IS-ENES3 Deliverable D9.1

ESMValTool version supporting irregular and unstructured grids

Reporting period: 01/07/2020 – 31/12/2021

Authors: Jan Griesfeller, Klaus Zimmermann, Alok Kumar Gupta, Mats Bentsen, Javier Vegas

Reviewer(s): Bryan Lawrence, Italo Epicoco
Release date: 27/11/2020

ABSTRACT

This report covers recent developments of the ESMValTool within the IS-ENES3 project. Covered topics are the support of irregular and unstructured grids, the current capabilities for ingestion of measurements including a list of supported data sets for user guidance and some information on provenance of observational data is handled.
A first release on the second major version of the ESMValTool has been made. Main general updates are a switch to a new a version of the Python programming language (Python 3), a renewed command line interface and a much more modern and maintainable software design.

Dissemination Level		
PU	Public	X
CO	Confidential, only for the partners of the IS-ENES3 project	

Revision table			
Version	**Date**	**Name**	**Comments**
Release for review	25/09/2020	J. Griesfeller	Iterations needed with NORCE following the reviewer's comments
Final version	27/11/2020	J. Griesfeller	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824084

Table of contents

Executive Summary

This report covers recent developments of the ESMValTool to version 2 within the IS-ENES3 project. Covered topics are the support of irregular and unstructured model grids from ocean models, the current capabilities for ingestion of measurements including a list of supported data sets for user guidance and some information on provenance of observational data is handled.

A first release on the second major version of the ESMValTool has been published. Main general updates are a switch to a new a version of the Python programming language (Python 3), a renewed command line interface and a much more modern and maintainable software design.

This report documents that the objectives in the workplan for the first period have mostly been reached (pilot support of ocean grids, easier way to download observational data, version information of observations used is now in the general provenance information). But it also shows room for improvement. Not all ocean grids can be used without errors and support of point based observational data is still missing and will not arrive short term.

Nevertheless, the overall targets of the deliverables have been reached.

1. Objectives

1.1. Pilot support for ocean data on original grids

A current need for the ESMValTool is to improve the handling of curvilinear and unstructured ocean grids and models with Lagrangian vertical coordinate (e.g. MICOM of NorESM, MOM6 of GFDL (and soon NCAR), HYCOM of FSU and GISS, AWI-CM, ICON), as ESMValTool does currently not support them. Pilot support shall be added to the Iris based preprocessor, and made usable from ESMValTool via the backend that is developed in Task 1 [D9.1]

1.2. Providing observations to the user

Despite the recent efforts (e.g., obs4mips, ana4mips) to deliver observations and reanalysis in CMIP format, the vast majority of observations are still released without compliance to any specific standard. The ESMValTool provides the user with reformatting routines but users still need to manually download the data from the corresponding sources and apply the reformatting routines. The goal of this task is to automate the process of accessing the original data source, downloading and processing the observational data in the correct format. The design of this automated procedure should also allow adding new observation sources in an easy way [D9.1].

1.3. Version documentation of utilized observations

Reprocessing of observational data is motivated by better data understanding but creates moving targets for model evaluation. Increasingly DOIs, better metadata and other data tagging mechanisms are implemented to allow for transparent model evaluation by attaching version information to observational datasets. These version numbers need to be carried through from the data source to the diagnostics and diagnostic products of the ESMValTool [D9.1].

2. Methodology and Results

2.1. Pilot support for ocean data on original grids

In ESMValTool, support for reading irregular grids (tri-polar grids) is implemented and interpolation/regridding methods for irregular grids (tripolar grids) are also implemented. These interpolation methods support three techniques: bilinear, nearest neighbour and area weighted (also called first order conservative). These can be called in the *recipe* in the *preprocessors* within the section *regrid* using scheme *linear*, *nearest* and *area_weighted*, respectively.

Partial example recipe:

```
 preprocessors:
   prep_timeseries:
     regrid:
       target_grid: 1x1
       scheme: linear
```

Regridding code was found in the esmvalcore repository (https://github.com/ESMValGroup/ESMValCore) at the file *esmvalcore/preprocessor/_regrid_esmpy.py* (*https://github.com/ESMValGroup/ESMValCore/blob/master/esmvalcore/preprocessor/_regrid_esmpy.py*)

In addition, there's support for irregular grids without interpolating to the target grid. Using *fx_variables* which uses arguments *'areacello'*, - area of the grid cells and *'volcello'* - volume of grid cells, one can plot the various statistics and generate the plot for various types. *fx_variables* can be called in the *recipe* in the *preprocessors* within the section area_statistics.
Partial example receipe:

```
preprocessors:
  prep_timeseries:
    area_statistics:
      fx_variables: ['areacello']
```

These techniques have been tested thoroughly for NEMO (EC-Earth's ocean model) and MICOM (NorESM's ocean model) tripolar grids for CMORized variables of CMIP5 data. The following datasets for irregular grids were tested (listed are filenames that can be found at ESGF):

tos_Omon_EC-EARTH_historical_r1i1p1_199101-200512.nc
tos_Omon_NorESM1-M_historical_r1i1p1_185001-200512.nc

All tests are performed for the simulation years 2001 to 2004.

1.2.1. NorESM1-M (MICOM)

Table 1 shows a monthly timeseries of global mean SST for the NorESM1-M model interpolated to a 1x1 degree target grid using the dataset from the file *tos_Omon_NorESM1-M_historical_r1i1p1_185001-200512.nc* for the years 2001 to 2004. The results of three interpolation techniques bilinear, nearest neighbour and area weighted are plotted in figure 1 to figure 3. figure 4 represents comparison of these three schemes with the not interpolated grid cell area method.

Table 1: Monthly timeseries of global mean SST of the NorESM1-M model

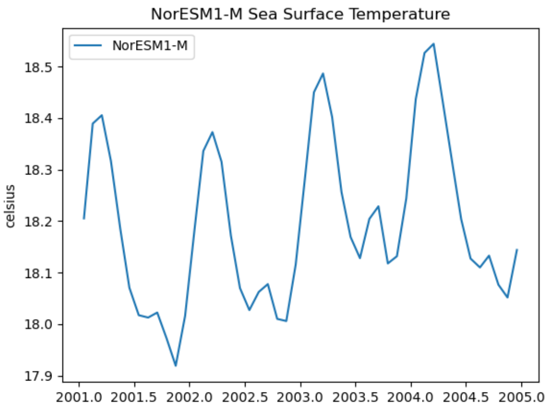


figure 1: bilinear

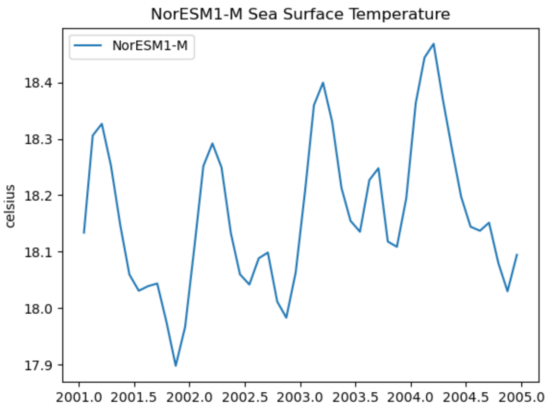


figure 2: nearest neighbour

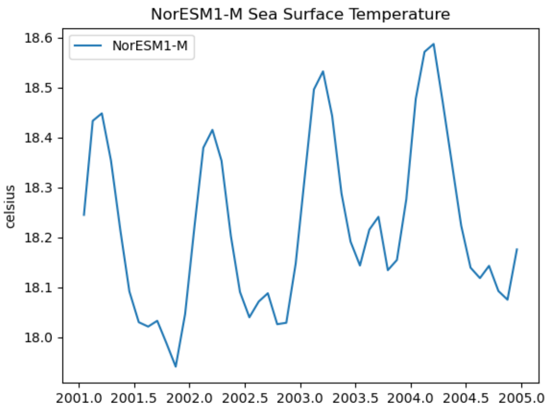


figure 3: area weighted

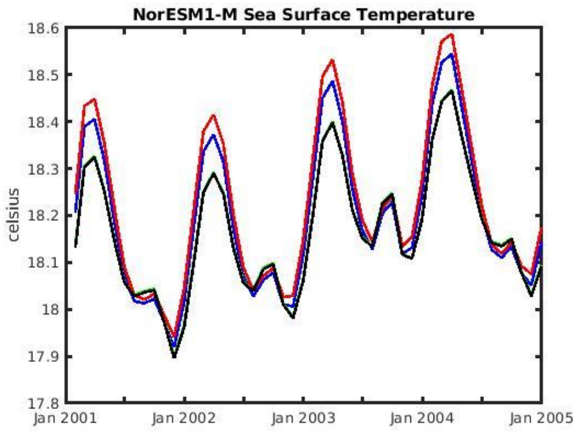


figure 4:Comparison of all three schemes with cell area
black: cell area method (not interpolated)
red: area weighted scheme
blue: bilinear scheme
green: nearest neighbour scheme

There is not a very significant difference between the three schemes of around 0.1 degrees centigrade at the maxima. This 0.1 degrees, for global average temperature seems to be very high.

But how these sea-land fractional area within a cell treated by various interpolation scheme always remains a challenge. Same results are noticed using tool CDO interpolation scheme.

1.2.2. EC-EARTH (NEMO)

Table 2 shows plots of monthly global SST averages derived using different interpolation schemes for the EC_EARTH model for the years 2001 to 2004. Figure 5, Figure 6 and Figure 7 represent these data respectively for bilinear, nearest neighbour and area weighted interpolation. The bilinear and nearest neighbour schemes are more or less identical as for the NorESM1-M model, but the area weighted scheme leads to about 1.5 degrees centigrade higher values (note the values on the y-axis).

There are some issues with NEMO grid interpolation and therefore, we have not performed comparison of various schemes for NEMO like MICOM as it will not be realistic. Presently, it is not advisable to use any regridded scheme for NEMO tripolar grid at least which is originating from IPSL and EC-EARTH model. A git issue has been opened at ESMValCore: https://github.com/ESMValGroup/ESMValCore/issues/863.

Table 2: Monthly timeseries of global mean SST of the EC-EARTH model

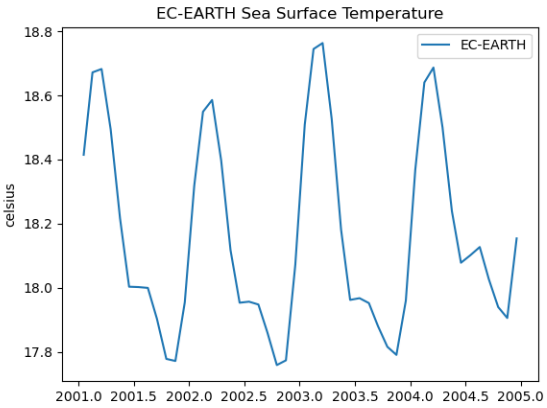


Figure 5: EC-EARTH model, bilinear interpolation

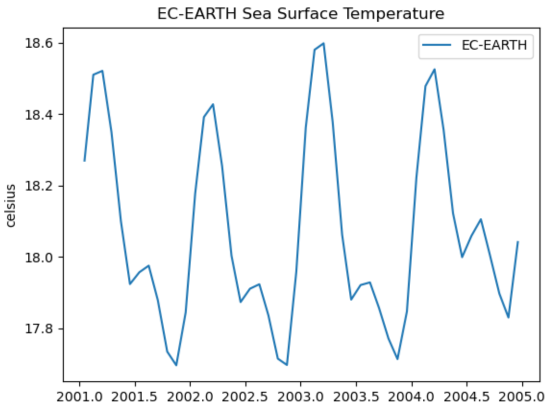


Figure 6: EC-EARTH model, nearest neighbour interpolation

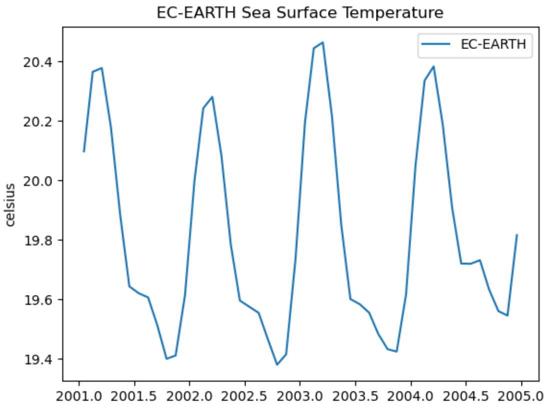


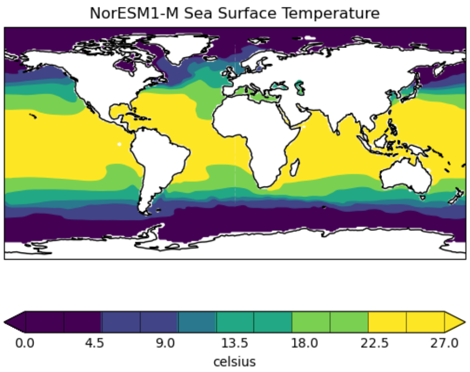
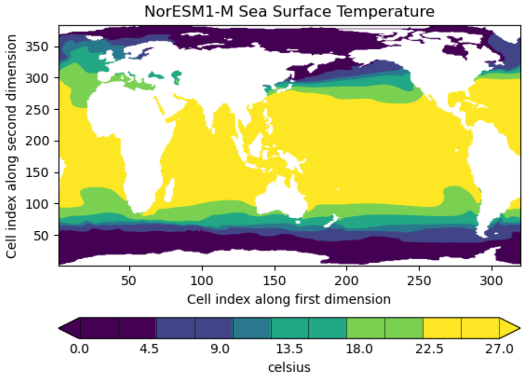
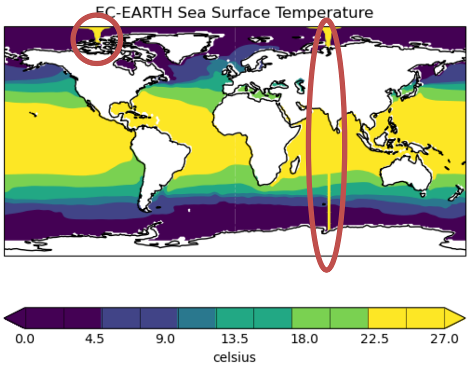
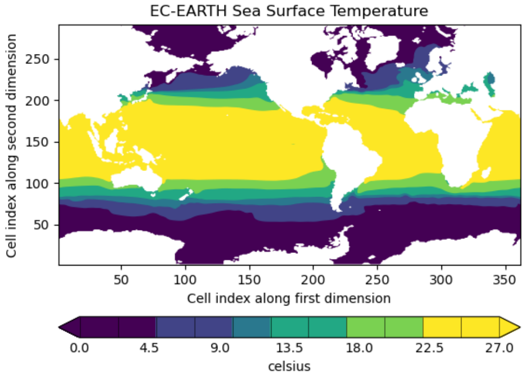
Figure 7: EC_EARTH model, area weighted interpolation

1.2.3. Maps

In this paragraph maps of one model month have been analyzed. Table 3 shows maps of the variable *tos* of one month of data of the NorESM1-M model (top) and the EC-EARTH model (bottom) interpolated using the area weighted interpolation scheme to a one by one degree grid (left hand side) and the original grid of the models (right hand side).

The figures show that the ESMValTool does support the native grids as well as interpolation NoeESM1-M's native grid using the area weighted interpolation scheme. Unfortunately, 10 shows that the area weighted interpolation does not work for the EC-EARTH model. Note the high value areas in the far north and the stripe of high values in the southern hemisphere (circled in red).

Table 3:Maps of SST interpolated and not interpolated

 NorESM1-M Sea Surface Temperature 0.0 4.5 9.0 13.5 18.0 22.5 27.0 celsius *Figure 8: Map of NorESM1-M area weighted scheme*	 NorESM1-M Sea Surface Temperature Cell index along second dimension 350 300 250 200 150 100 50 50 100 150 200 250 300 Cell index along first dimension 0.0 4.5 9.0 13.5 18.0 22.5 27.0 celsius *Figure 9: NorESM1-M no interpolation*
 EC-EARTH Sea Surface Temperature 0.0 4.5 9.0 13.5 18.0 22.5 27.0 celsius *Figure 10: EC-EARTH area weighted scheme*	 EC-EARTH Sea Surface Temperature Cell index along second dimension 250 200 150 100 50 50 100 150 200 250 300 350 Cell index along first dimension 0.0 4.5 9.0 13.5 18.0 22.5 27.0 celsius *Figure 11: EC-EARTH no interpolation*

Investigation of the EC-EARTH data shows that the problem lies in the source data. There is duplication along the edges, i.e. the two side edges overlap by 2 cells and the top edge overlaps with itself by one cell. The same behavior is found for NEMO grid with IPSL model. Since the data used in this comparison is the native NEMO data format, a fix is needed likely at the level of the iris Python package (https://github.com/SciTools/iris; used by the ESMValTool to read model data). Providing a fix is out of the scope of this deliverable report.

1.2.4. Comparison between nearest neighbour and area weighted interpolation

The following paragraph compares the results of the nearest neighbour and the area weighted interpolation schemes of the NorESM1-M model using the same data as in the earlier paragraphs.

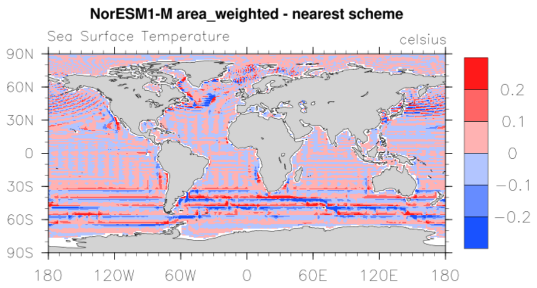


Figure 12: difference between area weighted and nearest neighbour interpolation

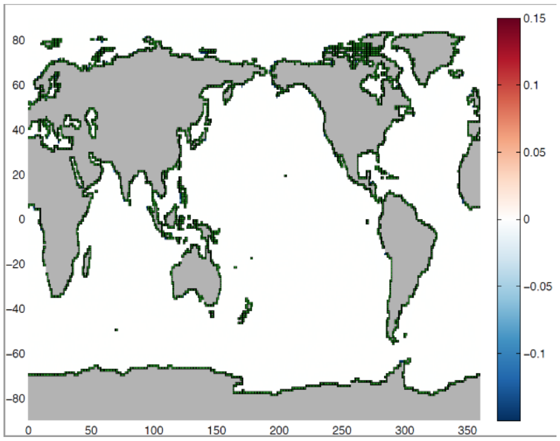


Figure 13: difference between area weighted scheme using ESMValTool and MATLAB interpolation

Figure 12 shows the difference between the nearest neighbour and area weighted interpolation scheme for the average global SST for the period from 2001 to 2004. A perfect match would produce an all green plot at all areas over ocean. The figure shows this besides some artefacts the interpolations produce identical results within a very small margin. Figure 13 shows, differences between first-order conservative scheme using MATLAB (self-created scripts) and ESMValTool. Green points represents points not considered by ESMValTool while we were able to interpolate on those using MATLAB.

2.2. Providing observations to the user

The managing of observational datasets in ESMValTool had several weaknesses that have to be addressed to make the user experience smoother. The work exposed on this deliverable, which is mostly contained in https://github.com/ESMValGroup/ESMValTool/pull/1657 (in approval process at the time of waiting this report), focuses in those two that were deemed as most critical: discoverability and automation of downloads.

At the current release of ESMValTool, the users are supposed to use for formatting an executable called *cmorize_obs*, a name that hides its relation with ESMValTool, but also to browse through the documentation to learn about which datasets are available and even to search in the formatting source code how to download them. All of this together makes it really impossible for a user to easily discover which datasets are available, how to get them and, in general, provide a bad experience when trying to get observational datasets to work with the ESMValTool.

To improve discoverability, a rework of ESMValTool CLI was submitted, (see https://github.com/ESMValGroup/ESMValCore/pull/605) so developers can add functionality to the *esmvaltool* command from any of the related packages. This allowed to integrate all the downloading and formatting functionalities into the *esmvaltool data* subcommand, so users can learn about it just calling *esmvaltool –help* on the command line. In addition to that, two informative commands were added:

- *esmvaltool data list*, which shows a list of all the supported datasets along with some basic info. Appendix 4.1 shows the output of this command at the time of writing this report.
- *esmvaltool data info DATASET_NAME*, which shows detailed info about a specific one, including how to download the data.

With these new functionalities, users can learn which datasets are available and how to get them just from using their installation of ESMValTool.

The second objective was to automatize the download procedure and to make it easy to adapt and extend to new datasets. In order to make it as flexible as possible, each supported dataset requires its own download script, but a set of helper classes has been created to manage most of the internal details of the connections. As a first step, five helper classes have been created to serve as an example on how to manage some common scenarios:

- An FTP downloader class and a WGET downloader class to handle downloads from standard FTP and HTTP servers
- A Climate Data Store downloader using cdsapi library to show how to integrate custom APIs
- Custom ESACCI and NASA downloaders to show how to derive from the basic downloader classes to simplify adding multiple datasets from a common source.

Using this approach, a new developer can start by looking at any of the working examples, choose one that is similar to its need and adapt it to its need. This approach has proved to be useful, as this work provides more automatic downloaders than expected at the start due to the easy development process (see Appendix 4.1)

Users can access both the automatic download and format functionalities from *esmvaltool data*:

- Download: *esmvaltool data download DATASET-NAME*
- Format: *esmvaltool data format DATASET-NAME*
- Download and format in one go: *esmvaltool data prepare DATASET-NAME*

At the time of the writing of this report, only gridded datasets were supported by the ESMValTool. Nevertheless, support of point-based station data will be included in the future. The way this will be achieved has caused some discussion within the user community of the ESMValTool during the run of the project. The basic choices were defining a NetCDF based data format for point-based observations and doing colocation at the ESMValTool level or using an existing external tool that will do all the needed tasks.
The former has been proposed in the ESMValTool github issue 496 (https://github.com/ESMValGroup/ESMValTool/issues/496), the latter in the github issue 1655 (https://github.com/ESMValGroup/ESMValTool/issues/1655).

After some discussion the ESMValTool community decided that the CIS tools (http://www.cistools.net/) will be used to integrate point based observational data. This approach is likely more general, in the long term less maintenance intensive and potentially more powerful. But in the short term less features are added to the ESMValTool because integrating the CIS tools into the ESMValTool will take some time and can only be done by the core programmers.

2.3. Version documentation of utilized observations

In order to be able to carry over available version information from the data source to the ESMValTool, the information has to be identified in the data products.

In the observational products in the CF/CMOR format (obs4mips) this is done by using the global attribute named ‘version’. There is no common rule of how this version number has to be formatted, but the most common practice is to put at least some sort of date string into this attribute. This version information is then also put into the diagnostic’s provenance file.

For other observations the authors of the cmorizer scripts are asked to provide a version number. If they carry over the version of the data or just the version of the cmorizer script lies in their responsibility. This version number is also put into the diagnostic's provenance file together with the source URL of the original data.

3. Conclusions and Recommendations

3.1. Pilot support for ocean data on original grids

Section 2.1 of this report has shown that the first release of the ESMValTool version 2 does support data from ocean models on their original grids. The problems shown there interpolating to a rectangular grid using data of the NEMO model needs to be investigated further. Likely this has to be done by persons more involved in the development than the ones contributing to this report. To facilitate this, a bug report has been created as a github issue. https://github.com/ESMValGroup/ESMValCore/issues/863

3.2. Providing observations to the user

That the stable synda version still uses the deprecated Python 2 makes the usage especially together with the ESMValTool (which is using Python 3) overly complicated. According to synda's github issue 111 (https://github.com/Prodiguer/synda/issues/111) a Python 3 based version should have been finished by now. Publishing a Python 3 based version of synda will make its usage much easier in current computer environments. Releasing that is therefore highly encouraged.

3.3. Version information of utilized observations

The version information of utilized observations kept in the provenance information of the ESMValTool has reached a helpful level. Nevertheless, for the cmorizer scripts, the version info in principle consists of two parts: the version information of the data and the version information of the cmorizer script. Right now, there are no rules about how to form this version information, with the effect that this represents mostly the version of the cmorizer script. Because a cmorizer script can be used as long as the native data format of an observation is not changed, its version number says little about the version number of the observations.

For observations with many updates (e.g. AERONET updates once a week, but also satellite data might have updates once a year) the information of the cmorizer script is much less interesting than the revision date of the data or the date of the data has been downloaded. The author of this report therefore recommends to put more emphasis on the data versions in addition to the version of the cmorizer script.

The CMIP (model) data on ESGF does not contain a standardized version information in the data files at all, although there are also different data versions. The versioning is only present in the file path information of some of the data structures used by the ESMValTool (DKRZ and Jasmin) and therefore only visible, if a user uses the same path structure at her site. The author of this report can’t see a reason why the version formation of an observational dataset has to carry the version information in the data files while the same version information is not present in the model files. This should be unified to both containing a common versioning string with a standardized set of information.

4. Appendix

4.1. Output of *esmvaltool data list*

```
| Dataset name                   | Tier | Auto-download | Last access |
-----------------------------------------------------------------------
| APHRO-MA                       |    3 | Yes           |  2020-03-06  |
| AURA-TES                       |    3 | Yes           |  2018-12-08  |
| BerkeleyEarth                  |    2 | Yes           |  2020-02-25  |
| CALIPSO-GOCCP                  |    2 | Yes           |  2020-01-27  |
| CDS-SATELLITE-ALBEDO           |    3 | Yes           |  2019-04-01  |
| CDS-SATELLITE-LAI-FAPAR        |    3 | Yes           |  2019-07-03  |
| CDS-SATELLITE-SOIL-MOISTURE    |    3 | Yes           |  2019-03-14  |
| CDS-UERRA                      |    3 | Yes           |  2019-11-04  |
| CDS-XCH4                       |    3 | Yes           |  2019-03-11  |
| CDS-XCO2                       |    3 | No            |  2019-03-19  |
| CERES-EBAF                     |    2 | No            |  2019-11-26  |
| CERES-SYN1deg                  |    3 | No            |  2019-02-07  |
| CowtanWay                      |    2 | Yes           |  2020-02-26  |
| CRU                            |    2 | Yes           |  2019-05-16  |
| CT2019                         |    2 | Yes           |  2020-03-23  |
| Duveiller2018                  |    2 | Yes           |  2019-04-30  |
| E-OBS                          |    2 | Yes           |  2020-02-25  |
| Eppley-VGPM-MODIS              |    2 | Yes           |  2019-05-15  |
| ERA-Interim-Land               |    3 | No            |  2019-11-04  |
| ERA-Interim                    |    3 | No            |  2019-09-05  |
| ESACCI-AEROSOL                 |    2 | Yes           |  2019-01-24  |
| ESACCI-CLOUD                   |    2 | Yes           |  2019-02-01  |
| ESACCI-FIRE                    |    2 | Yes           |  2019-01-24  |
| ESACCI-LANDCOVER               |    2 | No            |  2019-01-10  |
| ESACCI-OC                      |    2 | Yes           |  2019-02-27  |
| ESACCI-OZONE                   |    2 | Yes           |  2019-02-01  |
| ESACCI-SOILMOISTURE            |    2 | No            |  2019-02-01  |
| ESACCI-SST                     |    2 | No            |  2019-02-01  |
| ESRL                           |    2 | No            |  2020-06-30  |
| FLUXCOM                        |    3 | No            |  2019-07-27  |
| GCP                            |    2 | Yes           |  2019-10-17  |
| GHCN-CAMS                      |    2 | Yes           |  2020-03-04  |
| GHCN                           |    2 | Yes           |  2019-03-08  |
| GISTEMP                        |    2 | Yes           |  2020-03-03  |
| GPCC                           |    2 | Yes           |  2020-02-25  |
| HadCRUT3                       |    2 | Yes           |  2019-02-21  |
| HadCRUT4                       |    2 | Yes           |  2019-02-08  |
| HadISST                        |    2 | Yes           |  2019-02-08  |
| HALOE                          |    2 | Yes           |  2020-03-11 |
| HWSD                           |    3 | No            |  2019-10-15 |
| ISCCP-FH                       |    2 | Yes           |  2019-11-07 |
| JMA-TRANSCOM                   |    3 | No            |  2019-07-02 |
| LAI3g                          |    3 | No            |  2019-05-03 |
| LandFlux-EVAL                  |    3 | Yes           |  2019-05-16 |
| Landschuetzer2016              |    2 | Yes           |  2019-03-08 |
| MAC-LWP                        |    3 | No            |  2020-01-30 |
| MERRA2                         |    3 | Yes           |  2019-11-29 |
| MLS-AURA                       |    3 | No            |  2020-02-03 |
| MODIS                          |    3 | No            |  2019-02-09 |
| MTE                            |    3 | No            |  2019-05-07 |
| NCEP                           |    2 | Yes           |  2019-02-04 |
| NDP                            |    3 | No            |  2019-10-14 |
| NIWA-BS                        |    3 | No            |  2019-02-07 |
| NSIDC-0116-nh                  |    3 | Yes           |  2019-05-13 |
| OSI-450-nh                     |    2 | No            |  2019-05-02 |
| OSI-450-sh                     |    2 | No            |  2019-05-02 |
| PATMOS-x                       |    2 | Yes           |  2019-02-10 |
| PERSIANN-CDR                   |    2 | Yes           |  2020-04-22 |
| PHC                            |    2 | Yes           |  2019-01-31 |
| PIOMAS                         |    2 | No            |  2019-05-10 |
| REGEN                          |    2 | Yes           |  2020-02-26 |
| UWisc                          |    3 | No            |  2015-04-15 |
| WOA                            |    2 | Yes           |  2019-01-31 |
-----------------------------------------------------------------------
```