HEMCO: **H**armonized **Em**issions **Co**mponent as an emissions component for MUSICA/CAM-chem

Haipeng **Lin**, November 9, 2022

Rev. 1a (20221109)

https://doi.org/10.5194/gmd-14-5487-2021

Emissions are at the heart of atmospheric chemistry modeling

yet have several limitations in current models

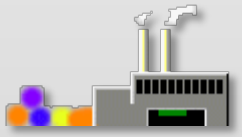
Need to preprocess inventories

to model grid

to target chemical scheme

with ad-hoc tools

using time and disk space



Difficult to share data and code

Source of human error

in pre-processing

for different cases/variations

in reproducing past results

Wide range of sources

possible inconsistent treatment

double counting/undercounting

HEMCO is a powerful on-line emissions component that can (mostly) be controlled by a text-based *configuration file*

```
# ----- REGIONAL INVENTORIES ----------------
    --> APEI                  :       false
    --> NEI2011_HOURLY        :       false
    --> NEI2011_MONMEAN       :       false
    --> MIX                   :       false
    --> DICE_Africa           :       false
# ----- GLOBAL INVENTORIES ------------------
    --> CEDS                  :       true
```

HEMCO Emissions Data Library

Gridded global and regional inventories

HEMCO Extensions

State dependent emission algorithms

(Biogenic, Dust, Lightning, …)

& GFED, Volcano, Ship Plumes…

CAM-Chem/MUSICA

CESM-GC

GEOS-Chem 'Classic'

GCHP

WRF-GC

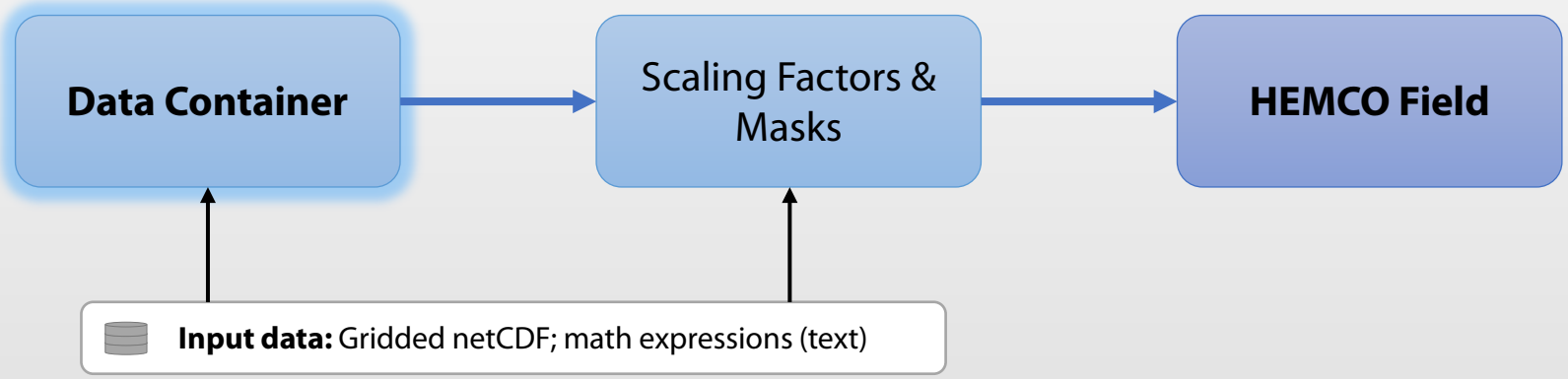
NASA GEOS (GEOS-GC)

Scaling, masking, and adding emissions *from different sources, regions, and species* at runtime *on a user-specified grid* without preprocessing!

Keller et al., 2014

How does HEMCO accomplish this?

"**HEMCO Fields**" are the basic building block




```
0 CEDS_NO_AGR   NO-em-total-anthro_CEDS_$YYYY.nc NO_agr 1970-2017/1-12/1/0 C
  Container name  Source netCDF file                 Variable  Date range (Y/M/D/H)  Cycling Opt

xy kg/m2/s NO                  25/1234          1         5
Dim Unit   Model Species Name  Scaling Factors  Category  Hierarchy
```

Example: Emissions data container in HEMCO.

Tokens like $YYYY, $MM, $DD, $HH are automagically replaced into file names

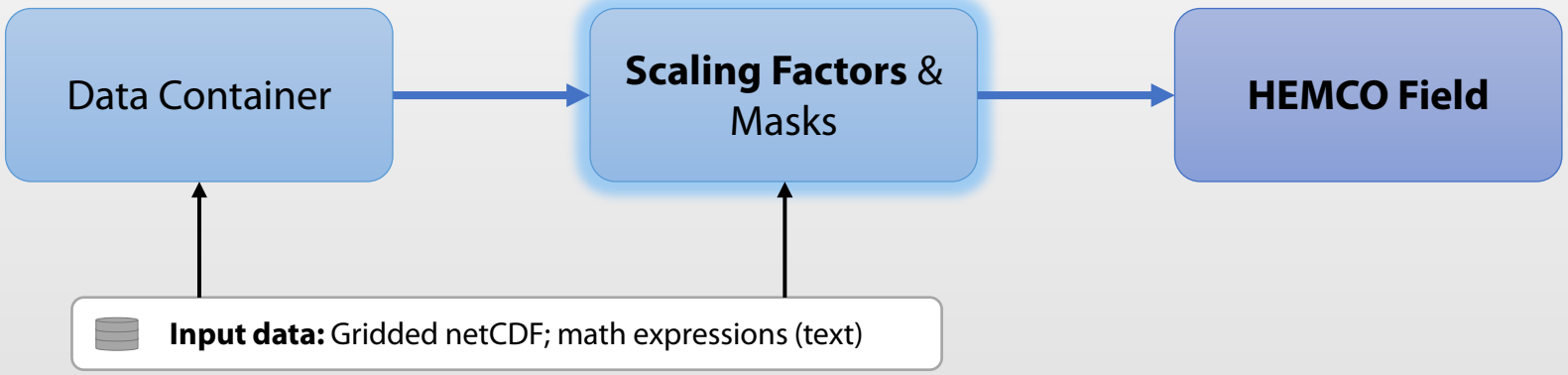
Cycling option: *e.g.,*

- C: cycle closest
- R: only use within range
- RA: range, otherwise avg
- A: average
- I: interpolate nearest two
- E: must be exact match

Haipeng **Lin** · @jimmielin · jimmielin.me · hplin@seas.harvard.edu

How does HEMCO accomplish this?

"**HEMCO Fields**" are the basic building block




```
115       NO2toNO   6.521739e-1    -     -     - xy unitless 1
```

Scale factor #, name *Expression* *Dim Unit* *Operation (1 = *, 2 = /, 3 = * scale^2)*

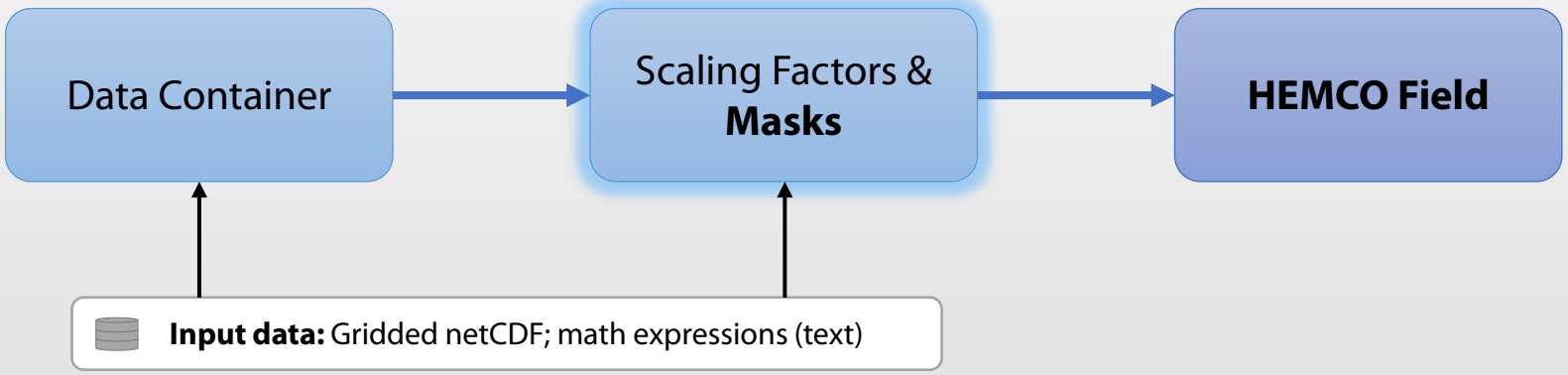
```
22    GEIA_DOW_HC   0.671/1.1102/1.1102/1.1102/1.1102/1.1102/0.768 - - - xy
unitless 1
```

Expression length: 7 = DOW (Sun/Mon/…/Sat); 12 = Jan/Feb/…/Dec; 24 = 00/01/02/…/23 (LT)

Example: Scaling factors in HEMCO.

How does HEMCO accomplish this?

"**HEMCO Fields**" are the basic building block




```
1007 CONUS_MASK   CONUS_Mask.01x01.nc          MASK    2000/1/1/0 C      xy 1 1 -
```

Mask #, name *netCDF File* *Variable* *Time Slice* *Cycling, Dim, Ignore, Ignore*

```
140/20/-50/60
```

Approximate lat/lon bounds of the mask (to speed up I/O)

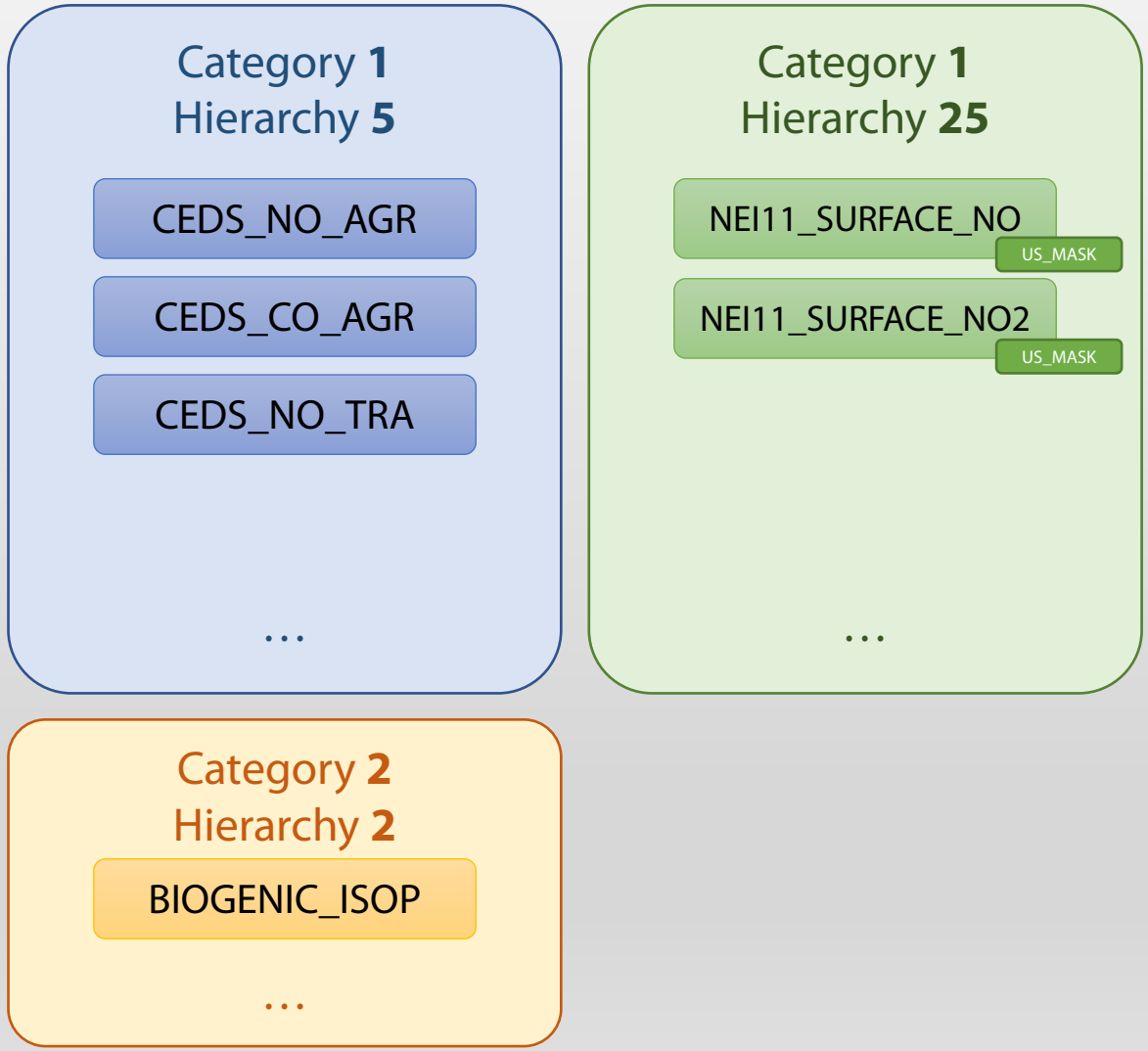
Example: Masks in HEMCO. Shown: CONUS mask (masks are binary)

Masks can be regional, 0 is assumed outside the provided region.

Haipeng **Lin** · @jimmielin · jimmielin.me · hplin@seas.harvard.edu

How does HEMCO accomplish this?

"**Category**" and "**Hierarchy**" are used to combine emission inventories



CEDS used over global domain
NEI11 used over US_MASK range

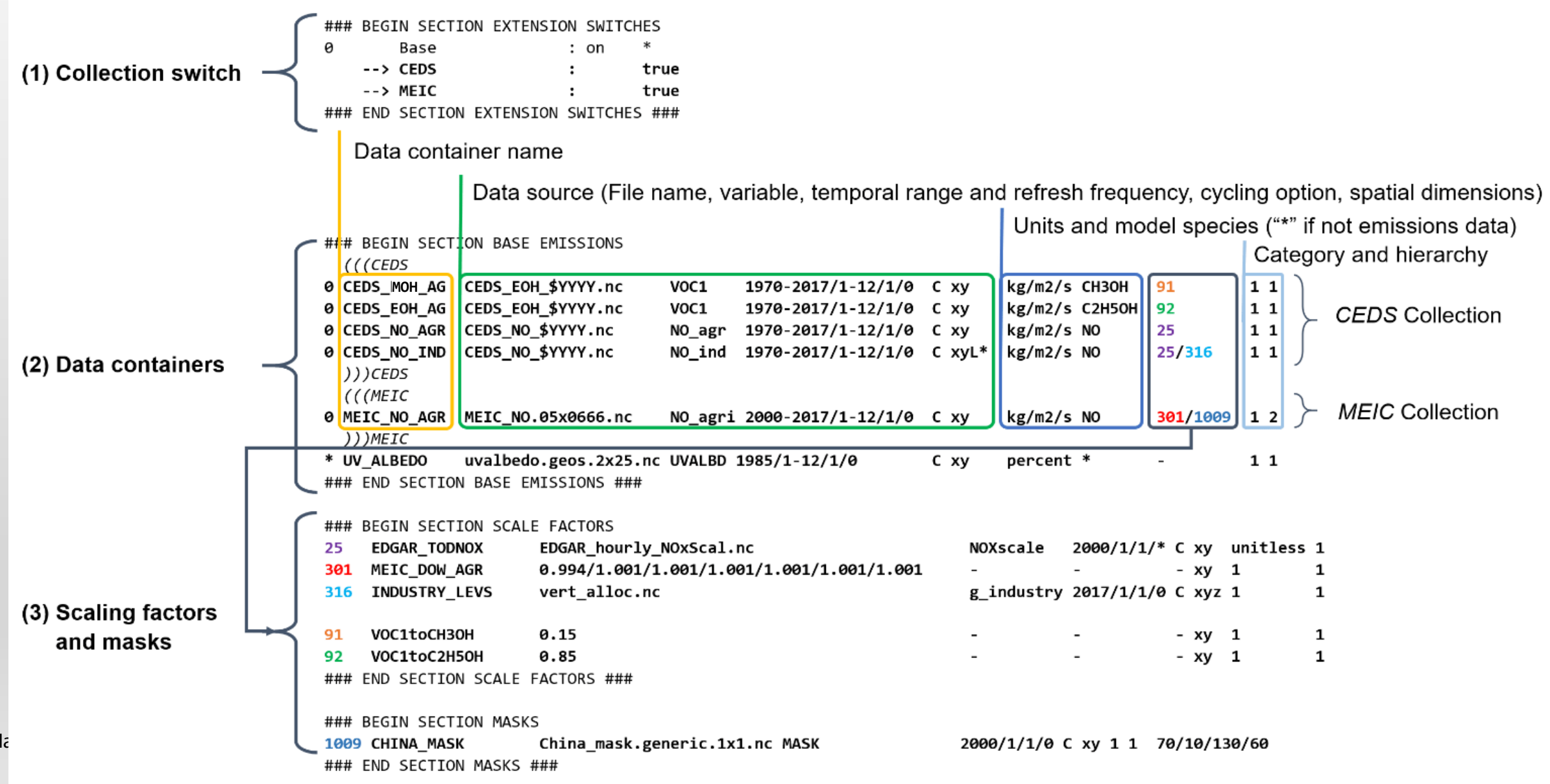
"Category 1 is anthropogenic emissions"

BIOGENIC_ISOP always used over available range

e.g., "Category 2 is biogenic emissions"

Haipeng **Lin** · @jimmielin · jimmielin.me · hplin@seas.harvard.edu

The anatomy of the **HEMCO configuration file**



HEMCO is **not magic**, *but it can describe emissions precisely*

- With HEMCO, you still have to
  - Think about the hierarchy between multiple inventories used
  - Specify the speciation (from source file → model emissions)
  - Process emissions data *once* to latitude-longitude input grid, in COARDS-compliant netCDF*
- But you won't have to
  - Re-process emissions input files *every time you change the model configuration*, e.g.
    - Sensitivity experiments (inventories on/off; scale emissions)
    - Apply scaling factors (diurnal; weekly; monthly; gridded time series; vertical)
    - Running for a different year (Interpolate? Use average? Cycle? Nearest time slice?)
  - Re-grid emissions *every time you change the model grid*

** See guides: https://hemco.readthedocs.io/en/latest/geos-chem-shared-docs/supplemental-guides/coards-guide.html and https://hemco.readthedocs.io/en/latest/geos-chem-shared-docs/supplemental-guides/netcdf-guide.html*

Case examples / exercises, *easy*

- Problem 1: Scaling up CEDS emissions by 2x
- Problem 2: Scaling up CEDS emissions *every Monday* by 2x

Case examples / exercises, *harder*

- Problem 1: Scaling up CEDS emissions by 2x
- Problem 2: Scaling up CEDS emissions *every Monday* by 2x
- Problem 3: Scaling up CEDS emissions *in China* by 2x

Case examples / exercises, *really?*

- Problem 4: Keeping CEDS emissions normal everywhere, *but use 2010 emissions from CEDS in China*

When you want a feature that cannot be described in text config file…

- **…use HEMCO extensions to describe them in code!**
- HEMCO uses "Extensions" to port-in third-party code, e.g., MEGAN
- Extensions can **read data from HEMCO and the model** and **return processed emissions**.

- e.g., "double the emissions every Tuesday after the first Monday of every month"

HEMCO extensions are used to port-in third party emissions code (e.g. MEGAN)

- Extensions can access meteorology data (e.g., wind speed) from the on-line model and read in datasets from HEMCO

Species*	Extension name and reference
Oceanic DMS, acetone, acetaldehyde, methyl nitrate, ethyl nitrate, methanol	SeaFlux (Johnson, 2010)
Ship plume NO_x, HNO_3, O_3	ParaNO_x (Vinken et al., 2011)
Lightning NO_x	LightNO_x (Murray et al., 2012; Ott et al., 2010)
Soil and fertilizer NO_x	SoilNO_x (Hudman et al., 2012)
Mineral dust aerosols	DEAD (Zender et al., 2003)
	Ginoux (Ginoux et al., 2001)
Sea salt aerosols	SeaSalt (Chin et al., 2002; Gong, 2003; Jaeglé et al., 2011)
Biogenic VOCs	MEGAN (Guenther et al., 2012)
Biomass burning	GFED (Akagi et al., 2011; Andreae et al., 2001; Giglio et al., 2013; Randerson et al., 2012; van der Werf et al., 2010)
	FINN (Wiedinmyer et al., 2011)
Volcanic SO_2	Volcano (Carn et al., 2015; Ge et al., 2016)
Inorganic iodine emissions: HOI, I_2	Inorg_Iodine (Carpenter et al., 2013; MacDonald et al., 2014)

Extensions have built-in scale capabilities (e.g., for sensitivity studies)

- For example, you can scale the output of the GFED Biomass Burning extension

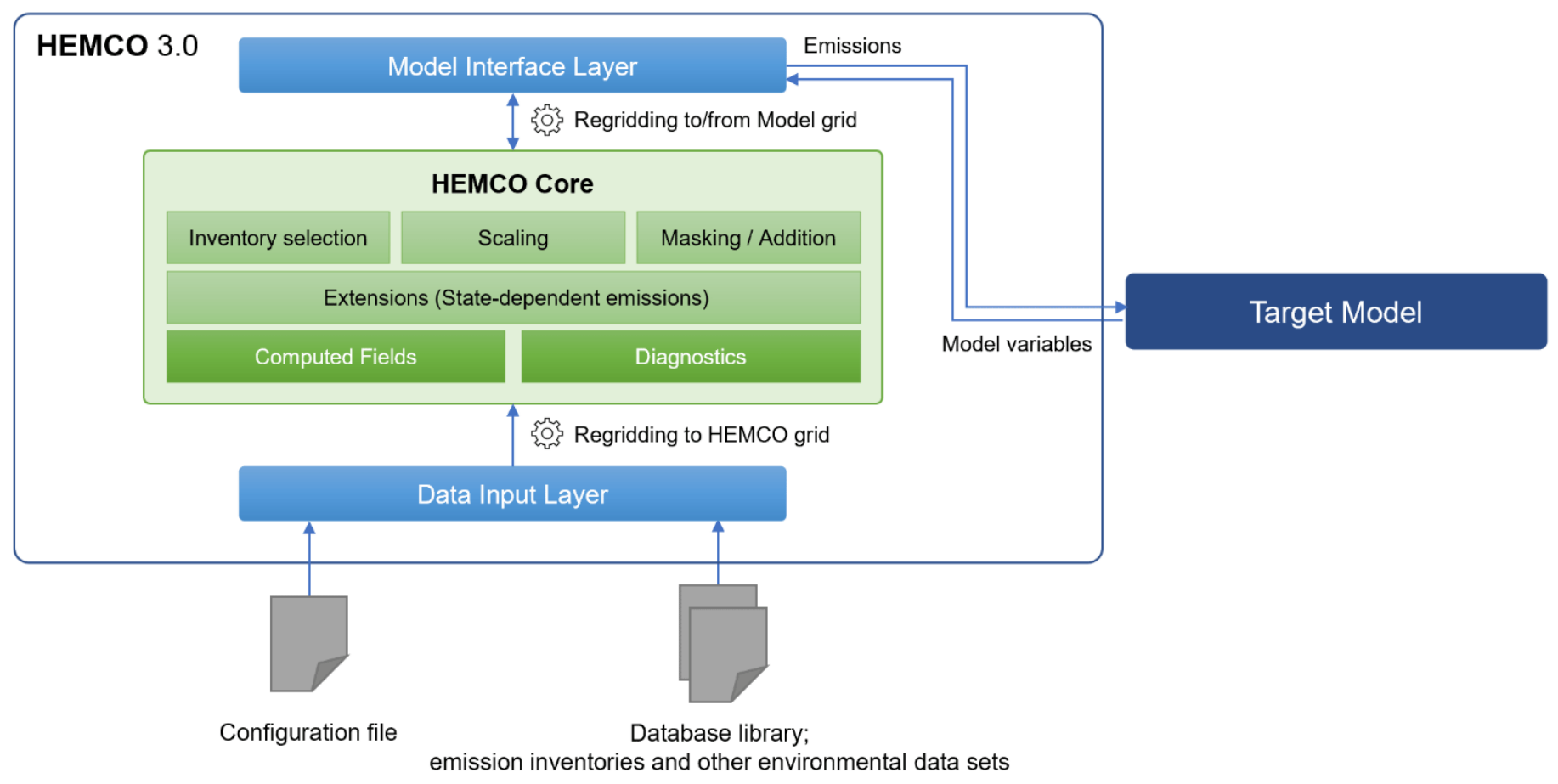
```
111     GFED                        : on
NO/CO/BIGALK/CH3COCH3/MEK/CH3CHO/C3H6/C2H2/C2H4/C3H8/CH2O/C2H6/SO2/NH3/bc_
a4/pom_a4/BENZENE/TOLUENE/XYLENES/C2H5OH/CH3OH
    --> GFED4                   :       true
    --> GFED_daily              :       false
    --> GFED_3hourly            :       false
    --> Scaling_CO              :       1.05
    --> Scaling_bc_a4           :       1.0
    --> Scaling_pom_a4          :       1.4
```

- We scaled the pom_a4 by 1.4* of what GFED extension computes by default

Can I use it in MUSICA today? **Yes**!

- …you will have to checkout my copy of CAM (`cam6_3_034`), though.
- It works with **CESM 2.2** and above with *any grid**
  - * *tested with 1.9x2.5, 0.9x1.25, ne30np4, ne0np4.KORUS03.ne30x16* but others should work
- Can I mix and match with current offline emissions? *Not directly*
  - You can use HEMCO to read the existing offline emissions on regular grids into the model instead
- Set-up instructions are available at https://github.com/ESCOMP/CAM/pull/560
- I'll also make a rough outline of the workflow here.

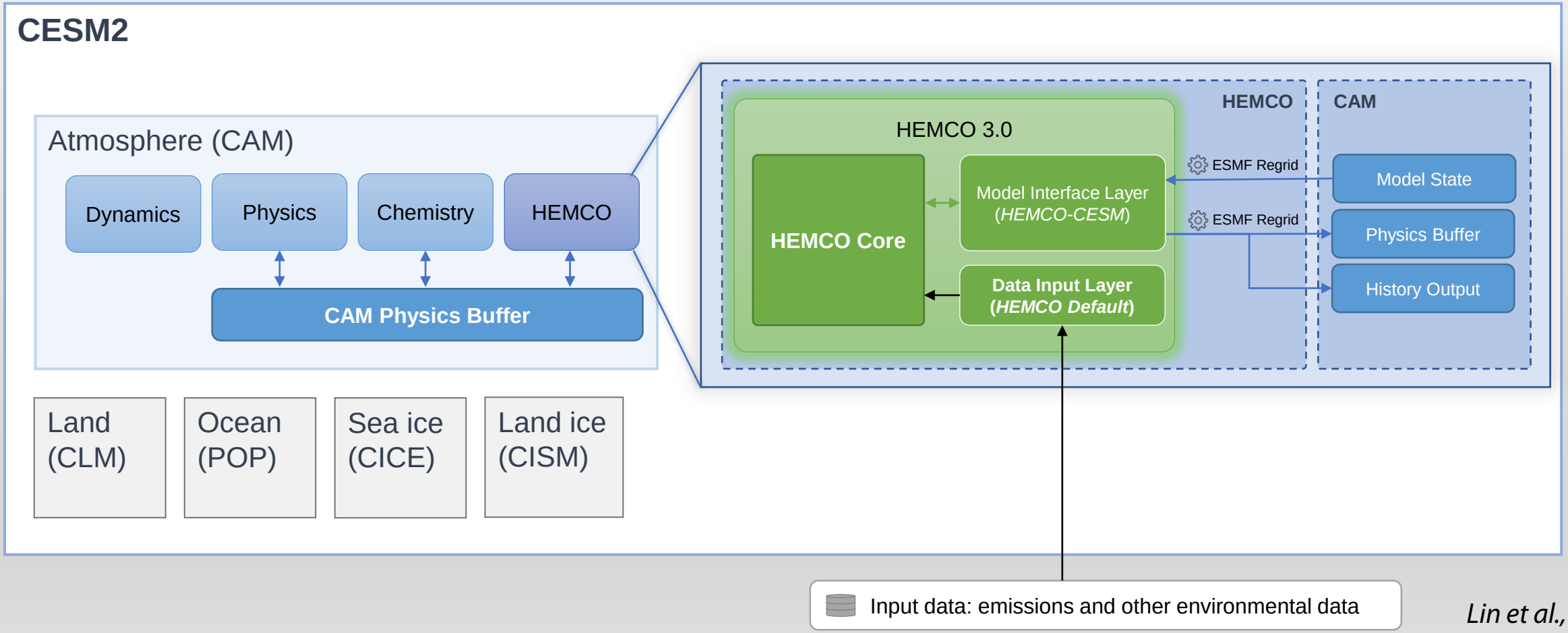
HEMCO restructured as a multi-model data tool



Lin et al., 2021

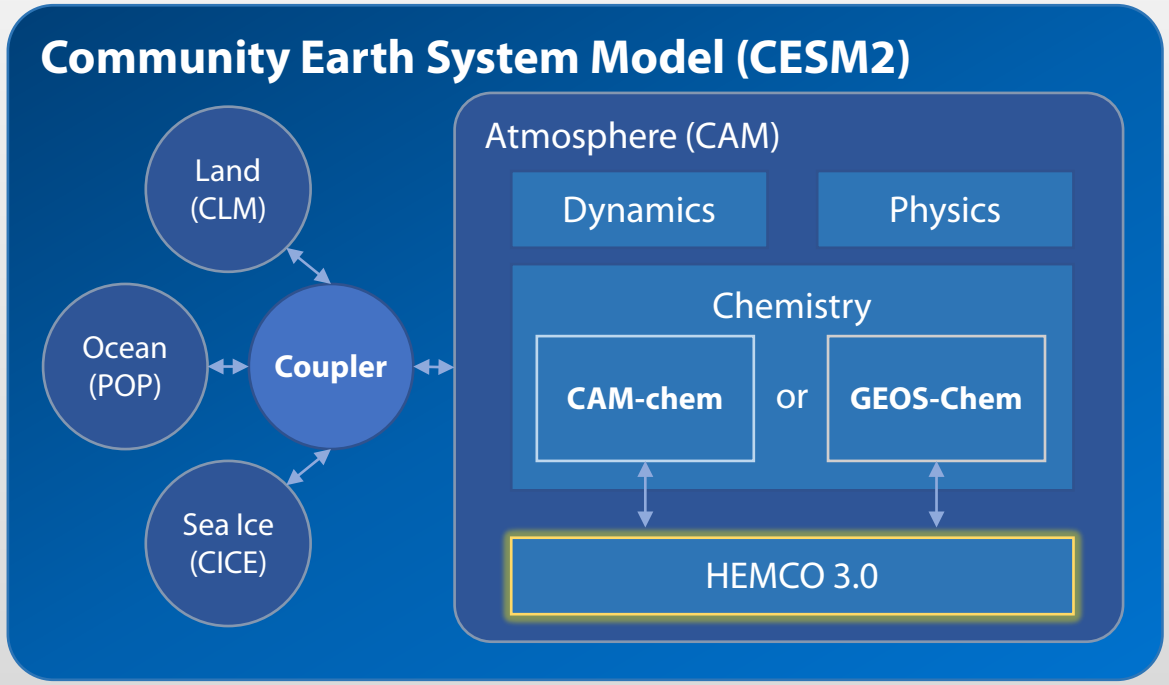
HEMCO implementation within CAM – as a separate component

HEMCO as a separate component in the atmosphere allows serving of data to any atmospheric component, independent of "the outside world"



Lin et al., 2021

HEMCO operates on a separate grid for accuracy and consistency



HEMCO is a separate component in CAM

- Has own grid ("HEMCO grid")
  - Consistent emissions regardless of model resolution
  - Re-grid computed results to model resolution at every time step
- Can serve both CAM-chem or GEOS-Chem chemistry in CESM

Adapted from NSF proposal; GEOS-Chem within CESM2 is Fritz et al., 2022 (in press)

Setting up: Installing CESM with HEMCO enabled

This is temporary until CESM "trunk" includes HEMCO.

The instructions are based off CESM-2.2 but if the CAM version is close (`cam6_3_045`), you may also be able to use your own version.

- Checkout ESCOMP/CESM at `release-cesm2.2.0` release.

```
$ git clone https://github.com/ESCOMP/CESM.git
$ cd CESM; git checkout release-cesm2.2.0
```

- Edit `Externals.cfg` to replace the [cam] external

```
[cam]
tag = cam6_3_045_hemco-cesm1_0_1_hemco3_5_1
protocol = git
repo_url = https://github.com/CESM-GC/CAM
local_path = components/cam
externals = Externals_CAM.cfg
required = True
```

Setting up: Installing CESM with HEMCO enabled

- Checkout externals

```
$ ./manage_externals/checkout_externals
```

- Create case with HEMCO compsets

  (FCSD_HCO, FCHIST_HCO, FCnudged_HCO, …)

```
$ cd cime/scripts
$ ./create_newcase --case ~/dev-2211_hemco_compset_test --compset FCSD_HCO --res f09_f09_mg17 --
run-unsupported --project project_id --mach cheyenne
```

* Special note: Remove two files from source

Remove two files in CESM/components/cam/src/hemco/HEMCO/src/Core

```
$ cd components/cam/src/hemco/HEMCO/src/Core
$ rm hcoio_read_mapl_mod.F90
$ rm hcoio_write_mapl_mod.F90
```

This will be fixed in a future version of CIME (`cime6.0.56`+), but not in CESM-2.2.

Setting up: Configuring CAM and the HEMCO-CESM interface

Modifying `user_nl_cam`

- Disable the off-line emissions:

```
ext_frc_specifier=''
srf_emis_specifier=''
```

- Configure HEMCO: (Grid: 288x201 = 0.9x1.25, 2400x1201 = 0.15x0.15…)

```
&hemco_nl
 hemco_config_file = '/glade/scratch/hplin/2211_camchem_2.2_ne0korus/run/HEMCO_Config.rc',
 hemco_grid_xdim = 288,
 hemco_grid_ydim = 201,
/
```

Setting up: Building the case

- Require building CESM with ESMF (* step won't be needed in future CESM versions)

```
$ ./xmlchange USE_ESMF_LIB=true
```

- Setup and build the case

```
$ ./case.setup --reset
$ qcmd -l select=1:ncpus=18 -q share -- ./case.build -v
```

Now patiently wait for as long as CESM usually takes to compile…

Setting up: Configuring CAM and the HEMCO-CESM interface

Modifying `user_nl_cam`

- Provide CESM with the ESMF mesh file describing the CAM grid (used for regridding)

```
cam_physics_mesh = '/glade/p/cesmdata/cseg/inputdata/atm/cam/coords/fv0.9x1.25_esmf_141008.nc'
```

Known grids:

- F09: `/glade/p/cesmdata/cseg/inputdata/atm/cam/coords/fv0.9x1.25_esmf_141008.nc`
- F19: `/glade/p/cesmdata/cseg/inputdata/atm/cam/coords/fv1.9x2.5_esmf_200428.nc`
- NE30NP4: `/glade/p/cesmdata/cseg/inputdata/atm/cam/coords/ne30np4_esmf_191122.nc`

Setting up: Configuring CAM and the HEMCO-CESM interface

Moving the configuration files to the $scratch/$case/run directory.

```
cd /glade/scratch/$USER/my_case_name/run
wget https://raw.githubusercontent.com/jimmielin/HEMCO_CESM_configs/master/CAM-Chem/HEMCO_Diagn.rc
wget https://raw.githubusercontent.com/jimmielin/HEMCO_CESM_configs/master/CAM-
Chem/CMIP6_emissions_2000climo/3.5.0/HEMCO_Config.rc
```

These include global default inventories from: (can always enable/disable any!)

- CEDSv2 *(Hoesly et al., 2018; McDuffie et al., 2020)*, 0.5x0.5deg
- AEIC 2019 Aircraft Emissions *(Eastham and Fritz)*, 0.5x0.625deg
- GEIA NH_3 *(Bouwman et al., 1997)*, Arctic seabird NH_3 *(Croft et al., 2016)*, Fossil fuel and biofuel C_2H_6 *(Tzompa-Soza et al., 2017)*, C_3H_8 *(Xiao et al., 2008)*, Bromocarbons *(Liang et al., 2010; Ordonez et al., 2012; Sherwen et al., 2016)*.
- GFED4 biomass burning emissions

Setting up: Configuring HEMCO in **HEMCO_Config.rc**

Root data location: paths beginning with $ROOT in the configuration file resolve here

```
ROOT:                          /glade/p/univ/umit0034/ExtData/HEMCO
```

You can get most default inventories from $ROOT but you can also use absolute paths for your own data:

```
0 DICE_CARS_CO    $ROOT/DICE_Africa/v2016-10/DICE-Africa-cars-2013-v01-4Oct2016.nc
CO                     2013/1/1/0         C xy    g/m2/yr  CO    26/1008              1   60

0 SOME_INVENTORY  /glade/u/home/hplin/some.nc
NO_agriculture         2000-2017/1-12/1/0 C xy    kg/m2/s  NO    301/306/1009         1/2 56
```

Setting up: Configuring HEMCO in **HEMCO_Config.rc**

Root data location: paths beginning with $ROOT in the configuration file resolve here

```
ROOT:                       /glade/p/univ/umit0034/ExtData/HEMCO
```

Inventory switches:

```
# ----- REGIONAL INVENTORIES ---------------------------------------------
    --> APEI                  :       false    # 1989-2014
    --> NEI2016_MONMEAN       :       false    # 2002-2020
    --> DICE_Africa           :       false    # 2013
# ----- GLOBAL INVENTORIES -----------------------------------------------
    --> CEDSv2                :       true     # 1750-2019
    --> EDGARv43              :       false    # 1970-2010
    --> HTAP                  :       false    # 2008-2010
…
```

Warning: Inventories were developed for GEOS-Chem species! Non-CEDSv2 inventories may need some work to map species correctly for MOZART-TS1

Notes on inventory support (support = mapping the species correctly!)

This might be trivial:

```
0 CEDS_NO_AGR  NO-em-anthro_CMIP_CEDS_$YYYY.nc NO_agr 1750-2019/1-12/1/0 C xy kg/m2/s NO 25 1 5
```

But this may not be!

```
0 CEDS_SO2_ENE SO2-em-anthro_CMIP_CEDS_$YYYY.nc SO2_ene 1750-2019/1-12/1/0 C xyL=100m:300m kg/m2/s
SO2 - 1 5
0 CEDS_SO4_ENE -                                -       -                   - xyL=100m:300m - so4_a1
8907  1 5

...
# --- CEDS to CAM-Chem scale factors ---
8907 CESM_SO4a1_ENE  0.025    - - - xy 1 1
```

Table S4. MAM4 aerosol parameters for emissions that are specified with emissions files (dust and sea salt are calculated online). Original species refers to labels in emissions inventories.

so4_a1	SO2*0.025 energy, industrial	115	1770	0.261	100-300m

HEMCO is not magic

It's up to you to specify the mapping Between netCDF file and model emiss.!

Then just run!

- Run CAM-chem/MUSICA as usual.

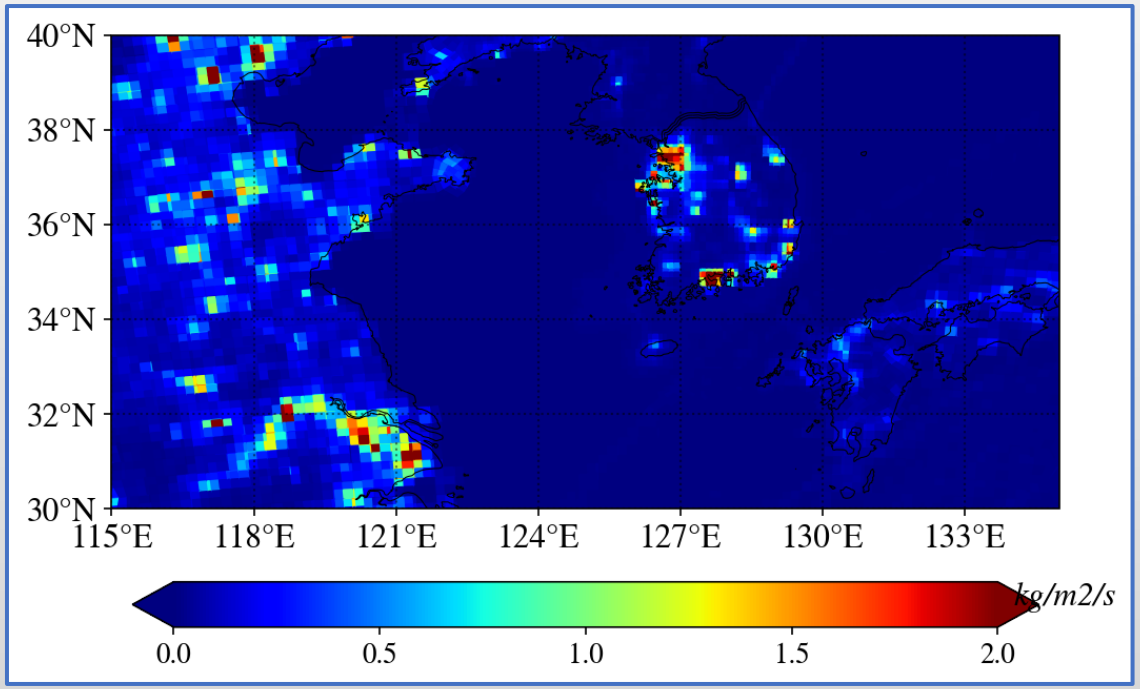
```
$ ./case.submit
```

- Where to look for HEMCO log output:
  - *atm.log…* will show major steps that HEMCO takes. If there is an error in the interface to CESM, it'll error out here.
  - *cesm.log…* and *HEMCO.log* will show HEMCO processing logs, including files read. If HEMCO itself crashes errors will be found here.
- HEMCO emissions are passed to chemistry starting from the **second** model timestep.

Can I see HEMCO output?

- All HEMCO emissions in kg/m2/s are available in the master field list
- HCO_NO, HCO_NO2, etc.
- Available over the entire vertical if emissions are 3-D.

HEMCO-CESM output (HCO_NO)
On MUSICA Korea grid, HEMCO internal res 0.15x0.15deg



Using Vivaldi-a package for Plot_2D
Credit to Duseong Jo for Korea MUSICA Grid

Are there checks and diagnostics available, apart from emission totals?

- Not currently in CESM, but can be implemented in the HEMCO interface (let me know if there is demand!)
- e.g., can be used to get emissions information per sector / extension

# Name	Spec	ExtNr	Cat	Hier	Dim	OutUnit	LongName
EmisCO_Total	CO	-1	-1	-1	3	kg/m2/s	CO_emission_flux_from_all_sectors
EmisCO_Aircraft	CO	0	20	-1	3	kg/m2/s	CO_emission_flux_from_aircraft
EmisCO_Anthro	CO	0	1	-1	3	kg/m2/s	CO_emission_flux_from_anthropogenic
EmisCO_BioBurn	CO	111	-1	-1	2	kg/m2/s	CO_emission_flux_from_biomass_burning
EmisCO_Ship	CO	0	10	-1	2	kg/m2/s	CO_emission_flux_from_ships

This is available in HEMCO code through a call to `subroutine GetHcoDiagn`, but some work to regrid it to CAM history and output is needed.

What if I want to run HEMCO *outside* of CESM?

https://hemco.readthedocs.io/en/stable/hco-sa-guide/intro.html

- HEMCO has an **off-line version** as well, if you just want to try it out.
- Can reuse all the HEMCO configuration file(s) from other models
- Specify species properties…

```
#ID NAME            MW      K0                  CR        PKA
1   NO              30.00   0.000000E+00        0.00      0.00
2   O3              48.00   0.000000E+00        0.00      0.00
3   PAN            121.00   0.000000E+00        0.00      0.00
4   CO              28.00   0.000000E+00        0.00      0.00
```

- Grid specification…

```
XMIN: -182.5
XMAX:  177.5
YMIN: -90.0
YMAX:  90.0
NX: 72 ...
```

OK, I want to add inventories/features to HEMCO/HEMCO-CESM!

- Thank you! ☺
- Git repositories:
  - **HEMCO** Core code, including Extensions: https://github.com/geoschem/HEMCO/
  - **HEMCO-CESM** interface: https://github.com/ESCOMP/HEMCO_CESM
  - **Configuration files** for CAM-chem/GEOS-Chem within CESM: https://github.com/jimmielin/HEMCO_CESM_configs/
- Please feel free to reach out to me with any questions or bugs!
- Generic HEMCO questions can be asked at the HEMCO GitHub as well (supported by the GEOS-Chem Support Team)

Many useful resources

- HEMCO User's Guide: https://hemco.readthedocs.io
- HEMCO references:
  - HEMCO 3.0 – implementation in several models, including CESM: https://gmd.copernicus.org/articles/14/5487/2021/ (Lin et al., 2021)
  - HEMCO 1.0 – original reference: https://gmd.copernicus.org/articles/7/1409/2014/ (Keller et al., 2014)
- HEMCO in MUSICA:
  - Pull request with setup instructions: https://github.com/ESCOMP/CAM/pull/560
- My email: hplin@seas.harvard.edu

What about point sources?

- HEMCO Core does not support point sources but the *Volcano* extension does:

```
###  LAT (-90,90), LON (-180,180), SULFUR [kg S/s], ELEVATION [m], CLOUD_COLUMN_HEIGHT [m]
### If elevation=cloud_column_height, emit in layer of elevation
### else, emit in top 1/3 of cloud_column_height
volcano::
10.030 -83.770 1.934297e+00 3340 3340
-37.860 -71.160 4.756469e+00 2800 2800
19.420 -155.290 2.736555e+01 1222 1222
3.170 98.390 1.192288e+01 2460 2460
-15.800 -71.860 1.331811e+00 5967 5967
-15.400 167.830 1.786847e+01 1395 1395
..
```