

APPENDIX A: Model output specifications

The general structure of output files for the TFHTAP common model experiments is described in the document *HTAP_CoordinatedStudies_ExpSet1.doc* (available at <http://aqm.jrc.it/HTAP/>). All model output shall be submitted in netcdf format and attributes should adhere to the CF conventions (see <http://www.cgd.ucar.edu/cms/eaton/cf-metadata>) where possible. Details on the naming conventions and units for the variables contained in the output files can be found in the Excel tables *HTAP_variables.xls* (also available at <http://aqm.jrc.it/HTAP/>). If possible, please include also the *standard_name* attribute for each variable as this will go a long way to making your files CF compliant and facilitate the later archiving and exchange with other scientists (CF is also being used for analysis of model data for the forthcoming 4th assessment report of IPCC, for example).

Example TFHTAP output files are provided at the TFHTAP web pages <http://aqm.jrc.it/HTAP/>.

We recommend the use of the **Climate Model Output Rewriter (CMOR)** to construct HTAP- and CF-compliant files and generate the file structure defined in TFHTAP. The use of CMOR guarantees that the data format requirements described below are followed. CMOR also contains a TFHTAP compliance checker which can be used to test output files which are otherwise produced.

The CMOR software package

CMOR has been developed by PCMDI (<http://www-pcmdi.llnl.gov/software-portal/cmor/>). It comprises a set of FORTRAN 90 functions that can be used to produce HTAP/CF-compliant netCDF files that fulfill the requirements of many of the climate community's standard model experiments. These experiments are collectively referred to as Model Intercomparison Experiments (MIPs) and include, for example, AMIP, CMIP, CFMIP, PMIP, APE, and IPCC scenario runs. The output resulting from CMOR is "self-describing" and facilitates analysis of results across models. CMOR has been adapted to the HTAP experiments by Christiane Textor and Michael Schulz.

The CMOR software and documentation can be downloaded from the PCMDI web site <http://www-pcmdi.llnl.gov/software-portal/cmor/>. Please follow the instructions provided in the CMOR documentation. A short guide to installing and using CMOR can be found below.

HTAP-specific MIP tables to be used by CMOR are provided at the TFHTAP web pages <http://aqm.jrc.it/HTAP/>. CMOR uses these tables to construct HTAP-compliant files, thereby reducing the programming effort required of the individual MIP contributors. All formatting requirements and recommendations listed below are taken care of by CMOR, once it has been properly installed and adapted to your model's simulation data.

CMOR produces files contain one variable but can contain several time steps. These files are concatenated with a nco tool provided at the TFHTAP web pages <http://aqm.jrc.it/HTAP/> (*create_htap_files.job*) to produce the TFHTAP file structure described in experiment description.

Data format requirements:

- Data must be written through the [netCDF](#) API (application program interface) and should conform to the [CF metadata standards](#). (Please note, that the standard names for aerosols and chemicals used with in TFHTAP are still under discussion for official approval in CF, but minor changes are expected.)
- The file structure must follow the **General description of model output files and Standard diagnostics and file identifiers**. Each file will include coordinate variables, attributes and other metadata as specified below.
- If the field is a function of time, the following rules should be followed:
 - monthly mean/integrated data: 12 records (= 1 year) per file
 - daily mean data: 1 record per (daily) file
 - hourly mean data: 24 records per (daily) file
 - instantaneous data (00 and 12UTC): 2 records per (daily) file
- For data that are a function of longitude and latitude, only grids representable as a Cartesian product of longitude and latitude axes are allowed. Model output on other grids, such as "thin" grids, grids with rotated poles, and irregular grids, must be mapped to a longitude-latitude Cartesian grid before being stored.
- Atmospheric fields that are functions of the vertical coordinate reside on the original model levels. There are different options to describe vertical coordinates with associated metadata and attribute requirements specified below for different types of vertical axes.
- The File size should never exceed 2 gigabytes.

Requirements for output variables:

- All output fields must be written as single precision floating point numbers (netCDF type "float").
- The variable names (in the netCDF file) must be the "output variable names" specified in the HTAP Excel or Standard MIP tables (e.g., ozone mole mixing ratio will be stored in a variable named "vmr_o3").
- The units required for the output fields are given in the HTAP Standard Output tables.
- The positive direction of vertical fluxes must be consistent with that specified in the HTAP Standard MIP table. For example, the positive direction for emissions and deposition fluxes have to be provided in the netCDF file.
- The order of array dimensions must be: time, level, latitude, longitude. For fields of lesser rank, one or more coordinate will be omitted, but the ordering must be preserved. The above order assumes the C language convention in which the last coordinate varies most rapidly (note that this is the opposite from the FORTRAN and IDL ordering). In the few cases where an additional dimension is required (e.g., the

"region" dimension used to distinguish different ocean basins), that dimension should appear just after the "time" dimension.

- If longitude is a coordinate, data must be stored west to east (in degrees), starting with the first grid point greater than or equal to 0 degrees east. All coordinate locations must be unique (e.g., don't include both 0 and 360 degrees east).
- If latitude is a coordinate, data must be stored south to north (in degrees).
- If there is a vertical coordinate, data must be stored starting with the grid point nearest the surface.
- If there is a time dimension, data must be stored with time increasing. Reference date is 2001-01-01 00:00.
- All "missing data" should be assigned the single precision floating point value of 1.e20. However, if a model does not generate a certain variable, please do not submit a field of missing values for this variable but rather omit it from the file.
- Variable attributes:
 - standard_name = a character string containing a HTAP/CF standard name that can be found in the HTAP Standard Output tables or in the HTAP Standard MIP tables.
 - units = a character string containing the units given in the HTAP Standard Output tables. The form should be consistent with [UNIDATA's Udunits package](#) (e.g., "kg m-2" instead of "kg/m^2").
 - If the variable has any "missing values", then define _FillValue = 1.e20, and assign that value to data that are "missing".
 - If the variable contains a time-mean, then the "cell_methods" attribute should be defined with the value "time: mean". If the variable is accumulated over a time period then the cell_methods attribute must include the string "time: sum", like for the monthly integrated precipitation.
- Recommended additional variable attributes:
 - long_name = a character string that might serve as a title for the variable on plots (e.g., "Surface Air Temperature").
 - history = how the variable was processed before outputting (e.g., give name(s) of the file(s) from which the data were read and indicate what calculations were performed, such as interpolating or changing the unit). This information should allow someone at the user's institute to reproduce the procedure that created the output. Note that this history attribute is variable-specific, whereas the global history attribute defined below (see "Recommended attributes" under "Requirements for global attributes") provides information concerning

the model simulation itself or refers to processing procedures common to all variables.

→ comment = a character string providing further information concerning the variable (e.g., for precipitation the comment might read "includes all types (rain, snow, large-scale, convective, etc.)").

Requirements for coordinate variables:

- All coordinate variables should be written as double precision floating point numbers (netCDF type "double"). For the time coordinate this is mandatory, while we will also accept files where lon, lat, and lev are single precision.
- A longitude coordinate must be named "lon", its CF standard name is "longitude", and its units must be "degrees_east". The values should be stored west to east starting with the first longitude greater than or equal to 0 degrees, but we will also accept files starting with -180 degrees longitude. If you wish to remap your files yourself, you can find the ncregrid.pro package for IDL on <ftp://ftp.retro.enes.org/pub/htap/tools/>.
- A latitude coordinate must be named "lat", its CF standard name is "latitude", and its units must be "degrees_north". The values should be stored south to north, but we will also accept files in opposite order. However, please note that we cannot accept data on "sparse" or "reduced" grids. If you wish to remap your files yourself, you can find the ncregrid.pro package for IDL on <ftp://ftp.retro.enes.org/pub/htap/tools/>.
- A time coordinate must be named "time", its CF standard name is "time", and its units must be "days since [basetime]", where [basetime] is "days since 2001-01-01 00:00". For time-mean data, it is recommended that the coordinate value be defined as the mid-point of the interval over which the average is computed (for example, the 15th of a month for monthly mean data).
- The dimensionless vertical coordinate (indicating model level) must be named "lev". Different types can be distinguished by their CF standard names that can be found in the HTAP Output Table and [Appendix D^{\[20\]}](#) of the CF-conventions. The values should be stored starting with the coordinate nearest the earth's surface. We will also accept files with vertical coordinates stored in opposite order (from the top of the atmosphere down). Please note that vertical coordinates are a frequent source of confusion and can lead to severe processing errors. Therefore, if in doubt, please send us a sample file with your model output including the vertical coordinate specification and we will tell you if this is acceptable.
- For the dimensionless vertical coordinates, the user must include in the file all the information needed to positively and uniquely indicate the location of the data. In this case the "formula_terms" attribute should be defined and additional variables or parameters may need to be stored in the file (e.g., surface pressure and pressure at the top of the model for an atmospheric-sigma coordinate system). See [section 4.3.2 of the CF-conventions^{\[22\]}](#) and [Appendix D^{\[23\]}](#) of the CF-conventions.

- It would be helpful if bounds on coordinate variables are also included in the file. For time, the bounds should be included if the data represent a time-mean (e.g., monthly mean) or time-integral (e.g., sum over a month). For model-level data, the cell bounds should be included, indicating the interfaces between model levels. For data that represent a “snap-shot” in time (e.g., instantaneous values), the time coordinate should have no “bounds” attribute. Note that CF requires cell bounds for 1-dimensional coordinates be stored in 2-dimensional arrays as described in [section 7.1 of the CF-conventions^{\[21\]}](#) and as shown in the examples below.
- Coordinate variable attributes:
 - `standard_name` = a character string containing the standard name (see coordinate standard names given above).
 - `units` = a character string containing the units (see allowable coordinate units given above). Note that in the case of a dimensionless vertical coordinate, this attribute may be omitted.
 - `axis` = "X", "Y", "Z", or "T" as appropriate (see [section 4 of the CF-conventions\[26\]](#)). In the few cases where an additional dimension is required (e.g., the "region" dimension used to distinguish different ocean basins), this attribute should be omitted (for this additional dimension).
 - `bounds` = a character string containing the name of the variable where the cell bounds are stored for the vertical axis, or alternatively a variable defined at the interfaces between model levels.
 - For dimensionless vertical coordinates the “positive” attribute must be included and should be set to either “down” or “up”, as appropriate.
 - For dimensionless vertical coordinates, the “`formula_terms`” attribute must be defined as described in [section 4.3.2 of the CF-conventions\[28\]](#) and as illustrated in example below. Further information can be found in [Appendix D\[29\]](#) of the CF-conventions.
 - For time coordinates, the “calendar” attribute must be included as described in [section 4.4.1 of the CF-conventions\[27\]](#).
-

Global attributes:

- Required attributes:
 - `institution` = a character string containing an abbreviated name or acronym of the institution that performed the model simulation, followed in parentheses by the fully spelled-out name and location (e.g., 'GFDL (Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA)')

→ source = a character string identifying the model version. Additional explanatory information may follow the required information.

→ project_id = "HTAP experiments YEAR "

→ experiment_id = the appropriate character string, identifying the experiment, taken from Table 1 of the PDF-document on ExpSet1, see also the list below.

▪ Recommended attributes:

→ Conventions = 'CF-1.0 - HTAP' .

→ contact = name and contact information (e.g., email, address, phone number) of person who should be contacted for more information about the data.

→ history = A character string containing an audit trail for modifications to the original data. Each modification is typically preceded by a "timestamp". The "history" attribute provided here will be a global one that should not depend on which variable is contained in the file. A variable-specific "history" can also be included as an attribute attached to the output variable.

→ comment = A character string containing additional information about the data or methods used to produce it. The user may include, for example, a description of how the initial conditions for a simulation were specified and how the model was spun-up (including the length of the spin-up period).

→ references = A character string containing a list of published or web-based references that describe the data or the methods used to produce it. Typically, the user should provide references describing the model formulation here.

→ title = "[*institute_name*] model output prepared for TFHAP [*experiment_id*]" where [*institute_name*] should be replaced by the contributing institution's acronym or name (e.g., "GFDL", "CCCma", "Hadley Centre", etc.) and [*experiment_id*] should be replaced by one of the following strings:

SR1	Base case simulation for year 2001.
SR2 (=TP2a)	CH4 global mixing ratio reduced to 1408 ppb (i.e. -20%)
SR3EA	Reduction of anthropogenic NOx emissions by 20% over East Asia
SR3EU	Reduction of anthropogenic NOx emissions by 20% over Europe
SR3NA	Reduction of anthropogenic NOx emissions by 20% over North America
SR3SA	Reduction of anthropogenic NOx emissions by 20% over South Asia
SR4EA	Reduction of anthropogenic NMVOC emissions by 20% over East Asia
SR4EU	Reduction of anthropogenic NMVOC emissions by 20% over Europe
SR4NA	Reduction of anthropogenic NMVOC emissions by 20% over North America
SR4SA	Reduction of anthropogenic NMVOC emissions by 20% over South Asia
SR5EA	Reduction of anthropogenic CO emissions by 20% over East Asia
SR5EU	Reduction of anthropogenic CO emissions by 20% over Europe
SR5NA	Reduction of anthropogenic CO emissions by 20% over North America
SR5SA	Reduction of anthropogenic CO emissions by 20% over South Asia
SR6EA	Reduction of anthropogenic emissions by 20% over East Asia
SR6EU	Reduction of anthropogenic emissions by 20% over Europe
SR6NA	Reduction of anthropogenic emissions by 20% over North America
SR6SA	Reduction of anthropogenic emissions by 20% over South Asia

SR7EA	Reduction of Hg emissions by 20% over East Asia
SR7EU	Reduction of Hg emissions by 20% over Europe
SR7NA	Reduction of Hg emissions by 20% over North America
SR7SA	Reduction of Hg emissions by 20% over South Asia
SR81EA	Reduction of PCB-153 emissions by 20% over East Asia
SR81EU	Reduction of PCB-153 emissions by 20% over Europe
SR81NA	Reduction of PCB-153 emissions by 20% over North America
SR81SA	Reduction of PCB-153 emissions by 20% over South Asia
SR82EA	Reduction of α -HCH emissions by 20% over East Asia
SR82EU	Reduction of α -HCH emissions by 20% over Europe
SR82NA	Reduction of α -HCH emissions by 20% over North America
SR82SA	Reduction of α -HCH emissions by 20% over South Asia
TP1	Passive CO tracer experiment

A sample title is: LSCE model output prepared for TFHTAP Base case simulation for year 2001'

Short guide to using CMOR

- Download the CMOR programs and MIP tables for HTAP: htap_cmor_<version>.tar
- Installation of CMOR

After gunzipping and extracting the files from the tar, you should read the file named "INSTALL" (in the CMOR directory). This tells you how to create the library. As explained in the INSTALL file, you can compile the test codes (called htap_test_codeA1[a-e].f90), which are in the HTAP_CMOR subdirectory. Then run the executables (htap_test_codeA[a-e]). If you need to alter the makefile (or the compile file), other than changing paths, please send us a copy of what you had to do, so others can avoid repeating your work.

- Additional files potentially needed for unit and time conversions

If the input data are provided in units different from what is required by HTAP, those units can be transformed to the correct units using the udunits software, then CMOR performs the transformation before writing the data. Otherwise, CMOR will return an error. To enable this capability, the user will have to obtain and build CMOR with the udunits library and, for time transformations, the cdms library. Please install these library before installing CMOR.

- install udunits

<http://www.unidata.ucar.edu/software/udunits/>

- install CDMS

http://www-pcmdi.llnl.gov/software-portal/esg_data_portal/cdunif-2-1-tar.gz/view

- After having successfully installed CMOR for HTAP
 - adapt all 5 htap_test_codeA1?.f90 programs for your model
 - run them all

- Recombine the CMOR-produced files (one variable per file) to HTAP requested file structure:

- install nco tools <http://nco.sourceforge.net/>
- adapt create_htap.job to your file system
- run create_htap_files.job

- A detailed documentation of CMOR and the related programs can be found at the PCMDI web site:

- http://www-pcmdi.llnl.gov/software/cmor/cmor_users_guide.pdf
- <http://www-pcmdi.llnl.gov/software-portal/cmor/>.

- CMOR can handle the following dimensionless vertical coordinates:

hybrid sigma pressure coordinate	$lev(n,k,j,i) = a(k)*p0 + b(k)*ps(n,j,i)$
sigma coordinate	$lev(n,k,j,i) = ptop + sigma(k)*(ps(n,j,i) - ptop)$
alternate hybrid sigma pressure coordinate	$lev(n,k,j,i) = ap(k) + b(k)*ps(n,j,i)$
hybrid height coordinate	$lev(n,k,j,i) = a(k) + b(k)*orog(j,i)$

The user must declare the coordinate types by the axis identifiers within CMOR (cmor_axis: table_entry). Sigma coordinates and alternate hybrid sigma pressure coordinates will be transferred to hybrid sigma pressure coordinates by CMOR.

- If you have a different vertical coordinate type, please contact christiane.textor@aero.jussieu.fr

Acknowledgement:

We would like to thank Karl Taylor (Program, Climate Model Diagnosis and Intercomparison Lawrence Livermore National Laboratory) for support and for adapting CMOR for the HTAP experiments.

APPENDIX B: Technical Summary of submissions with deadlines and output files

Deadline for submission of model results: 02 Oct 2006

Task no.	Output files
SR1.1	tracerm metm emim depn
SR1.2	sfc<ddd> sfc1<ddd>
SR1.3	aerosolm aerosolaod
SR1.4	depn budgetm
SR1.5	vertprof

Deadline for submission of model results: 15 Nov 2006

Task no.	Output files
SR2.1	tracerm depn budgetm
SR2.2	sfc<ddd>

Deadline for submission of model results: 15 Nov 2006

Task no.	Output files
SR3.1 SR4.1	tracerm depn budgetm
SR3.2 SR4.2	sfc<ddd>

Deadline for submission of model results: 15 Jan 2007

Task no.	Output files
SR6.1	tracerm depn budgetm
SR6.2	sfc<ddd>
SR6.3	aerosolm aerosolaod

Deadline for submission of model results: 15 Jan 2007

Task no.	Output files
SR7.1	Sfc1<ddd>
SR7.2	tracerm dep budgetm

Deadline for submission of model results: 15 Jan 2007

SR8 – POP perturbation runs no detailed information available yet

Task no.	Output files
SR8.1EA	tracerm dep Sfc1<ddd>
SR8.1EU	tracerm dep Sfc1<ddd>
SR8.1NA	tracerm dep Sfc1<ddd>
SR8.1SA	tracerm dep Sfc1<ddd>
SR8.2EA	tracerm dep Sfc1<ddd>
SR8.2EU	tracerm dep Sfc1<ddd>
SR8.2NA	tracerm dep Sfc1<ddd>
SR8.2SA	tracerm dep Sfc1<ddd>

Deadline for submission of model results: 15 Jan 2007

Task no.	Output files
TP1.1	tracerm
TP1.2	tracerm
TP1.3	tracerm