints placed on the file format. This recipe should also be a simple and straightforward as possible in its presentation, with all the relevant information accessible in the same place to ensure that barriers to implementation/use of this information are reduced as much as possible. In some contexts, e.g. in ESMValTool, it may be possible to generate common code or functions that are able to manipulate uncertainty information given appropriate inputs. Where these are widely applicable for different geophysical variables, it would provide consistency in analysis and intercomparisons.

An outline detailing the possible format of such a 'recipe' is provided below (note this example covers only the uncertainty components and assumes that these are Gaussian in distribution – some additional information could be added about how to use standard deviation and standard error fields. This example is written based on the assumption that Obs4MIPs data are always spatially complete. **Note: step 2 in this example is an approximation for the propagation of the systematic component and not the full propagation. This may not be the recommended formulation in an actual 'recipe' but is given here for illustrative purposes only.**

A simple recipe for using uncertainty information in Obs4MIPs datasets

Obs4MIPs datasets can be provided with four possible uncertainty fields: a total, and then a breakdown of uncertainty components into independent, structured and common components. In some cases, more than one structured component may exist.

If the Obs4MIPs data are already at the spatiotemporal resolution you require then you can use the total uncertainty to inform your analysis. The total uncertainty describes the degree to which the measurement is 'in doubt' and can vary in both space and time. These should be considered when making comparisons with model output, rather than assuming that the observations are without uncertainty.

If you need to make any modifications to the resolution of the Obs4MIPs data in order to compare it with the model output, then you need to use the breakdown of uncertainty components in order to do this. If your Obs4MIPs data contain the total uncertainty in addition to the components, the sum of the components in quadrature is equal to the total:

$$\sigma_{total} = \sqrt{\sigma_{ind}^2 + \sigma_{str}^2 + \sigma_{com}^2} \quad (1)$$

Here the subscripts 'ind', 'str' and 'com' refer to independent, structured and common respectively. Note the file may contain more than one structured component. Each should be added in the same way.

To calculate the uncertainty for an average of the geophysical variable over a given space/time scale you should do the following. First calculate the individual uncertainty components for your averaged value. Then add them in quadrature following equation (1) to give you a total uncertainty.

1. ***Calculate the independent uncertainty for your average***

Given that you have 'n' observations contributing to your geophysical variable average, calculate the independent uncertainty (σ_{ind_output}) as follows:

$$\sigma_{ind\_output} = \sqrt{\frac{1}{\sqrt{n}} \sum_{1}^{n} \sigma_{ind}^{2}} \tag{2}$$

2. ***Calculate the structured uncertainty for your average***

Given that you have 'n' observations contributing to your geophysical variable average, calculate the structured uncertainty (σ_{str_output}) as follows:

a) *First work out the number of independent pieces of information (m) that you have:*

$$m = \frac{r}{b} \tag{3}$$

Where 'r' is the size of your region e.g. the total domain over which you are calculating the average. This may have both time and space dimensions. 'b' is the size of the box over which uncertainty component is correlated. Calculate this using the correlation length scales provided as metadata with the variable.

b) *Then calculate the propagated structured uncertainty using the input information and 'm' defined in step a).*

$$\sigma_{str\_output} = \sqrt{\frac{1}{\sqrt{m}} \sum_{1}^{n} \sigma_{str}^{2}} \tag{4}$$

If your file contains more than one structured uncertainty component do this for each one individually, using the appropriate length scale.

3. ***Calculate the common uncertainty for your average***

Given that you have 'n' observations contributing to your geophysical variable average, calculate the common uncertainty (σ_{com_output}) as follows:

$$\sigma_{com\_output} = \sqrt{\frac{1}{n} \sum_{1}^{n} \sigma_{com}^{2}} \tag{5}$$

4. ***Finally add the components in quadrature to get a total value***

Use equation (1) given above to calculate the total uncertainty for your averaged value. Add each structured component individually.

Recommendation 7: Improve the provision of uncertainties and documentation on how to use them

Recommendation 7.1: Include both total uncertainty information with the data and full breakdown of uncertainty components with guidance on how to sum these to obtain the total. Also, guidance on how to translate to different resolutions.

Recommendation 7.2: Prioritise the work on exploring options to include ancillary information, including uncertainties which has been assigned to obs4MIPs TT3.

Recommendation 7.3: have a series of optional additional 'variables' or 'supplementary/auxiliary' fields (dependent on the preferred nomenclature) that enable producers of observational datasets to include the most appropriate uncertainty information for their product.

Recommendation 7.4: scope of the accompanying technical note would be expanded, to enable the data provider to specify some more information about the total uncertainty and breakdown of uncertainty components, where these are provided.

Recommendation 7.5 a universal recipe explaining how to use observational uncertainty to be provided by obs4MIPs for modellers.

6. Summary of recommendations

Recommendation 1: Obs4MIPs to be publicised more widely with a clear statement of scope especially the extent to which the variable format is defined by the project ie:

- **which variables are currently included and any restrictions for future variables to be added**
- **resolution (spatial and temporal)**
- **length of timeseries**
- **level of quality control carried out by obs4MIPs**

Recommendation 1.1: Regular (quarterly) obs4MIPs newsletter which contains:

- a list of all current datasets available
- any recent updates to these
- plans for the next period

This newsletter could be subscribed to, and all major modelling centres could be contacted through CORDEX and CMIP IPOs to publicise when this is set up.

Recommendation 1.2: the top level landing page when searching on the internet for obs4MIPs should lead to a list of content, clearly outlined, subdivided by climate domain (atmosphere, cryosphere etc) and with links direct to the technical notes for each dataset

Recommendation 1.3: this top level landing page needs to be kept up to date frequently. CMUG has found through the interviews that out of data information is one of the main barriers to obs4MIPs use. Either out of data descriptions, or out of date datasets

Recommendation 2: Improved documentation and metadata

Recommendation 2.1: Technical notes to continue in current format but to include more links to detailed documentation

Recommendation 2.1: Technical notes to be linked from top level landing page/table of contents for obs4MIPs (see recommendation 1.2)

Recommendation 2.2: Technical notes to be live documents updated as quality issues come to light, or perhaps with link to data quality tables for each version as exemplified by LST_CCI

Recommendation 2.3: Standardised metadata and QC flags.

Recommendation 2.4: CF compliance is important, but if more information can be added to the metadata within this format this would be welcome. Minimum top level metadata should include DOI and version number for dataset. Adding grid box by grid box metadata would also be welcome.

Recommendation 2.5: Regular literature searches should be carried out by the observation providers or the obs4MIPs team and recent published peer reviewed papers which use the obs4MIPs datasets should be linked to the tech note to allow climate researchers to easily access information on how the data have been used.

Recommendation 2.6: A feedback form on the web page to allow users to raise issues that they encounter around access or data availability. This is only worthwhile if the feedback is read and acted on.

Recommendation 3: data access and manipulation tools should be provided for the obs4MIPs datasets.

Recommendation 3.1: All data and metadata formats should be completely consistent across obsMIPs.

Recommendation 3.2: ODS2.5[2] should be linked from each tech note and the link should be updated when the document is updated.

Recommendation 3.3: All live datasets should be updated once per year (preferably at the same time) to include the latest time period. All datasets (static or live) should have the ability to be updated quickly if bugs or errors are fixed or other improvements made.

Recommendation 3.4: Closer working relationships between obs4MIPs and ESMValTool should be built – potentially each observation provider could be required to provide a small recipe within ESMValTool to manipulate their data correctly providing a starting point for those doing analysis

Recommendation 3.5: subsetting tools should be provided to allow smaller data volumes to be downloaded, this should allow subsetting by spatial extent or by time period

Recommendation 3.6: regridding tools should be provided to correctly map the data (and uncertainties, see Section 5) to different resolutions and grids (e.g. rotated pole, unstructured).

Recommendation 3.6: better explanation of grid labels.

Recommendation 3.7: a cloud computing environment such as JASMIN would be most useful for those in the global south, to mitigate against the connectivity issues.

Recommendation 5: Obs4MIPs steering panel should identify key datasets and invite the producers to contribute.

Recommendation 6: obs4MIPs to encourage observation simulators to be developed through partnerships between observation providers and users and for obs4MIPs to facilitate the sharing of these to users who need them (through links from documentation, or hosting with other tools which have been recommended, recommendation 3).

Recommendation 7: Improve the provision of uncertainties and documentation on how to use them

Recommendation 7.1: Include both total uncertainty information with the data and full breakdown of uncertainty components with guidance on how to sum these to obtain the total. Also guidance on how to translate to different resolutions.

Recommendation 7.2: Prioritise the work on exploring options to include ancillary information, including uncertainties which has been assigned to obs4MIPs TT3.

Recommendation 7.3: have a series of optional additional 'variables' or 'supplementary/auxiliary' fields (dependent on the preferred nomenclature) that enable producers of observational datasets to include the most appropriate uncertainty information for their product.

Recommendation 7.4: scope of the accompanying technical note would be expanded, to enable the data provider to specify some more information about the total uncertainty and breakdown of uncertainty components, where these are provided.

Recommendation 7.5: a universal recipe explaining how to use observational uncertainty is provided by obs4MIPs for modellers.

7. Conclusions

This report summarises the results of 36 interviews with climate observation users involved in climate research. In addition to this a full gap analysis has been carried out and detailed thoughts and recommendations on the treatment of uncertainties with obs4MIPs have been discussed.

While best efforts were made to sample a good cross section of climate science in the interviews it was not easy to get participation from busy scientists and geographical limitations were an issue, therefore, this is more of a snapshot of current user opinions rather than a widespread sampling.

Nevertheless, CMUG believe that the cross section of users sampled is representative of the wider community and that action on the recommendations above will make an extremely positive impact on the uptake of the obs4MIPs datasets.

8. References

Ablain, M., Meysignac, B., Zawadzki, L., Jugier, R., Ribes, A., Spada, G., Benveniste, J., Cazenave, A., Picot, N. (2019). Uncertainty in satellite estimates of global mean sea-level changes, trend and acceleration. Earth System Science Data, 11, 1189-1202.

Bulgin, C. E., Embury, O., Corlett, G., Merchant, C. J. (2016a). Independent uncertainty estimates for coefficient based sea surface temperature retrieval from the Along-Track Scanning Radiometer. Remote Sensing of Environment, 178, 213-222.

Bulgin, C. E., Embury, O., Merchant, C. J. (2016b). Sampling uncertainty in gridded sea surface temperature products and Advanced Very High Resolution Radiometer (AVHRR) Global Area Coverage (GAC) data. Remote Sensing of Environment, 177, 287-294.

Bulgin, C. E., Green, P., Gruber, A., Macintosh, C., Mittaz, J., Ramsay, A., Rayner, N. A. (2024). The importance of scale in the definition of uncertainties: how do we best communicate this to data users? Surveys of Geophysics, in preparation.

Copernicus Catalogue – Climate Data Store, https://cds.climate.copernicus.eu/datasets. Accessed 08/12/2024.

Dorigo, W., Preimesberger, W., Stradiotti, P., Kidd, R., Van der Schalie, R., van der Vilet, M., Rodriguez-Fernandez, N., Madelon, R., Baghdadi, N. (2023). Algorithm Theoretical Basis Document (ATBD) supporting product version 08.1. ESA Climate Change Initiative Plus – Soil Moisture.

Ferraro, R.; Waliser, D.E.; Gleckler, P.; Taylor, K.E.; Eyring, V. Evolving Obs4MIPs to Support Phase 6 of the Coupled Model Intercomparison Project (CMIP6). *Bull. Am. Meteorol. Soc.* **2015**, *96*, Es131–Es133.

Ghent, D., Veal, K., Trent, T., Dodd, E., Sembhi, H., Remedios, J. (2019). A new approach to defining uncertainties for MODIS land surface temperature. Remote Sensing, 11.

Gleckler, P. J., Ferraro, R., Waliser, D. E. 2011: Improving use of satellite data in evaluation climate models. *Eos T. Am. Geophys. Un.*, 92, 172, https://doi.org/10.1029/2011eo200005, 2011.

Gleckler, P., Taylor, K. E., Durack, P. J., Nadeau, D., Biard, J. C., Elsaesser, G., Ferraro, R., Finkensieper, S., Hassler, B., Manaster, A., Mears, C., Pinnock, S., Stevens, S., Tuma, M., Turner, B., Waterfall, A. and Willet, K. (2024). Obs4MIPs Data Specifications version 2.5 (OSD2.5). 10.581/zenodo.11500474.

Joint Committee for Guides in Metrology (JCGM), (2008). Evaluation of measurement data – Guide to the expression of uncertainty in measurement. Organisation Internationale de Metrologie Legale.

Joint Committee for Guides in Metrology (JCGM), (2012). International vocabulary of metrology – Basic and general concepts and associated terms (VIM). 3rd Edition.

Merchant, C. J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., de Leeuw, G., Mittaz, J., Poulsen, C., Povey, A. C., Reuter, M., Sathyendranath, S., Sandven, S., Sofieva, V. K., Wagner, W. (2017). Uncertainty information in climate data records from Earth observation. Earth Syst. Sci. Data. 9, 511-527. https://doi.org/10.5194/essd-9-511-2017.

Mittaz, J., Merchant, C. J., Woolliams, E. R. (2019). Applying principles of metrology to historical observations from satellites. Metrologia, 56 (3).

Obs4MIPs Technical Note Guidance and Template, Obs4MIPs Steering Panel, https://zenodo.org/records/14276263. Accessed 10/12/2024.

Plummer, S., Lecomte, P., Doherty, M. (2018). The ESA Climate Change Initiative (CCI): A European contribution to the generation of the Global Climate Observing System. Remote Sensing of Environment, 203.

Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., Vegas-Regidor, J., Bock, L., Brötz, B., de Mora, L., Diblen, F., Dreyer, L., Drost, N., Earnshaw, P., Hassler, B., Koldunov, N., Little, B., Loosveldt Tomas, S., and Zimmermann, K. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0 – technical overview, *Geosci. Model Dev*., 13, 1179–1199, https://doi.org/10.5194/gmd-13-1179-2020

Teixeira, J.; Waliser, D.; Ferraro, R.; Gleckler, P.; Lee, T.; Potter, G. (2014). Satellite Observations for CMIP5 The Genesis of Obs4MIPs. *Bull. Am. Meteorol. Soc. 95*, 1329–1334 https://doi.org/10.1175/BAMS-D-12-00204.1

Waliser, D, Gleckler, P. J., Ferraro, R., Taylor, K. E., Ames, S., Biard, J., Bosilovich, M. G., Brown, O., Chepfer, H., Cinquini, L, Durack, P. J., Eyring, V., Mathieu, P-P., Lee, T., Pinnock, S., Potter, G. L., Rixen, M., Saunders, R., Schulz, J., Thepaut, J-N., and Tuma, M. (2020). Observations for Model Intercomparison Project (Obs4MIPs): status for CMIP6. Geosci. Model Dev., 13, 2945-2958.

9. Glossary

Terms	
Data assimilation	Observations directly influence the model initial state taking into account their error characteristics during every cycle of a model. This is used for reanalysis, NWP, which includes seasonal and decadal forecasting.

Model validation	Observations are compared with equivalent model fields to assess the accuracy of the model. This can be on short time scales for process studies or long time scales for climate trends.
Climate monitoring	This describes the use of a satellite only dataset to monitor a particular atmospheric or surface variable over a period > 15yrs to investigate whether there is a trend due to climate change.
Initialisation	To initialise prognostic quantities of the model with reasonable values at the beginning of the simulation but do not continuously update.
Prescribe boundary conditions	Prescribe boundary conditions for a model run for variables that are not prognostic (e.g. land cover, ice caps etc).
Accuracy	Accuracy is the measure of the non-random, systematic error, or bias, that defines the offset between the measured value and the true value that constitutes the SI absolute standard.
Stability	Stability is a term often invoked with respect to long-term records when no absolute standard is available to quantitatively establish the systematic error – the bias defining the time-dependent (or instrument-dependent) difference between the observed quantity and the true value.
Precision	Precision is the measure of reproducibility or repeatability of the measurement without reference to an international standard so that precision is a measure of the random and not the systematic error. Suitable averaging of the random error can improve the precision of the measurement but does not establish the systematic error of the observation.
Acronyms	
CCI	Climate Change Initiative
CDR	Climate Data Record
CMIP	Climate Model Intercomparison Project
CMIP6	Climate Model Intercomparison Project-6
CMIP7	Climate Model Intercomparison Project-7
CMUG	Climate Modelling Users Group
IPO	International Project Office
ISCCP	International Satellite Cloud Climatology Project
LAI	Leaf Area Index
NWP	Numerical Weather Prediction
PCMDI	Program for Climate Model Diagnosis and Intercomparison

10. Annex 1: Contributors

With thanks to all those who kindly agreed to be interviewed and contributed to this report.

Constantin Ardilouze, CNRM (France)
Angela Benedetti, ECMWF (UK)
Segolene Berthou, Met Office (UK)
Lisa Bock, DLR (Germany)
Alejandro Boda-Salcedo, Met Office, (UK)
Erasmo Buonomo, Met Office, (UK)

Jean-Christophe Calvet, Meteo-France, CNRM (France)
Juan Ignacio Caragunis, Department of Atmospheric and Oceanic Sciences, University of Buenos Aires (Argentina)
David Ford, Met Office (UK)
Baylor Fox-Kemper, Brown University (USA)
Nicola Gedney, Met Office (UK)
Dan Grosvenor, Met Office (UK)
Rafiq Hamdi, Royal Meteorological Institute of Belgium (Belgium)
Glen Harris, Met Office, (UK)
Andy Hartley, Met Office (UK)
Debbie Hemming, Met Office (UK)
Helene Hewitt, Met Office (UK)
Awel Haji Ibrahim, Arba Minch University (Ethiopia)
Axel Lauer, DLR (Germany)
Abhishek Lodh, SMHI (Sweden)
Rob King, Met Office (UK)
Elizaveta Malinina, Environment and Climate Change Canada (Canada)
Camilla Matheson, Met Office (UK)
Jane Mulcahy, Met Office (UK)
James Murphy, Met Office (UK)
Christine Nam, Climate Service Centre (Germany)
Tamzin Palmer, Met Office (UK)
Rob Parker, NCEO, University of Leicester (UK)
Charlotte Pascoe, CEDA (UK)
Ed Pope, Met Office (UK)
Eddy Robertson, Met Office (UK)
Steven Rumbold, Met Office (UK)
Kirsti Salonen, ECMWF (UK)
Christian Steger, DWD (Germany)
Hamish Steptoe, Met Office (UK)
Ulrika Willén, SMHI (Sweden)

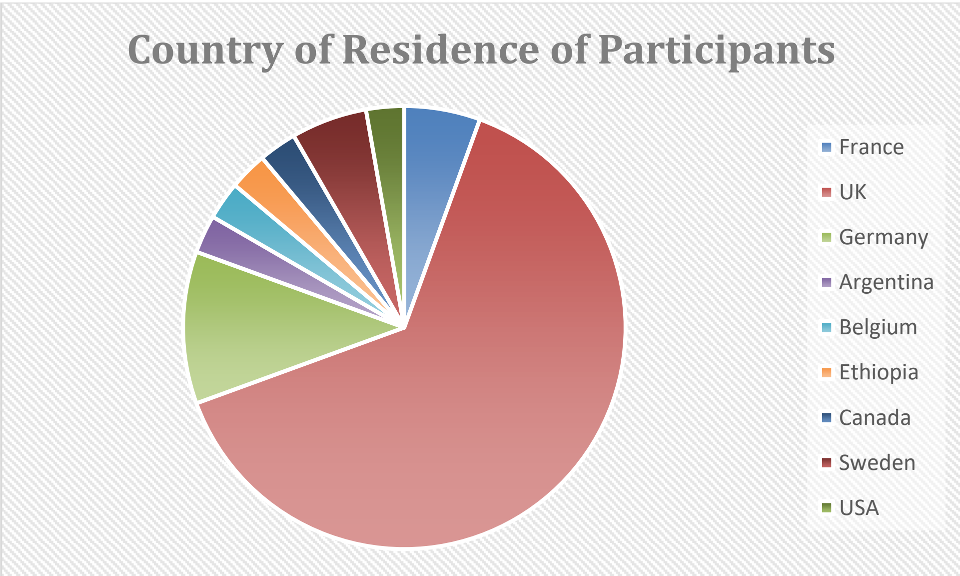


Figure A.1 country of residence of the interviewees

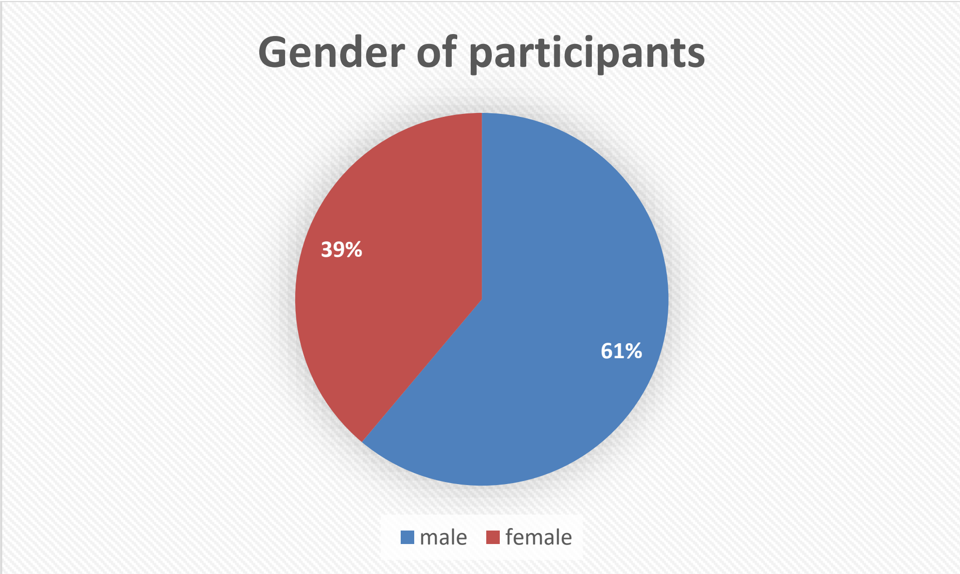


Figure A.2: Gender of interviewees

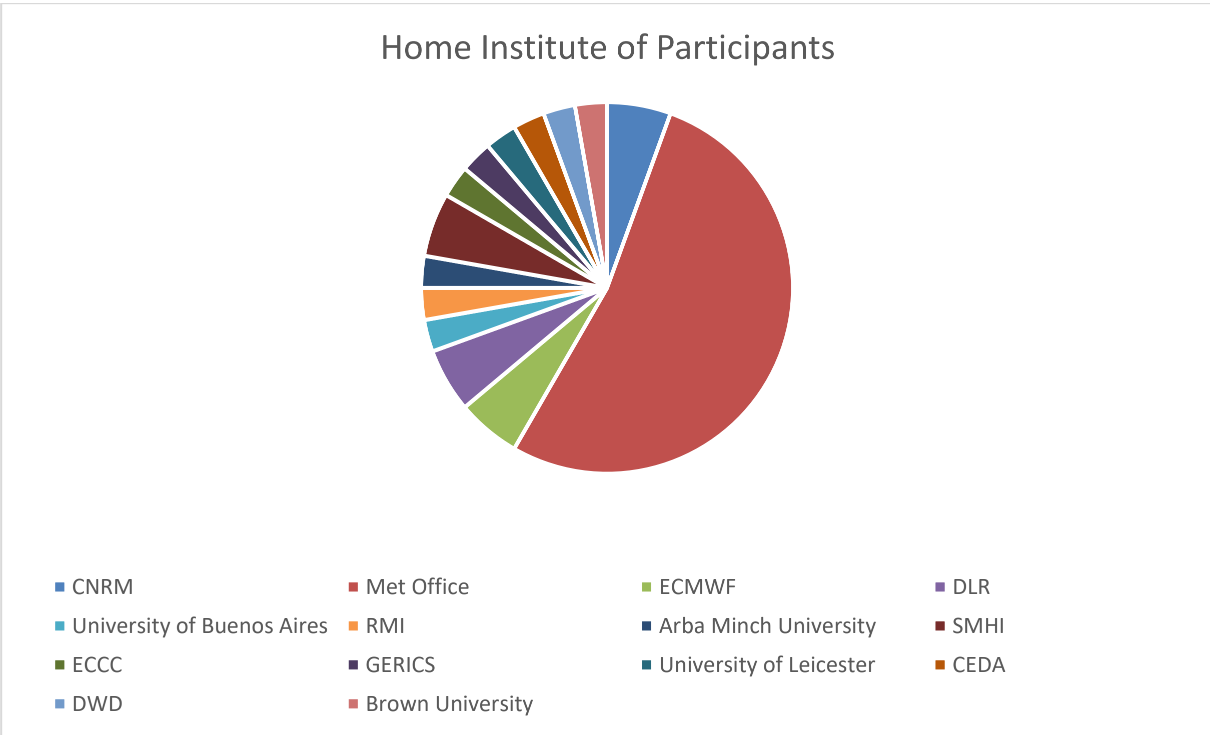


Figure A.3: home institute of interviewees

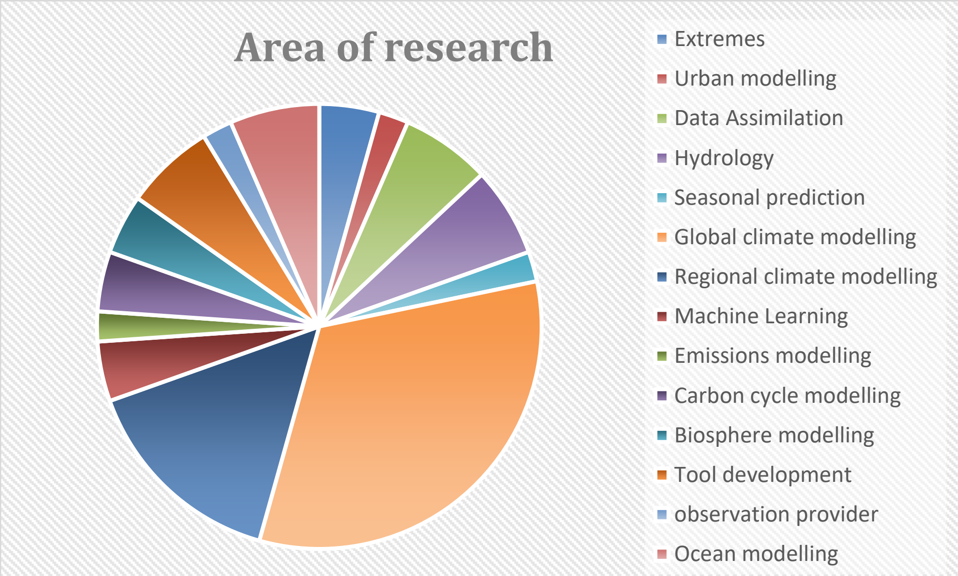


Figure A.4: Area of research of interviewees

Table A.1: Uses of observations by interviewees

downscaling
training machine learning models
model tuning
model bias correction
model evaluation
model baselining
model initialisation
constraining projections
using models to assess observation quality

11. Annex 2: Obs4MIPs datasets

Table A.2: List of Obs4MIPs datasets analysed for this report (as of 3/12/2024)

source_id	title	variable	variable_long_name	version
['AIRS-1-0']	obs4MIPs.NASA-JPL.AIRS-1-0.mon.hus.gn	['hus']	['Specific Humidity']	20110608
	obs4MIPs.NASA-JPL.AIRS-1-0.mon.ta.gn	['ta']	['Air Temperature']	20110608
['AIRS-2-0']	obs4MIPs.NASA-JPL.AIRS-2-0.mon.hur.gn	['hur']	['Relative Humidity']	20180307

	obs4MIPs.NASA-JPL.AIRS-2-0.mon.hus.gn	['hus']	['Specific Humidity']	20180307
	obs4MIPs.NASA-JPL.AIRS-2-0.mon.ta.gn	['ta']	['Air Temperature']	20180307
['AIRS-2-1']	obs4MIPs.NASA-JPL.AIRS-2-1.mon.hur.gn	['hur']	['Relative Humidity']	20201110
	obs4MIPs.NASA-JPL.AIRS-2-1.mon.ta.gn	['ta']	['Temperature']	20201110
	obs4MIPs.NASA-JPL.AIRS-2-1.mon.hus.gn	['hus']	['Specific Humidity']	20201110
['ARC-SST-1-1']	obs4MIPs.UOE.ARC-SST-1-1.mon	['tos']	['sea surface temperature']	1
['ATSR2-AATSR']	obs4mips.SU.ATSR2-AATSR.od550aer.mon	['crs', 'lat', 'lon', 'od550aer', 'time', 'time_bnds']	['', 'latitude', 'longitude', 'Ambient Aerosol Optical Thickness at 550 nm', 'time', '']	20160922
['Aura-MLS-v04-2']	obs4MIPs.NASA-JPL.Aura-MLS-v04-2.mon.cli.gn	['cli', 'cliNobs', 'cliStderr']	['Mass Fraction of Cloud Ice', 'Mass Fraction of Cloud Ice Number of Observations', 'Mass Fraction of Cloud Ice Standard Error']	20160504
	obs4MIPs.NASA-JPL.Aura-MLS-v04-2.mon.hus.gn	['hus']	['Specific Humidity']	20111025
	obs4MIPs.NASA-JPL.Aura-MLS-v04-2.mon.ta.gn	['ta']	['Air Temperature']	20111025
['AVISO-1-0']	obs4MIPs.CNES.AVISO-1-0.mon.zos.gn	['zos']	['Sea Surface Height Above Geoid']	20180305
['C3S-GTO-ECV-9-0']	obs4MIPs.DLR-BIRA.C3S-GTO-ECV-9-0.mon.toz.gn	['toz']	['Total Column Ozone']	20231115
['CALIPSO']	obs4MIPs.IPSL.CALIPSO.clcalipso.night	['clcalipso']	['CALIPSO Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.clccalipso.day	['clccalipso']	['CALIPSO Clear Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.cfad2Lidarsr532.night	['cfad2Lidarsr532']	['CALIPSO Scattering Ratio']	20110323

	obs4MIPs.IPSL.CALIPSO.cfadLidarsr532.day	['cfadLidarsr532']	['CALIPSO Scattering Ratio']	20110323
	obs4MIPs.IPSL.CALIPSO.cfadLidarsr532.mon	['cfadLidarsr532']	['CALIPSO Scattering Ratio']	20110323
	obs4MIPs.IPSL.CALIPSO.cfadLidarsr532.night	['cfadLidarsr532']	['CALIPSO Scattering Ratio']	20110323
	obs4MIPs.IPSL.CALIPSO.clcalipso.day	['clcalipso']	['CALIPSO Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.clcalipso.mon	['clcalipso']	['CALIPSO Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.cfad2Lidarsr532.day	['cfad2Lidarsr532']	['CALIPSO Scattering Ratio']	20110323
	obs4MIPs.IPSL.CALIPSO.cfad2Lidarsr532.mon	['cfad2Lidarsr532']	['CALIPSO Scattering Ratio']	20110323
	obs4MIPs.IPSL.CALIPSO.clhcalipso.night	['clhcalipso']	['CALIPSO High Level Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.cllcalipso.day	['cllcalipso']	['CALIPSO Low-Level Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.clhcalipso.day	['clhcalipso']	['CALIPSO High Level Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.clhcalipso.mon	['clhcalipso']	['CALIPSO High Level Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.clccalipso.mon	['clccalipso']	['CALIPSO Clear Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.clccalipso.night	['clccalipso']	['CALIPSO Clear Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.clmcalipso.night	['clmcalipso']	['CALIPSO Mid Level Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.clrcalipso.day	['clrcalipso']	['CALIPSO 3D Clear fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.cltcalipso.night	['cltcalipso']	['CALIPSO Total Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.uncalipso.day	['uncalipso']	['CALIPSO 3D Undefined fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.uncalipso.mon	['uncalipso']	['CALIPSO 3D Undefined fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.uncalipso.night	['uncalipso']	['CALIPSO 3D Undefined fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.clrcalipso.mon	['clrcalipso']	['CALIPSO 3D Clear fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.clrcalipso.night	['clrcalipso']	['CALIPSO 3D Clear fraction']	20110323

	obs4MIPs.IPSL.CALIPSO.cllcalipso.mon	['cllcalipso']	['CALIPSO Low-Level Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.cllcalipso.night	['cllcalipso']	['CALIPSO Low-Level Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.clmcalipso.day	['clmcalipso']	['CALIPSO Mid Level Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.clmcalipso.mon	['clmcalipso']	['CALIPSO Mid Level Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.cltcalipso.day	['cltcalipso']	['CALIPSO Total Cloud Fraction']	20110323
	obs4MIPs.IPSL.CALIPSO.cltcalipso.mon	['cltcalipso']	['CALIPSO Total Cloud Fraction']	20110323
['CERES-EBAF']	obs4MIPs.NASA-LaRC.CERES-EBAF.atmos.mon	['rlut', 'rlutcs', 'rsdt', 'rsut', 'rsutcs']	['TOA Outgoing Longwave Radiation', 'TOA Outgoing Clear-Sky Longwave Radiation', 'TOA Incident Shortwave Radiation', 'TOA Outgoing Shortwave Radiation', 'TOA Outgoing Clear-Sky Shortwave Radiation']	20160610
['CERES-EBAF_Surface']	obs4MIPs.NASA-LaRC.CERES-EBAF_Surface.atmos.mon	['rlds', 'rldscs', 'rlus', 'rsds', 'rsdscs', 'rsus', 'rsuscs']	['Surface Downwelling Longwave Radiation', 'Surface Downwelling Clear-Sky Longwave Radiation', 'Surface Upwelling Longwave Radiation', 'Surface Downwelling Shortwave Radiation', 'Surface Downwelling Clear-Sky Shortwave Radiation', 'Surface Upwelling Shortwave Radiation', 'Surface Upwelling Clear-Sky Shortwave Radiation']	20160610
['CERES-EBAF-4-2']	obs4MIPs.NASA-LaRC.CERES-EBAF-4-2.mon.rsuscs.gn	['rsuscs']	['Surface Upwelling Clear-Sky Shortwave Radiation']	20231205
	obs4MIPs.NASA-LaRC.CERES-EBAF-4-2.mon.rt.gn	['rt']	['Top of Atmosphere Net Radation']	20240513

	obs4MIPs.NASA-LaRC.CERES-EBAF-4-2.mon.rsdscs.gn	['rsdscs']	['Surface Downwelling Clear-Sky Shortwave Radiation']	20240513
	obs4MIPs.NASA-LaRC.CERES-EBAF-4-2.mon.rsutcs.gn	['rsutcs']	['TOA Outgoing Clear-Sky Shortwave Radiation']	20240513
	obs4MIPs.NASA-LaRC.CERES-EBAF-4-2.mon.rstcre.gn	['rstcre']	['Top of Atmosphere Shortwave CRE']	20240513
	obs4MIPs.NASA-LaRC.CERES-EBAF-4-2.mon.rsut.gn	['rsut']	['TOA Outgoing Shortwave Radiation']	20240513
	obs4MIPs.NASA-LaRC.CERES-EBAF-4-2.mon.rsus.gn	['rsus']	['Surface Upwelling Shortwave Radiation']	20240513
	obs4MIPs.NASA-LaRC.CERES-EBAF-4-2.mon.rsdt.gn	['rsdt']	['TOA Incident Shortwave Radiation']	20240513
	obs4MIPs.NASA-LaRC.CERES-EBAF-4-2.mon.rldscs.gn	['rldscs']	['Surface Downwelling Clear-Sky Longwave Radiation']	20240513
	obs4MIPs.NASA-LaRC.CERES-EBAF-4-2.mon.rsds.gn	['rsds']	['Surface Downwelling Shortwave Radiation']	20240513
	obs4MIPs.NASA-LaRC.CERES-EBAF-4-2.mon.rlds.gn	['rlds']	['Surface Downwelling Longwave Radiation']	20240513
	obs4MIPs.NASA-LaRC.CERES-EBAF-4-2.mon.rlutcs.gn	['rlutcs']	['TOA Outgoing Clear-Sky Longwave Radiation']	20240513
	obs4MIPs.NASA-LaRC.CERES-EBAF-4-2.mon.rlut.gn	['rlut']	['TOA Outgoing Longwave Radiation']	20240513
	obs4MIPs.NASA-LaRC.CERES-EBAF-4-2.mon.rltcre.gn	['rltcre']	['Top of Atmosphere Longwave CRE']	20240513
	obs4MIPs.NASA-LaRC.CERES-EBAF-4-2.mon.rlus.gn	['rlus']	['Surface Upwelling Longwave Radiation']	20240513
['CloudSat']	obs4MIPs.PCMDI.CloudSat.missingdatafraction.mon	['missingdatafraction']	['Missing data fraction due to the effects of ground clutter and surface elevation']	20130503
	obs4MIPs.PCMDI.CloudSat.overpasses.mon	['overpasses']	['Number of CloudSat Profiles Contributing to the Calculation']	20130503

	obs4MIPs.PCMDI.CloudSat.cfadDbze94.mon	['cfadDbze94']	['CloudSat Radar Reflectivity CFAD']	20130503
	obs4MIPs.PCMDI.CloudSat.cltcloudsat.mon	['cltcloudsat']	['CloudSat 94GHz radar Total Cloud Fraction']	20130503
['ERA-interim']	obs4MIPs.ECMWF.ERA-interim.atmos.mon	['ua', 'va']	['Eastward Wind', 'Northward Wind']	20160614
['ESACCI-CLOUD-ATSR2-AATSR-3-0']	obs4MIPs.DWD.ESACCI-CLOUD-ATSR2-AATSR-3-0.mon.clCCI.gr	['clCCI']	['CCI Cloud Area Fraction']	20200106
	obs4MIPs.DWD.ESACCI-CLOUD-ATSR2-AATSR-3-0.mon.clivi.gr	['clivi']	['Ice Water Path']	20200106
	obs4MIPs.DWD.ESACCI-CLOUD-ATSR2-AATSR-3-0.mon.cltCCI.gr	['cltCCI']	['CCI Total Cloud Fraction']	20200106
	obs4MIPs.DWD.ESACCI-CLOUD-ATSR2-AATSR-3-0.mon.clwCCI.gr	['clwCCI']	['CCI Liquid Cloud Area Fraction']	20200106
	obs4MIPs.DWD.ESACCI-CLOUD-ATSR2-AATSR-3-0.mon.clwtCCI.gr	['clwtCCI']	['CCI Total Liquid Cloud Area Fraction']	20200106
	obs4MIPs.DWD.ESACCI-CLOUD-ATSR2-AATSR-3-0.mon.clwvi.gr	['clwvi']	['Condensed Water Path']	20200106
	obs4MIPs.DWD.ESACCI-CLOUD-ATSR2-AATSR-3-0.mon.pctCCI.gr	['pctCCI']	['CCI Mean Cloud Top Pressure']	20200106
['ESACCI-CLOUD-AVHRR-AM-3-0']	obs4MIPs.DWD.ESACCI-CLOUD-AVHRR-AM-3-0.mon.clCCI.gr	['clCCI']	['CCI Cloud Area Fraction']	20190918
	obs4MIPs.DWD.ESACCI-CLOUD-AVHRR-AM-3-0.mon.clivi.gr	['clivi']	['Ice Water Path']	20190918
	obs4MIPs.DWD.ESACCI-CLOUD-	['cltCCI']	['CCI Total Cloud Fraction']	20190918

	AVHRR-AM-3-0.mon.cltCCI.gr			
	obs4MIPs.DWD.ESACCI-CLOUD-AVHRR-AM-3-0.mon.clwCCI.gr	['clwCCI']	['CCI Liquid Cloud Area Fraction']	20190918
	obs4MIPs.DWD.ESACCI-CLOUD-AVHRR-AM-3-0.mon.clwtCCI.gr	['clwtCCI']	['CCI Total Liquid Cloud Area Fraction']	20190918
	obs4MIPs.DWD.ESACCI-CLOUD-AVHRR-AM-3-0.mon.clwvi.gr	['clwvi']	['Condensed Water Path']	20190918
	obs4MIPs.DWD.ESACCI-CLOUD-AVHRR-AM-3-0.mon.pctCCI.gr	['pctCCI']	['CCI Mean Cloud Top Pressure']	20190918
['ESACCI-CLOUD-AVHRR-PM-3-0']	obs4MIPs.DWD.ESACCI-CLOUD-AVHRR-PM-3-0.mon.clCCI.gr	['clCCI']	['CCI Cloud Area Fraction']	20190918
	obs4MIPs.DWD.ESACCI-CLOUD-AVHRR-PM-3-0.mon.clivi.gr	['clivi']	['Ice Water Path']	20190918
	obs4MIPs.DWD.ESACCI-CLOUD-AVHRR-PM-3-0.mon.cltCCI.gr	['cltCCI']	['CCI Total Cloud Fraction']	20190918
	obs4MIPs.DWD.ESACCI-CLOUD-AVHRR-PM-3-0.mon.clwCCI.gr	['clwCCI']	['CCI Liquid Cloud Area Fraction']	20190918
	obs4MIPs.DWD.ESACCI-CLOUD-AVHRR-PM-3-0.mon.clwtCCI.gr	['clwtCCI']	['CCI Total Liquid Cloud Area Fraction']	20190918
	obs4MIPs.DWD.ESACCI-CLOUD-AVHRR-PM-3-0.mon.clwvi.gr	['clwvi']	['Condensed Water Path']	20190918
	obs4MIPs.DWD.ESACCI-CLOUD-AVHRR-PM-3-0.mon.pctCCI.gr	['pctCCI']	['CCI Mean Cloud Top Pressure']	20190918

['ESA-CCI-SST-v2-1']	obs4MIPs.UReading.ESA-CCI-SST-v2-1.mon.tos.gn	['tos']	['Sea Surface Temperature']	20201130
['GERB-HR-ED01-1-0']	obs4MIPs.ImperialCollege.GERB-HR-ED01-1-0.1hrCM.rlut.gn	['rlut']	['TOA Outgoing Longwave Radiation']	20200506
	obs4MIPs.ImperialCollege.GERB-HR-ED01-1-0.1hrCM.rsut.gn	['rsut']	['TOA Outgoing Shortwave Radiation']	20200506
['GERB-HR-ED01-1-1']	obs4MIPs.ImperialCollege.GERB-HR-ED01-1-1.1hrCM.rlut.gn	['rlut']	['TOA Outgoing Longwave Radiation']	20231221
	obs4MIPs.ImperialCollege.GERB-HR-ED01-1-1.1hrCM.rsut.gn	['rsut']	['TOA Outgoing Shortwave Radiation']	20231221
['GNSS-RO-1-3']	obs4MIPs.NASA-JPL.GNSS-RO-1-3.mon.zg.gn	['zg']	['Geopotential Height']	20160601
	obs4MIPs.NASA-JPL.GNSS-RO-1-3.mon.ta.gn	['ta']	['Air Temperature']	20160601
['GPCP-Daily-3-2']	obs4MIPs.NASA-GSFC.GPCP-Daily-3-2.mon.pr.gn	['pr']	['Precipitation']	20231205
['GPCP-Monthly-3-2']	obs4MIPs.NASA-GSFC.GPCP-Monthly-3-2.mon.pr.gn	['pr']	['Precipitation']	20231205
['GPCP-V1.2']	obs4MIPs.NASA-GSFC.GPCP-V1.2.atmos.day	['pr']	['Precipitation']	20180518
['GPCP-V1.3']	obs4MIPs.NASA-GSFC.GPCP-V1.3.atmos.day	['pr']	['Precipitation']	20180519
['GPCP-V2.2']	obs4MIPs.NASA-GSFC.GPCP-V2.2.atmos.mon	['pr', 'prStderr']	['Precipitation', 'Precipitation Standard Error']	20180518
['GPCP-V2.3']	obs4MIPs.NASA-GSFC.GPCP-V2.3.atmos.mon	['pr']	['Precipitation']	20180519
['IMERG-v06B-Final']	obs4MIPs.NASA-GSFC.IMERG-v06B-Final.mon.pr.gn	['pr']	['Precipitation']	20240408

	obs4MIPs.NASA-GSFC.IMERG-v06B-Final.3hr.pr.2x2	['pr']	['Precipitation']	20210812
['ISCCP']	obs4MIPs.PCMDI.ISCCP.clisccp.mon	['clisccp']	['ISCCP Cloud Area Fraction (Joint histogram of optical thickness and cloud top pressure)']	20130503
	obs4MIPs.PCMDI.ISCCP.cttisccp.mon	['cttisccp', 'cttisccpunweighted']	['ISCCP Mean Cloud Top Temperature (Cloud-fraction weighted & daytime only)', 'ISCCP Mean Cloud Top Temperature (Unweighted & daytime only)']	20130503
	obs4MIPs.PCMDI.ISCCP.pctisccp.mon	['pctisccp', 'pctisccpunweighted']	['ISCCP Mean Cloud Top Pressure (Cloud-fraction weighted & daytime only)', 'ISCCP Mean Cloud Top Pressure (Unweighted, daytime only)']	20130503
	obs4MIPs.PCMDI.ISCCP.albisccp.mon	['albisccp', 'albisccpunweighted']	['ISCCP Mean Cloud Albedo (Cloud-fraction weighted & daytime only)', 'ISCCP Mean Cloud Albedo (Unweighted & daytime only)']	20130503
	obs4MIPs.PCMDI.ISCCP.cltisccp.mon	['cltisccp']	['ISCCP Total Cloud Fraction (daytime only)']	20130503
['MISR']	obs4MIPs.NASA-GSFC.MISR.atmos.mon	['od550aer', 'od550aerNobs', 'od550aerStdv']	['Ambient Aerosol Optical Thickness at 550 nm', 'Ambient Aerosol Optical Thickness at 550 nm Number of Observations', 'Ambient Aerosol Optical Thickness at 550 nm Standard Deviation']	20160614
	obs4MIPs.UW.MISR.clMISR.mon	['clMISR']	['Cloud Fraction retrieved by MISR']	20131113
	obs4MIPs.UW.MISR.samplesMISR.mon	['samplesMISR']	['Number of MISR Samples']	20131113
['MODIS']	obs4MIPs.NASA-GSFC.MODIS.atmos.mon	['fpar', 'od550aer', 'od550aer	['Fraction of Absorbed Photosynthetically Active Radiation',	20160614

		Nobs', 'od550aerStdv']	'Ambient Aerosol Optical Thickness at 550 nm', 'Ambient Aerosol Optical Thickness at 550nm Number of Observations', 'Ambient Aerosol Optical Thickness at 550nm Standard Deviation']	
	obs4MIPs.NASA-GSFC.MODIS.land.mon	['lai']	['Leaf Area Index']	20160614
['MODIS-1-0']	obs4MIPs.NASA-GSFC.MODIS-1-0.mon.clt.gn	['clt']	['Total Cloud Fraction']	20180305
['OISST']	obs4MIPs.NCEI.OISST.tos.mon	['tos']	['Sea Surface Temperature']	20160401
	obs4MIPs.NCEI.OISST.tos.day	['tos']	['Sea Surface Temperature']	20160401
['OLR']	obs4MIPs.NCEI.OLR.rlut.mon	['rlut']	['TOA Outgoing Longwave Radiation']	20160401
	obs4MIPs.NCEI.OLR.rlut.day	['rlut']	['TOA Outgoing Longwave Radiation']	20160401
['PARASOL']	obs4MIPs.LOA-IPSL.PARASOL.parasolRefl.day	['parasolRefl']	['PARASOL Reflectance']	20110323
	obs4MIPs.LOA-IPSL.PARASOL.parasolRefl.mon	['parasolRefl']	['PARASOL Reflectance']	20110323
	obs4MIPs.LOA-IPSL.PARASOL.sza.day	['sza']	['solar zenith angle']	20110323
	obs4MIPs.LOA-IPSL.PARASOL.sza.mon	['sza']	['solar zenith angle']	20110323
['PMSIC']	obs4MIPs.NCEI.PMSIC.SH.sic.day	['sic']	['Sea Ice Area Fraction']	20160401
	obs4MIPs.NCEI.PMSIC.NH.sic.day	['sic']	['Sea Ice Area Fraction']	20160401
	obs4MIPs.NCEI.PMSIC.SH.sic.mon	['sic']	['Sea Ice Area Fraction']	20160401
	obs4MIPs.NCEI.PMSIC.NH.sic.mon	['sic']	['Sea Ice Area Fraction']	20160401
['QuikSCAT-v20110531']	obs4MIPs.NASA-JPL.QuikSCAT-v20110531.mon.sfcWind.gn	['sfcWind']	['Near-Surface Wind Speed']	20120411

	obs4MIPs.NASA-JPL.QuikSCAT-v20110531.mon.uas.gn	['uas']	['Eastward Near-Surface Wind']	20120411
	obs4MIPs.NASA-JPL.QuikSCAT-v20110531.mon.vas.gn	['vas']	['Northward Near-Surface Wind']	20120411
['RSS-PRW-v07r02']	obs4MIPs.RSS.RSS-PRW-v07r02.mon.prw.gn	['prw']	['Water Vapor Path']	20231205
['RSS-v7']	obs4MIPs.RSS.RSS-v7.mon.prw.gn	['latitude_bounds', 'longitude_bounds', 'prw', 'time_bounds']	['Latitude_bounds', 'Longitude_bounds', 'monthly average atmosphere water vapor content over ice-free oceans', 'time_bounds']	20180305
	obs4MIPs.RSS.RSS-v7.mon.tos.gn	['tos']	['Sea Surface Temperature']	20180305
	obs4MIPs.RSS.RSS-v7.mon.sfcWind.gn	['latitude_bounds', 'longitude_bounds', 'sfcWind', 'time_bounds']	['Latitude_bounds', 'Longitude_bounds', 'Monthly Average Near-Surface Wind Speed', 'time_bounds']	20180305
['SSMI-MERIS']	obs4MIPs SSMI-MERIS Water Vapor Path L3 Monthly Data	['prw']	['Water Vapor Path']	20140616
['TES-1-0']	obs4MIPs.NASA-JPL.TES-1-0.mon.tro3.gn	['tro3']	['Mole Fraction of O3']	20110608
['TRMM']	obs4MIPs.NASA-GSFC.TRMM.atmos.3hr	['pr']	['Precipitation']	20160613
	obs4MIPs.NASA-GSFC.TRMM.atmos.mon	['pr', 'prStderr']	['Precipitation', 'Precipitation Standard Error']	20160613
['XCH4_CRDP3']	obs4MIPs.IUP.XCH4_CRDP3.xch4.mon	['xch4']	['column-average dry-air mole fraction of atmospheric methane']	100
['XCO2_CRDP3']	obs4MIPs.IUP.XCO2_CRDP3.xco2.mon	['xco2']	['column-average dry-air mole fraction of atmospheric carbon dioxide']	100

12. Annex 3: Interview responses on barriers to the use of obs4MIPs

obs4MIPs barriers	Number of reports
Datasets not up to date	7
downloading datasets/dataset size	5
already have sources of data/work directly with data providers	4
navigating through the datasets	3
Needs to be clear about USP = what is it trying to do?	3
Poor documentation (non-existent)	3
Not as convenient as reanalysis	3
need higher temporal res	2
search for link did not take straight to the correct page	2
slow internet connection	2
steep learning curve for platform use	2
unfamiliar terminology	2
No contents list	2
Thought it only contained atmospheric and ocean variables, not land	2
Takes too long and too much effort for data providers to update their versions or even get it on there in the first place	2
Some data will never fit into o4m format	2
Lack of publicity	2
data licenses	2
o4m grid labels hard to understand, not intuitive	1
difficult to understand structure	1
Lack of tools to manipulate the data	1
No quality analysis or indication of which data to use for which application	1
4 different versions of CCI LC on obs4M	1
no good for extremes	1
MIP standard not flexible enough	1
need higher spatial res	1
Format not useful for GHG data users	1
No clear way to share uncertainties	1
Gridded data can't use averaging kernels	1
Need more guidance on how to use the data	1
It should be required for participation in CMIP	1
QC and comparison of data would be useful	1
Not always CMORised correctly	1
Metadata not well supported in CF conventions	1
unwieldy variable names not standard ncdf names	1
Need L1 or L2 data	1
need in situ data	1
ocean grids are often very different	1