

# Information required from each model group for cmorizing and uploading TRACMIP to ESGF

Aiko Voigt, 18 September 2017

## Background

From the very beginning TRACMIP was designed as a public dataset. However, the use of TRACMIP has been lower than we had hoped. In response to this, Michela and I would like to upload the TRACMIP dataset to ESGF, as we believe this will increase its usage within the TRACMIP community as well as the wider community. ESGF allows for a more convenient access of the data in close analogy of the CMIP5 archive, which many in the climate community are used to.

As a first step in this direction I have recently transferred all TRACMIP data from the University of Miami server to the DKRZ supercomputer in Hamburg, Germany, and I have established contacts at DKRZ regarding the data upload.

However, the data needs to be cmorized before it can be uploaded to ESGF. Some of you have done this already (thanks!), but not all, and even when data was cmorized some of the required metadata might be missing or the latitudes might not be in the correct ordering. This is for example the case for the ECHAM6.1 data that I generated.

To produce data compliant with the cmor and ESGF requirements, I have developed a script that uses the cdo cmor operator on the DKRZ supercomputer (the implementation of cmor2 is used, which was the cmor version applied for CMIP5). The cdo cmor operator is a handy wrapper to the CMOR (Climate Model Output Rewriter) program developed by LLNL. The script allows for batch processing of files based on information provided in so-called infofiles, mapping tables and MIP tables.

Changes to simulation names required for ESGF

To be more easily find at the ESGF database, we have decided to rename the simulations so that they start with a lower-case letter. In this way people looking for aquaplanet simulations will not only find the CMIP5 aquaplanets but also the TRACMIP aquaplanets. Moreover, to distinguish TRACMIP from CMIP5 and CMIP6 simulations with the same or a very similar name, the word “TRACMIP” is added at the end. The ESGF simulation names are (<ESGFSIMULATIONNAMES>):

- aquaControlTRACMIP
- aqua4xCO2TRACMIP
- landControlTRACMIP
- land4xCO2TRACMIP
- landOrbitTRACMIP.

## Request and action needed

cdo cmor requires three kinds of input files:

1. A outputstream file specific to the output stream Amon, Aday and A3hr: there is one file for each output stream, these files are called TRACMIP\_<OUTPUTSTREAM>, where <OUTPUTSTREAM> is either Amon, Aday or A3hr.
2. A cdocmorinfo file specific to the model and the simulation: there is one file for each model and simulation, it is named cdocmorinfo\_TRACMIP\_<ESGFMODELNAME>\_<ESGFSIMULATIONNAME>. The values for <ESGFMODELNAME> are given in the first column of the table below, please note that this differ slightly from the names used in the Tracmip 2016 JAMES introduction paper by Voigt et al. The values for <ESGFSIMULATIONNAME> were given in the section above.
3. A mapping table specific to each model: there is one such file per model. The file is called mt\_TRACMIP\_<ESGFMODELNAME>.txt.

The outputstream files TRACMIP\_<OUTPUTSTREAM> have been generated by me based on the CMIP5\_Amon table. You do not need to change anything regarding these files. Yet, they define the properties of the cmorized variables, such as long name, units, time-averaging (snapshot or mean over some period). For us, the most important thing to know is that the files define if a variable has a “direction”. This is defined by the keyword “positive”, which can be “up” or “down”. See, for example, the definition of the variable rsut in the file. For variables with a “positive” keyword, the mapping table mt\_TRACMIP\_<ESGFMODELNAME>.txt needs to include the attribute “p”. I will explain this in more detail below, here you need to remember that some variables in the mapping table are defined to have a direction in the cmorized data version. If so, the mapping table needs to say what direction these variables have in the model’s native data version.

The cdocmorinfo files and mapping tables are specific to each model and/or simulation. Assuming that these files and tables are available, scripting the cdo cmor processing is doable for me in a reasonable amount of time. Creating these model-specific files for each of the 14 models is beyond what I can do, however, and after all you know this information better than me. Therefore, I am asking each of the model groups to create these files and submit them to me. Specific descriptions are given in the next section.

As an overview and to keep track the following table lists all models and the responsible persons. For models with more the one responsible person feel free to distribute the work as you see fit, and for the CAM models it might be good to join forces. But please make sure that the files are consistent among each other (e.g., how the model reference publication is given in the infofiles, upper- and lower-case letters etc.).

The table also contains the model names as they are planned for publication in ESGF. Model names are adapted from the model names used in the TRACMIP introduction paper of Voigt et al., 2016, JAMES, but with blanks and dots removed (minus signs are kept).

| <b>ESGFMODELNAME</b> | Model name (as in Voigt et al., 2016, JAMES) | Model name as on UMiami server | Responsible person(s)                          |
|----------------------|----------------------------------------------|--------------------------------|------------------------------------------------|
| <b>AM21</b>          | AM2.1                                        | AM2                            | Sarah Kang<br>Elizabeth Maroon<br>Jeongbin Seo |
| <b>CAM3</b>          | CAM3                                         | CAM3                           | Ross Dixon                                     |
| <b>CAM4</b>          | CAM4                                         | CAM4                           | Brian Rose                                     |

|                     |              |                 |                                                  |
|---------------------|--------------|-----------------|--------------------------------------------------|
| <b>CAM5Nor</b>      | CAM5Nor      | NorESM2         | Thomas Toniazzo                                  |
| <b>CNRM-AM5</b>     | CRNM-AM5     | CNRM-AM6-DIA-v2 | Romain Roehrig                                   |
| <b>ECHAM61</b>      | ECHAM6.1     | ECHAM-6.1       | Aiko Voigt                                       |
| <b>ECHAM63</b>      | ECHAM6.3     | ECHAM-6.3       | Jürgen Bader<br>Jong-yeon Park                   |
| <b>GISS-ModelE2</b> | GISS ModelE2 | GISS_ModelE     | Sonali McDermid<br>Gary Russell<br>Jeffrey Jonas |
| <b>LMDZ5A</b>       | LMDZ5A       | IPSL-CM5A       | Francis Codron                                   |
| <b>MetUM-CTL</b>    | MetUM-CTL    | MetUM-GA6-CTL   | Nicholas Klingaman                               |
| <b>MetUM-ENT</b>    | MetUM-ENT    | MetUM-GA6-ENT   | Nicholas Klingaman                               |
| <b>MIROC5</b>       | MIROC5       | MIROC5          | Masakazu Yoshimori                               |
| <b>MPAS</b>         | MPAS         | MPAS            | Jian Lu<br>Ruby Leung                            |
| <b>CALTECH</b>      | CALTECH      | CaltechGray     | Simona Bordoni<br>Ho-Hsuan Wei                   |

## Detailed description of required info files and tables

The following is needed:

- 1 cdocmorinfo file for each model and simulation (e.g., 5 info files for ECHAM61 corresponding to the 5 TRACMIP simulations)
- 1 mapping table for each model (recall that the mapping table does not depend on the simulation)

### Info file:

The cdocmorinfo files are named

cdocmorinfo\_TRACMIP\_<ESGFMODELNAME\_ESGF>\_<ESGFSIMULATIONNAME>.

The infofiles define many of the metadata that need to be included in the cmorized netcdf files. I am attaching the cdomorinfo files for ECHAM61 for aquaControlTRACMIP and aqua4xCO2TRACMIP (note that lines starting with # are ignored). Here, I also explain the meaning of the keywords by the comments in the line above. You should use the ECHAM61 infofiles as templates and adapt the lines that are marked by ADAPT. Please do not change the lines marked by DO NOT CHANGE.

```
# simulation name; DO NOT CHANGE
EXPERIMENT_ID="aquaControlTRACMIP"
```

```
# number of realization, initialization, and physics package; DO NOT CHANGE
MEMBER=r1i1p1
```

```
# the name of the project; DO NOT CHANGE
PROJECT_ID=TRACMIP
```

# DO NOT CHANGE

PRODUCT=output

# a general comment of the project that the data belongs to; DO NOT CHANGE

COMMENT="aqua planet control of TRACMIP; for TRACMIP see Voigt et al., 2016, The Tropical Rain belts with an Annual cycle and a Continent Model Intercomparison Project: TRACMIP, 9, 1868–1891, doi:10.1002/2016MS000748"

# DO NOT CHANGE

HISTORY="N/A"

# the following three lines specify from which simulation this simulation was restarted

# from aquaControlTRACMIP, all entries are N/A; but for aqua4xCO2TRACMIP, the entries would  
# be

# PARENT\_EXPERIMENT="aquaControlTRACMIP"

# PARENT\_EXPERIMENT\_RIP="r1i1p1"

# PARENT\_EXPERIMENT\_ID="aquaControlTRACMIP"

# ADAPT ACCORDING TO YOUR SIMULATION

PARENT\_EXPERIMENT="N/A"

PARENT\_EXPERIMENT\_RIP="N/A"

PARENT\_EXPERIMENT\_ID="N/A"

# this is either CTRL, CO2, or Orbit depending on the simulation; ADAPT TO SIMULATION

FORCING="CTRL"

# the time at which the simulation is restarted from the parent experiment, for ECHAM61 and

# aqua4xCO2TRACMIP this is for example 0046-01-01; ADAPT TO SIMULATION

BRANCH\_TIME="N/A"

# defines the time when your aquaControlTRACMIP simulation was started; ADAPT TO YOUR

# AQUACONTROLTRACMIP SIMULATION

REQUIRED\_TIME\_UNITS="days since 0001-1-1 00:00:00"

# model info:

# put ESGF model name here; ADAPT TO ESGFMODELNAME LISTED IN TABLE ABOVE

MODEL\_ID=ECHAM61

# include reference for the model; this should be the same reference(s) as in column 2 of Table 3

# of the TRACMIP introduction paper; ADAPT TO YOUR MODEL

REFERENCES="ECHAM6.1: Stevens et al., 2013, Atmospheric component of the MPI-M Earth System Model: ECHAM6, JAMES, 5, 146-172, doi:10.1002/jame.2015;"

# some information about the model, this should be the information given in columns 3 and 4

# of Table 3 of the TRACMIP introduction paper; ADAPT TO YOUR MODEL

SOURCE="ECHAM6.1, T63 (1.9 deg lat x 1.9 deg lon), 47 levels; atmosphere component of MPI-ESM"

# specify the calendar you used, available options are: 'gregorian', 'proleptic gregorian', '360 day',

# 'noleap' and 'all leap'; ADAPT TO YOUR MODEL

CALENDAR=360\_day

# institution and contact info:

```
# short and long name of the institution, define a contact for the data set (name and email address)
# ADAPT TO YOUR MODEL
INSTITUTE_ID="LDEO"
INSTITUTION="Lamont-Doherty Earth Observatory, Columbia University; Karlsruhe Institute of
Technology"
CONTACT="aiko@ldeo.columbia.edu; aiko.voigt@kit.edu"
```

## **Mapping table**

There is one mapping table `mt_TRACMIP_<MODELNAME_ESGF>.txt` per model. The table contains the following information for each variable and output stream (Amon, Aday and A3hr):

- the variable name in the model mapped to the cmor variable name
- the direction of the variable assumed in the model; this is an optional key and allows cdo cmor to reverse the sign if needed; the key must be given if a variable in the corresponding TRACMIP\_<OUTPUSTREAM> file is registered as “directed” (see above; recall that this is the case for example for radiative fluxes or the zonal and meridional surface wind stress)
- the variable units as used in the model
- the time averaging method (mean or snapshot)

Thus, there is one line for each combination of variable and output stream.

The ECHAM61 mapping table is attached. Lines starting with ! are ignored. As an illustration, I explain the meaning of some of the files:

```
&parameter name=cl      cmor_name=cl      units="%"      cell_methods="m"
mip_table=Amon /
```

- this maps the model variable `cl` to the cmor variable `cl`, the units in the model are `%`, the variable is a time-averaged quantity and belongs to the Amon stream and so the TRACMIP\_Amon file will be used

```
&parameter name=cl      cmor_name=cl      units="%"      cell_methods="m"
mip_table=Aday /
```

- this is the same as above except for the fact that the handling of the variable for the Aday stream is defined here and so the TRACMIP\_Aday file will be used

```
&parameter name=cl      cmor_name=cl      units="%"      cell_methods="m"
mip_table=A3hr /
```

- the same again, but now for the A3hr stream

```
&parameter name=hfls    cmor_name=hfls    p="u" units="W m-2"    cell_methods="m"
mip_table=Amon /
```

- `hfls`=surface latent heat flux
- the direction keyword “`p`” is needed for some of the variables, `p=“u”` means that `hfls` is defined positive upward in the uncmorized ECHAM61 output
- note that `p` is the direction used in the model (not in CMIP5)
- the keyword `p=d/u` allows cdo cmor to reverse the sign if needed (e.g., if a model has a variable defined as positive downward, as indicated by `p=“d”` here, but the cmorized variable is defined as positive upward, as given in the TRACMIP\_<OUTPUTSTREAM> file)

&parameter name=rlds cmor\_name=rlds p="d" units="W m-2" cell\_methods="m"  
mip\_table=Amon /

- rlds= longwave radiation downward at the surface,
- now p="d" means that the rlds is defined positive downward in ECHAM61

&parameter name=ua cmor\_name=ua units="m/s" cell\_methods="p"  
mip\_table=A3hr /

- ua = 3-d zonal wind
- cell\_methods="p" means that this is a snapshot for the 3hr stream

&parameter name=ua cmor\_name=ua units="m/s" cell\_methods="m"  
mip\_table=Aday /

- this is again for ua but now for the day stream, note that cell\_methods="m" here, which is correct as the day stream in TRACMIP should be daily-averages for all variables

You should only need to adapt name, units and p. There should be no need to adapt the keyword cell\_methods.

## Vertical grids

If your data has not been interpolated to the 17 CMIP5 pressure levels, I believe can quite easily do this with cdo if you send me information about the vertical grid and how it relates to atmospheric pressure.

## List of variables

Please send me a list of the variables that are available for your model. I need this for proper scripting. For this list please use the model variable names (not the CMIP5 cmor names). I am asking for this list as not all the CMIP5 Amon variables might be available for all models.

## Information regarding the simulation times

For each simulation, please send me the start date and the end date. E.g., for ECHAM61 this would be:

AquaControl: 0001-01-01 --- 00451-12-30  
Aqua4xCO2: 0046-01-01 --- 0090-12-30  
LandControl: 0046-01-01 --- 0090-12-30  
Land4xCO2: 0091-01-01 --- 0135-12-30  
LandOrbit: 0091-01-01 --- 0135-12-30

Please also indicate for each simulation, which output stream is available for which time period. That is:

AquaControl Amon 0016-01-01 – 0045-12-30  
AquaControl Aday 0036-01-01 – 0045-12-30  
AquaControl A3hr 0043-01-01 – 0045-12-30

And the same for all other simulations. Again, I need this info for scripting the cmorizing.

## **Wrap-up of requested input from you**

Please send me the following files and information for your model:

- 1 cdocmorinfofile for each simulation
- 1 mapping table mt\_TRACMIP\_<ESGFMODELNAME>.txt for your model
- if your data is not on CMIP5 pressure levels, then information concerning the vertical grid
- information about simulation times and times for which the individual output streams are available

If there is anything else that you think I should know about your data, please include that as well.

## **Attachments**

- ECHAM61 infofiles for all 5 simulations
- ECHAM61 mapping table
- TRACMIP\_Amon, TRACMIP\_Aday and TRACMIP\_A3hr

**Some technical infos for those interested** (might or might not be helpful to you, but for sure helpful to Aiko)

- I am using the following cdo version on mistral.dkrz.de:  
/work/bm0021/cdo\_incl\_cmor/cdo\_08\_30\_2017\_cmor2/cdo\_gitcmor3branch/src/cdo
- Some info on the cdo cmor operator:  
<https://code.mpimet.mpg.de/projects/cdo/wiki/CMOR>
- The CMIP5\_Amon table can be found here for example:  
<https://github.com/PCMDI/cmip5-cmor-tables/tree/master/Tables>  
As defined in the TRACMIP introduction paper, we request, if possible, all variables from this table, except the chemical composition variables. It is not a problem if some of the variables are not available (this is the case for ECHAM61 for example).
- Some info on the CMOR operator can be found here:  
<https://github.com/PCMDI/cmor>  
Note that this describes the version cmor3 that will be used for CMIP6, while the cdo cmor operator I am using implements the version cmor2 (also used for CMIP5).