GEOS-Chem for advanced users and introduction to HEMCO

Christoph Keller and Bob Yantosca
School of Engineering and Applied Sciences
Harvard University
Cambridge, MA, USA
IGC7, 04 May 2015

Topics

- Quick overview of how HEMCO works
- Everything about the HEMCO configuration file
  - Base emissions, scale factors, masking, overlaying, combining emissions
  - Live examples
- Working on the GEOS-Chem HEMCO interface
  - Extensions, diagnostics, etc.
- NetCDF file format and manipulation
  - COARDS netCDF conventions

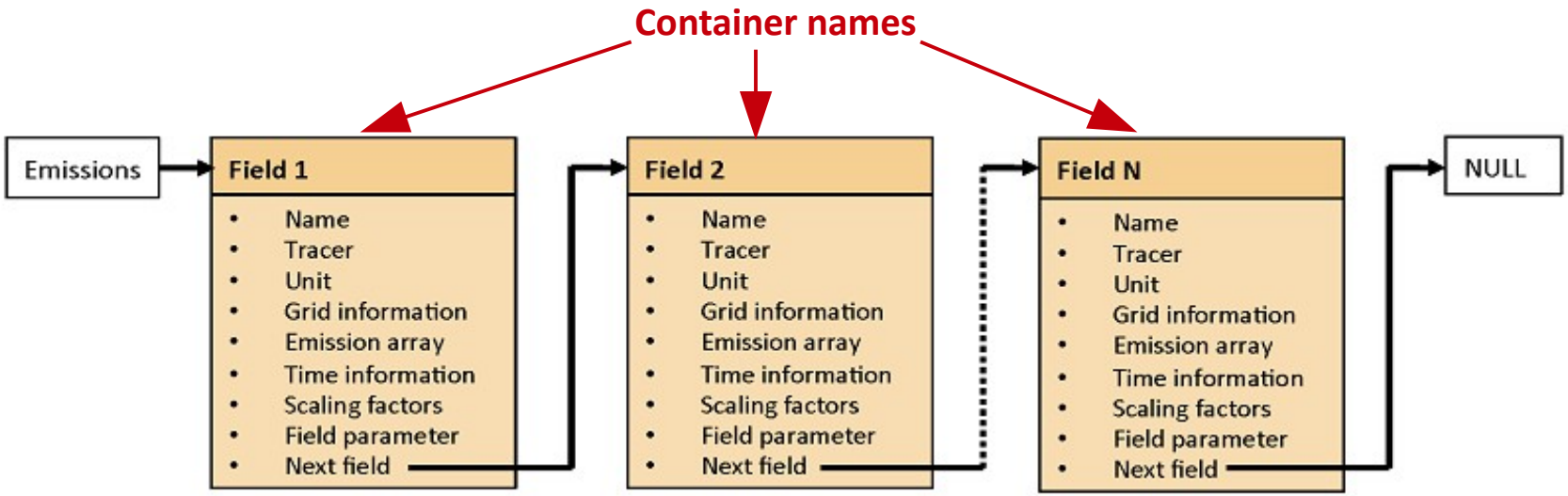
We are going to be throwing a lot of information at you now. Don't worry if you don't absorb it all at once.

We will post this presentation online for your reference. So don't worry about writing it all down now...

HEMCO: Harvard-NASA emissions component

- HEMCO now supplies emissions to GEOS-Chem (in v10-01+)
  - Reads data from COARDS-netCDF files (more about this later)
  - HEMCO reads emissions data at the highest possible resolution. Data is regridded in the horizontal/vertical.
- HEMCO can read:
  - Emissions data (fluxes in kg/m2/s, scale factors, etc.)
  - Non-emissions data (OH conc's, P/L rates, etc.)
- HEMCO can archive diagnostics to netCDF if desired
  - G-C v10-01 will still use binary format diagnostic output
  - G-C will eventually use HEMCO's data structure for all diagnostics!

The emissions linked list in HEMCO

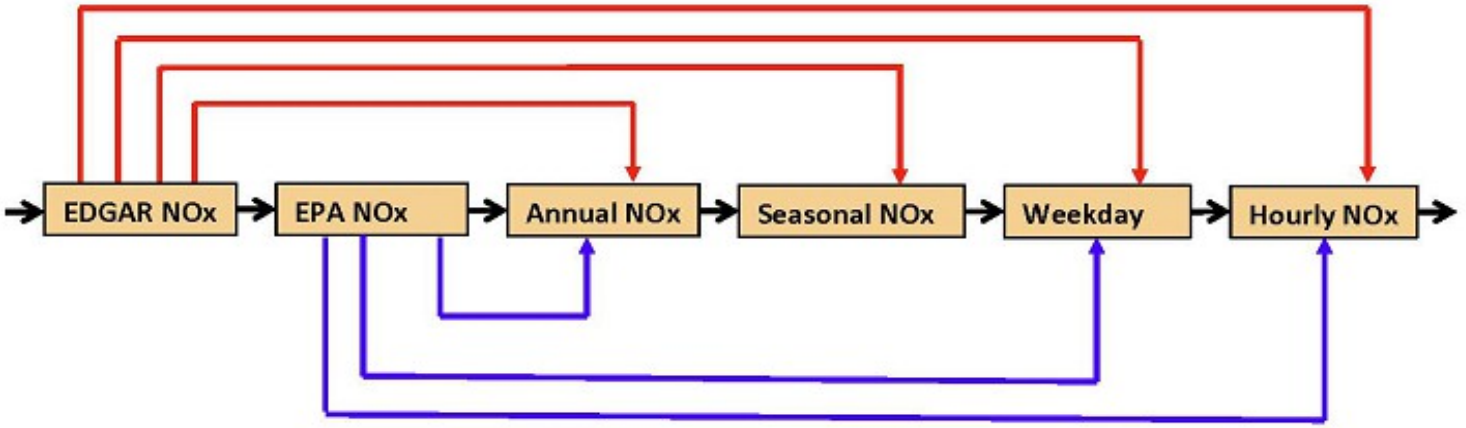


- List with all emission data
- Flexible length
- One variable ('Emissions') leads to all content
- Contains all information required to calculate the emissions

The heart and soul of HEMCO is the **linked list** structure. The linked list consists of multiple **containers**. Each **container** can store an **emissions inventory** (global or regional), a **non-emissions data set**, a **scale factor**, a **mask**, the **total emissions for a given species**, or a **diagnostic output**.

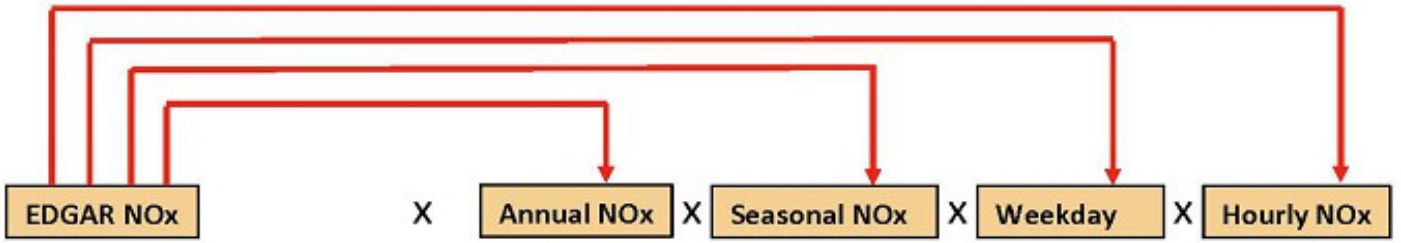
Courtesy of Christoph Keller

Applying the scale factors: Pointers!



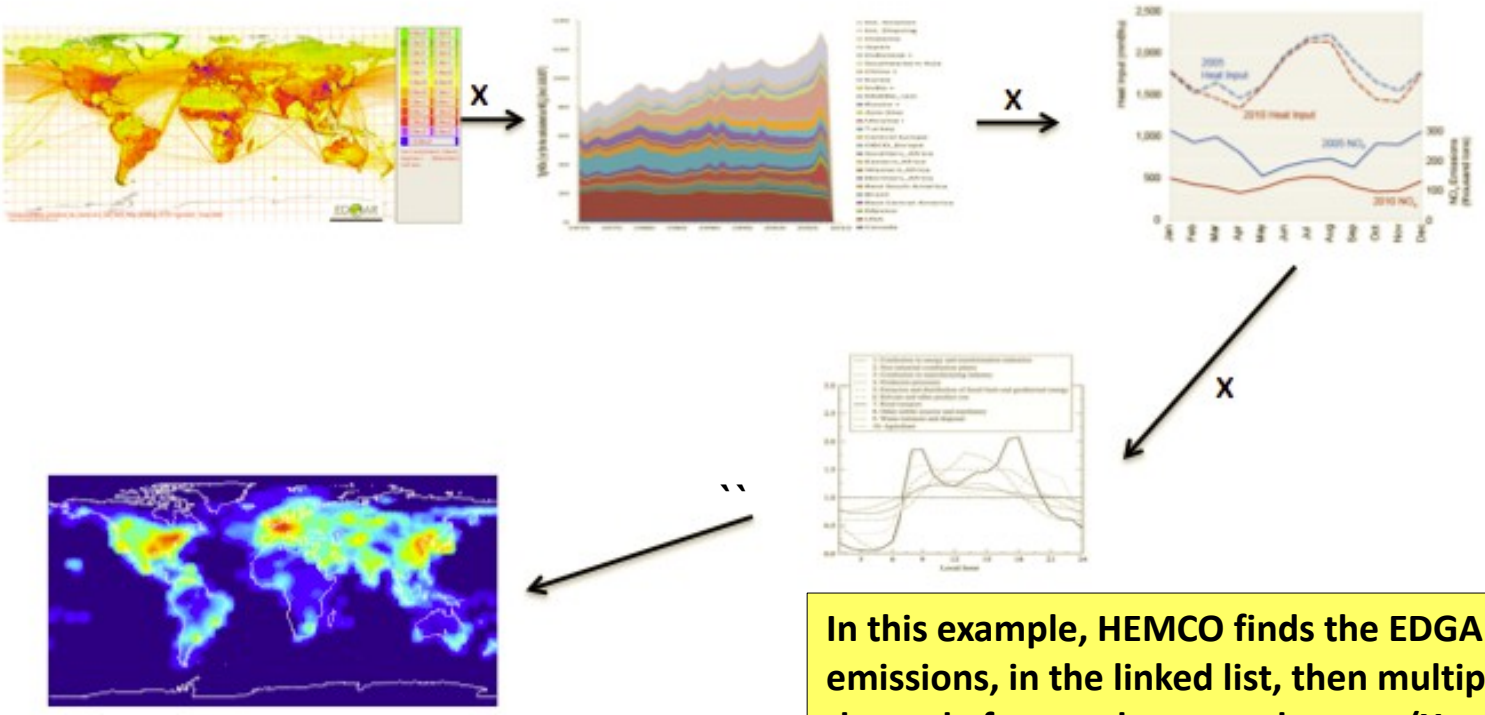
For each inventory, HEMCO applies the appropriate masks and scale factors by "pointing" to the containers where they are located.

EDGAR NOx for given time step:



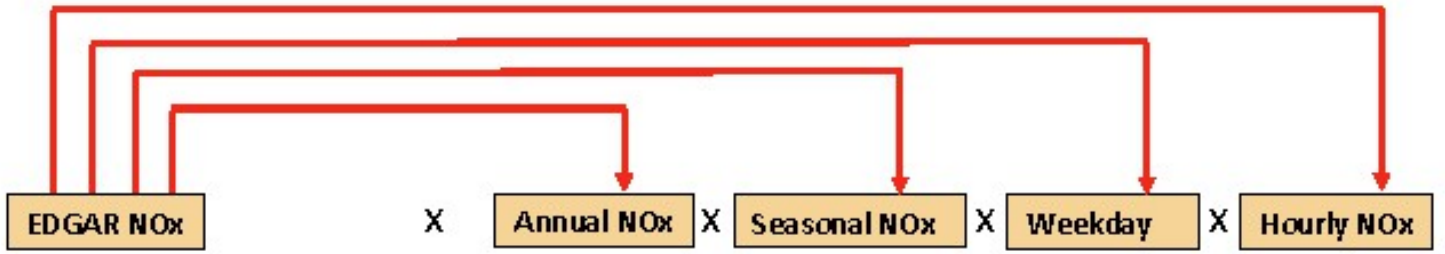
Courtesy of Christoph Keller

Applying the scale factors: Pointers!



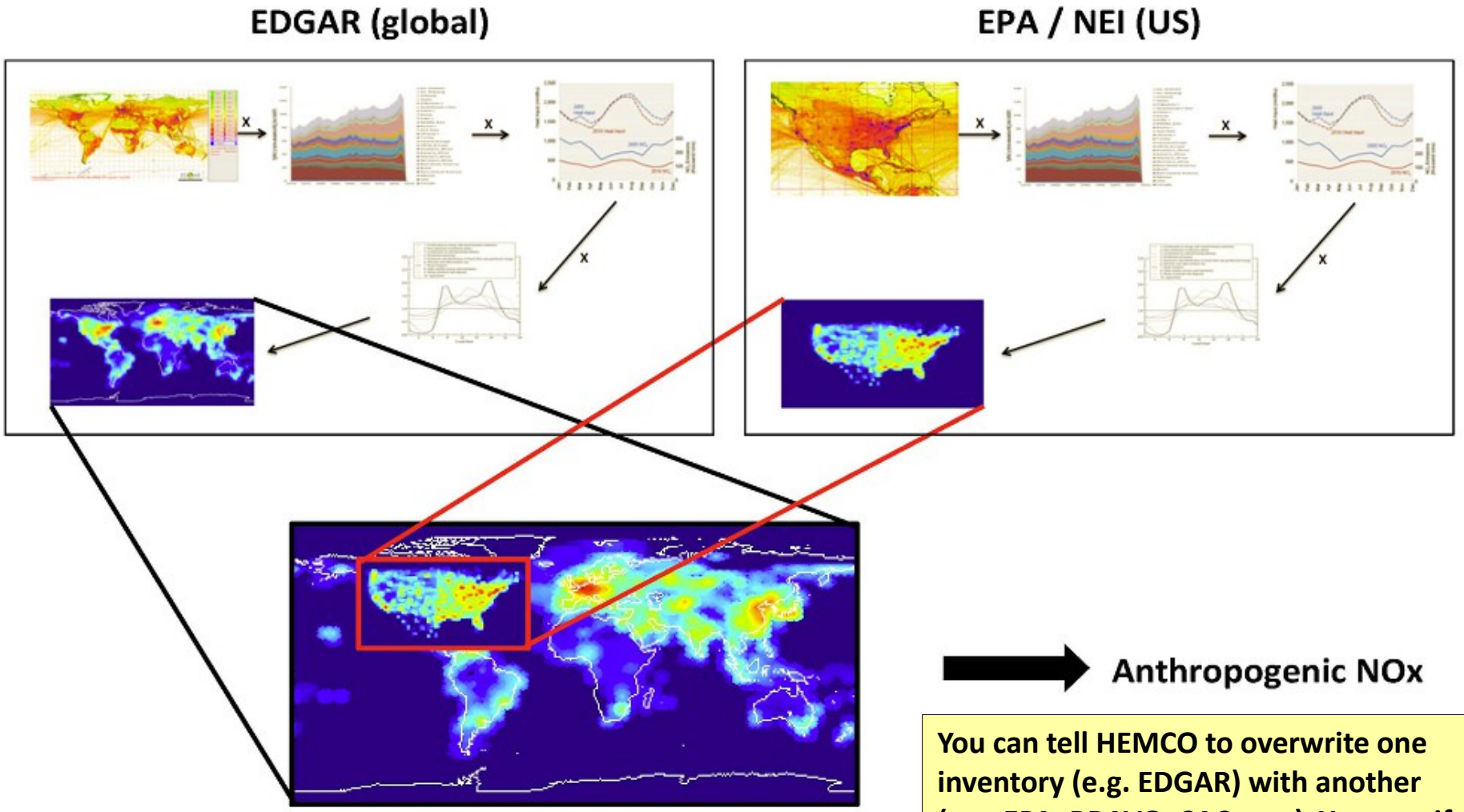
In this example, HEMCO finds the EDGAR base emissions, in the linked list, then multiplies it only by the scale factors that are relevant. (You specify scale factors for each inventory in the configuration file.)

EDGAR NOx for given time step:



Courtesy of Christoph Keller

Combining regional and global inventories with HEMCO



You can tell HEMCO to overwrite one inventory (e.g. EDGAR) with another (e.g. EPA, BRAVO, CAC, etc.) You specify these settings in the configuration file.

Courtesy of Christoph Keller

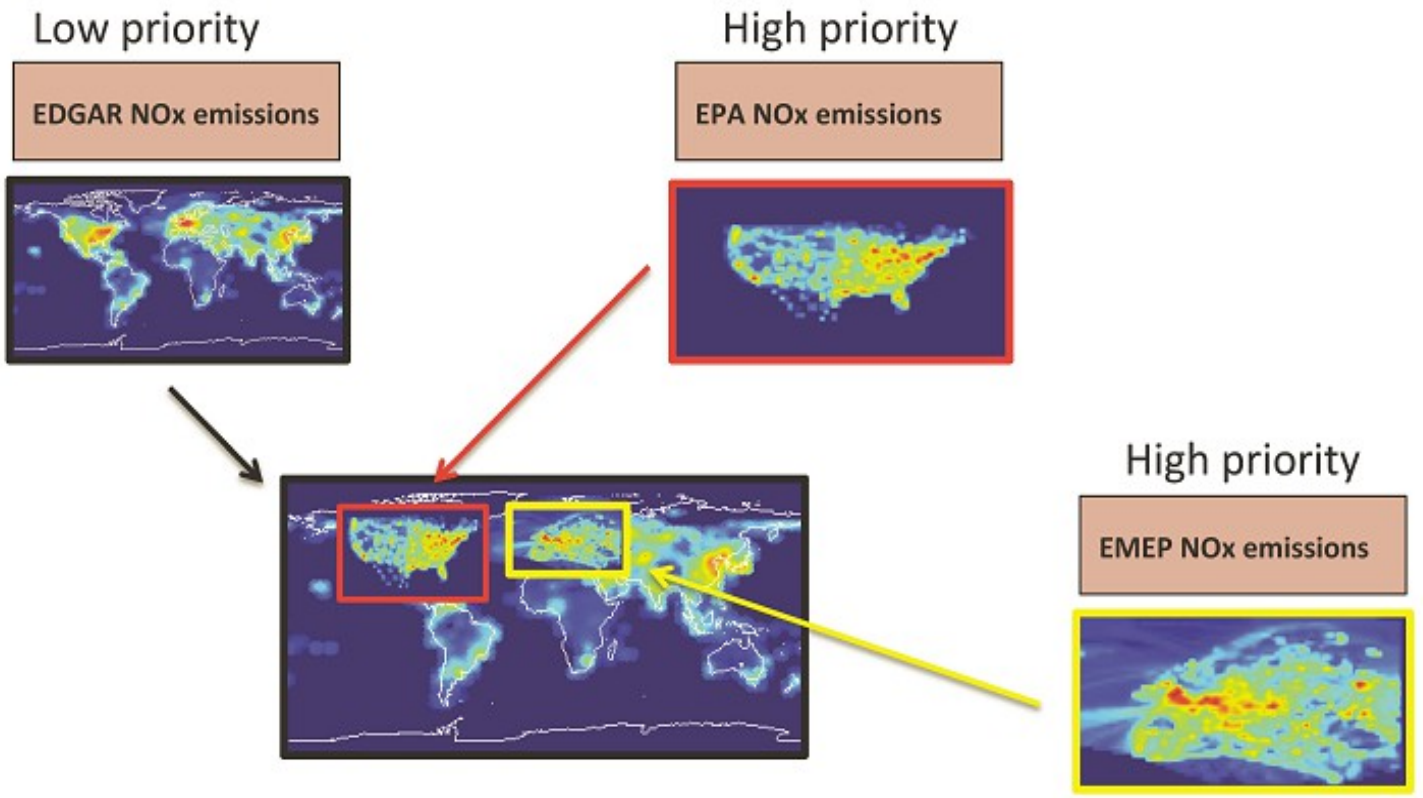
Field priorities

- Hierarchy of emissions defined through field priorities
- High-priority emissions overwrite low-priority emissions

In the HEMCO configuration file, you can specify the priorities for each species in each inventory.

Emissions of different *categories* are added together (e.g. anthro + biomass NO).

Emissions in the same category (e.g. anthro NO) may be assigned a *hierarchy* (higher values overwrite lower values).



Courtesy of Christoph Keller

HEMCO uses a separate configuration file

- The `input.geos` file now looks like this:

```
------------------------+----------------------------------
%%% EMISSIONS MENU %%%  :
Turn on emissions?      : T
Emiss timestep (min)    : 60
HEMCO Input file        : HEMCO_Config.rc
...etc...
```

- The `input.geos` file now only specifies the emission timestep and points to the HEMCO configuration file.
- The `HEMCO_Config.rc` file now contains all emissions options. You can edit this for your simulation.
- Complexity is now contained in the `HEMCO_Config.rc` file, instead of being hardwired in source code (as in v9-02 etc.)

HEMCO configuration file: *Base Emissions Switches*

```
# ExtNr ExtName           on/off  Species
0       Base              : on    *
    --> AEIC              :       true
    --> BIOFUEL           :       true
    --> BOND              :       true
    --> BOND_BIOMASS      :       false
    --> BRAVO             :       true
    --> CAC               :       true
    --> C2H6              :       true
    --> EDGAR             :       true
    --> HTAP              :       false
    --> EMEP              :       true
    --> GEIA              :       true
    --> LIANG_BROMOCARB   :       true
    --> NEI2005           :       false
    --> NEI2011           :       true
    --> NEI2011_SHIP      :       false
    --> RETRO             :       true
    --> SHIP              :       true
    --> SHIPNO_BASE       :       false
    --> MIX               :       true
    --> STREETS           :       false
    --> VOLCANO_ERUPT     :       true
    --> VOLCANO_DEGAS     :       true
    --> RCP_3PD           :       false
    --> RCP_45            :       false
    --> RCP_60            :       false
    --> RCP_85            :       false
```

HEMCO computes emissions in two ways:

Base Emissions are emissions that can just be read from disk. Most emission inventories fall into this category.

Some emissions are implemented as **HEMCO Extensions**. In this case HEMCO has to compute these emissions using meteorological inputs (P, T, RH, U, V, etc.). More on these in the next few slides.

The list at left shows the various input options in the **Base Emissions**. This list appears at the top of the ***Base Emissions Switches*** section in the `HEMCO_Config.rc` file. You can toggle individual emission inventories on or off.

For example, if you wanted to turn off the AEIC aircraft inventory, you would change its setting under the Species column from true to false.

HEMCO configuration file: *Base Emissions Data*

```
#==============================================================================
# --- AEIC aircraft emissions ---
#
# ==> Now emit aircraft BC and OC into hydrophilic tracers BCPI and OCPI.
#==============================================================================
(((AEIC
0 AEIC_NO   $ROOT/AEIC/v2014-10/aeic_2005.geos.1x1.47L.nc  NO       2005/1-12/1/0 C xyz kg/m2/s NO   110/115 20 1
0 AEIC_CO   $ROOT/AEIC/v2014-10/aeic_2005.geos.1x1.47L.nc  CO       2005/1-12/1/0 C xyz kg/m2/s CO   110     20 1
0 AEIC_SO2  $ROOT/AEIC/v2014-10/aeic_2005.geos.1x1.47L.nc  FUELBURN 2005/1-12/1/0 C xyz kg/m2/s SO2  111     20 1
0 AEIC_SO4  -                                              -        -             - -   -       SO4  112     20 1
0 AEIC_BCPI -                                              -        -             - -   -       BCPI 113     20 1
0 AEIC_OCPI -                                              -        -             - -   -       OCPI 113     20 1
0 AEIC_ACET $ROOT/AEIC/v2014-10/aeic_2005.geos.1x1.47L.nc  HC       2005/1-12/1/0 C xyz kg/m2/s ACET 114/101 20 1
0 AEIC_ALD2 -                                              -        -             - -   -       ALD2 114/102 20 1
0 AEIC_ALK4 -                                              -        -             - -   -       ALK4 114/103 20 1

0 AEIC_C2H6 -                                              -        -             - -   -       C2H6 114/104 20 1
0 AEIC_C3H8 -                                              -        -             - -   -       C3H8 114/105 20 1
0 AEIC_CH2O -                                              -        -             - -   -       CH2O 114/106 20 1
0 AEIC_PRPE -                                              -        -             - -   -       PRPE 114/107 20 1
0 AEIC_MACR -                                              -        -             - -   -       MACR 114/108 20 1
0 AEIC_RCHO -                                              -        -             - -   -       RCHO 114/109 20 1
)))AEIC
```

Further down in the HEMCO_Config.rc file (in the ***Base Emissions Data*** section), you will find the entry for the AEIC aircraft emissions. Each of these lines indicates that you want HEMCO to read a specific species from the AEIC inventory. For each species, you list **the container name**, **the data file** to be read, the **name of the variable** in the file, the **time range**, the **units**, the **G-C species to which these emissions will be added**, any **scale factors** you wish to apply, and the **category and hierarchy** of the data.

All of the entries between the brackets (((AEIC and)))AEIC will be ignored if you turned off AEIC emissions in the ***Base Emissions Switches*** section shown on the previous slide.

HEMCO configuration file: *Extension Switches*

```
# ------------------------------------------------------------------------------
100     Custom            : off   -
101     SeaFlux           : on    DMS/ACET
102     ParaNOx           : on    NO/NO2/O3/HNO3
    --> LUT data format   :       nc
    --> LUT source dir    :       $ROOT/PARANOX/v2015-02
103     LightNOx          : on    NO
    --> OTD-LIS factors   :       true
    --> CDF table         :       $ROOT/LIGHTNOX/v2014-07/light_dist.ott2010.dat
104     SoilNOx           : on    NO
    --> Use fertilizer NOx:       true
105     DustDead          : on    DST1/DST2/DST3/DST4
106     DustGinoux        : off   DST1/DST2/DST3/DST4
107     SeaSalt           : on    SALA/SALC/Br2
    --> SALA lower radius :       0.01
    --> SALA upper radius :       0.5
    --> SALC lower radius :       0.5
    --> SALC upper radius :       8.0
    --> Emit Br2          :       true
    --> Br2 scaling       :       1.0
108     MEGAN             : on    ISOP/ACET/PRPE/C2H4/ALD2
    --> Isoprene scaling  :       1.0
109     MEGAN_Mono        : on    CO/OCPI/MONX
110     MEGAN_SOA         : off   MTPA/MTPO/LIMO/SESQ
111     GFED              : on    NO/CO/ALK4/ACET/MEK/ALD2/PRPE/C3H8/CH2O/C2H6/SO2/NH3/BCPO/BCPI/OCPO/OCPI/POA1/NAP
    --> GFED3             :       false
    --> GFED4             :       true
    --> GFED_daily        :       false
    --> GFED_3hourly      :       false
    --> CO scale factor   :       1.05
    --> POA scale factor  :       1.27
    --> NAP scale factor  :       2.75e-4
    --> hydrophilic BC    :       0.2
    --> hydrophilic OC    :       0.5
...etc...
```

The ***Extension Switches Section*** (back up near the top of the HEMCO_Config.rc file) lets you select options for the various HEMCO Extensions. You specify the **extension number and name**, whether you want to **turn an extension on or off**, which **species to use with that extension**, and other **options used by the extension**.

HEMCO configuration file: *Extension data*

```
#==============================================================================
# --- GFED biomass burning emissions (Extension 111)
# NOTE: These are the base emissions in kgDM/m2/s.
#==============================================================================
111 GFED_HUMTROP    $ROOT/GFED3/v2014-10/GFED3_humtropmap.nc              humtrop           2000/1/1/0            C xy 1          * - 1 1

(((GFED3
111 GFED_WDL        $ROOT/GFED3/v2014-10/GFED3_gen.1x1.$YYYY.nc           GFED3_BB__WDL_DM  1997-2011/1-12/01/0   C xy kgDM/m2/s * - 1 1
111 GFED_AGW        $ROOT/GFED3/v2014-10/GFED3_gen.1x1.$YYYY.nc           GFED3_BB__AGW_DM  1997-2011/1-12/01/0   C xy kgDM/m2/s * - 1 1
111 GFED_DEF        $ROOT/GFED3/v2014-10/GFED3_gen.1x1.$YYYY.nc           GFED3_BB__DEF_DM  1997-2011/1-12/01/0   C xy kgDM/m2/s * - 1 1
111 GFED_FOR        $ROOT/GFED3/v2014-10/GFED3_gen.1x1.$YYYY.nc           GFED3_BB__FOR_DM  1997-2011/1-12/01/0   C xy kgDM/m2/s * - 1 1
111 GFED_PET        $ROOT/GFED3/v2014-10/GFED3_gen.1x1.$YYYY.nc           GFED3_BB__PET_DM  1997-2011/1-12/01/0   C xy kgDM/m2/s * - 1 1
111 GFED_SAV        $ROOT/GFED3/v2014-10/GFED3_gen.1x1.$YYYY.nc           GFED3_BB__SAV_DM  1997-2011/1-12/01/0   C xy kgDM/m2/s * - 1 1
)))GFED3

(((GFED4
111 GFED_WDL        $ROOT/GFED4/v2015-03/GFED4_gen.025x025.$YYYY.nc       WDL_DM            2000-2013/1-12/01/0   C xy kg/m2/s   * - 1 1
111 GFED_AGW        $ROOT/GFED4/v2015-03/GFED4_gen.025x025.$YYYY.nc       AGW_DM            2000-2013/1-12/01/0   C xy kg/m2/s   * - 1 1
111 GFED_DEF        $ROOT/GFED4/v2015-03/GFED4_gen.025x025.$YYYY.nc       DEF_DM            2000-2013/1-12/01/0   C xy kg/m2/s   * - 1 1
111 GFED_FOR        $ROOT/GFED4/v2015-03/GFED4_gen.025x025.$YYYY.nc       FOR_DM            2000-2013/1-12/01/0   C xy kg/m2/s   * - 1 1
111 GFED_PET        $ROOT/GFED4/v2015-03/GFED4_gen.025x025.$YYYY.nc       PET_DM            2000-2013/1-12/01/0   C xy kg/m2/s   * - 1 1
111 GFED_SAV        $ROOT/GFED4/v2015-03/GFED4_gen.025x025.$YYYY.nc       SAV_DM            2000-2013/1-12/01/0   C xy kg/m2/s   * - 1 1
)))GFED4

(((GFED_daily
111 GFED_FRAC_DAY   $ROOT/GFED3/v2014-10/GFED3_dailyfrac_gen.1x1.$YYYY.nc GFED3_BB__DAYFRAC 2002-2011/1-12/1-31/0  C xy 1          * - 1 1
)))GFED_daily

(((GFED_3hourly
111 GFED_FRAC_3HOUR $ROOT/GFED3/v2014-10/GFED3_3hrfrac_gen.1x1.$YYYY.nc   GFED3_BB__HRFRAC  2002-2011/1-12/01/0-23 C xy 1          * - 1 1
)))GFED_3hourly
```

Further down in the `HEMCO_Config.rc` file, there is an ***Extension Data Section***. In this section you list entries for data files that pertain to each of the HEMCO extensions. (You use the same syntax as for the ***Base Emissions Data*** section).

Here the GFED extension data is shown (GFED4 is highlighted). The brackets `(((GFED4` and `)))GFED4` are used to turn the GFED4 emissions on or off (depending on the settings in ***Extension Switches***).

HEMCO configuration file: *Scale Factors*

```
#==============================================================================
# --- temporary scale factors for comparisons
#==============================================================================
919 NO_ratio $ROOT/AnnualScalar/v2014-07/NO_ratio_2005_2002.nc NOXscalar 2005/1/1/0 C xy 1 1
918 CO_ratio $ROOT/AnnualScalar/v2014-07/CO_ratio_2005_1985.nc COscalar 2005/1/1/0 C xy 1 1

#==============================================================================
# --- day-of-week scale factors ---
# ==> data is Sun/Mon/.../Sat
#==============================================================================
20 GEIA_DOW_NOX 0.784/1.0706/1.0706/1.0706/1.0706/1.0706/0.863 - - - xy 1 1
21 GEIA_DOW_CO  0.683/1.1076/1.0706/1.0706/1.0706/1.0706/0.779 - - - xy 1 1
22 GEIA_DOW_HC  0.671/1.1102/1.1102/1.1102/1.1102/1.1102/0.768 - - - xy 1 1

#==============================================================================
# --- diurnal scale factors ---
#==============================================================================
25 EDGAR_TODNOX $ROOT/EDGARv42/v2015-02/NO/EDGAR_hourly_NOxScal.nc NOXscale 2000/1/1/HH C xy 1 1
26 GEIA_TOD_FOSSIL
0.45/0.45/0.6/0.6/0.6/0.6/1.45/1.45/1.45/1.45/1.4/1.4/1.4/1.4/1.45/1.45/1.45/1.45/0.65/0.65/0.65/0.65/0.45
/0.45 - - - xy 1 1
```

In the ***Scale Factors*** section of the `HEMCO_Config.rc` file, you may specify the scale factors that will be applied to the various emissions inventories listed in the ***Base Emissions Data*** and ***Extensions Data*** sections.

Scale factors can be:

- Gridded lon-lat data (read from a netCDF file)
- A list of values (i.e. global values for a sequence of years, months, or days)
- A single value

HEMCO configuration file: *Masks*

```
#==============================================================================
# Country/region masks
#==============================================================================
1000 EMEP_MASK    $ROOT/MASKS/v2014-07/EMEP_mask.geos.1x1.nc             MASK     2000/1/1/0 C xy 1 1 -30/30/45/70
1001 MEXICO_MASK  $ROOT/MASKS/v2014-07/BRAVO.MexicoMask.generic.1x1.nc   MASK     2000/1/1/0 C xy 1 1 -118/17/-95/33
1002 CANADA_MASK  $ROOT/MASKS/v2014-07/Canada_mask.geos.1x1.nc           MASK     2000/1/1/0 C xy 1 1 -141/40/-52/85
1003 SEASIA_MASK  $ROOT/MASKS/v2014-07/SE_Asia_mask.generic.1x1.nc       MASK     2000/1/1/0 C xy 1 1  60/-12/153/55
1004 NA_MASK      $ROOT/MASKS/v2014-07/NA_mask.geos.1x1.nc               MASK     2000/1/1/0 C xy 1 1 -165/10/-40/90
1005 USA_MASK     $ROOT/MASKS/v2014-07/usa.mask.nei2005.geos.1x1.nc      MASK     2000/1/1/0 C xy 1 1 -165/10/-40/90
1006 ASIA_MASK    $ROOT/MASKS/v2014-07/MIX_Asia_mask.generic.025x025.nc  MASK     2000/1/1/0 C xy 1 1  46/-12/180/82
1007 NEI11_MASK   $ROOT/MASKS/v2014-07/USA_LANDMASK_NEI2011_0.1x0.1.nc   LANDMASK 2000/1/1/0 C xy 1 1 -140/20/-50/60
```

In the ***Masks*** section, (near the end of the `HEMCO_Config.rc` file), you can define geographical (lon-lat) masks that are used with the regional emissions inventories. A mask is set to 1 where a regional inventory is used, and 0 where it is not used.

To specify a mask for a rectangular region, you can simply give the lon and lat of at the lower left and upper right corners. For irregularly-shaped regions, it is better to create a netCDF file for the mask.

For each mask, you specify the following:

- **A number and container name for the mask**
- **The file where the mask resides (optional)**
- **The name of the variable in the file**
- **The time range**
- **The units**
- **The longitude and latitude range of the mask.**

HEMCO configuration file: *Non-Emissions Data*

```
# --- Time zones (offset to UTC) ---
* TIMEZONES $ROOT/TIMEZONES/v2015-02/timezones_1x1.nc UTC_OFFSET 2000/1/1/0 C xy count * - 1 1

# --- UV albedo, for photolysis (cf Hermann & Celarier, 1997) ---
(((+UValbedo+
* UV_ALBEDO $ROOT/UVALBEDO/v2015-03/uvalbedo.geos.2x25.nc UVALBD 1985/1-12/1/0 C xy 1 * - 1 1
)))+UValbedo+

# --- TOMS/SBUV overhead ozone colums, for photolysis ---
(((+TOMS_SBUV_O3+
* TOMS_O3_COL   $ROOT/TOMS_SBUV/v2015-03/TOMS_O3col_$YYYY.geos.1x1.nc TOMS   1971-2010/1-12/1/0 C xy dobsons     * - 1 1
* DTOMS1_O3_COL -                                                     DTOMS1 -                  - -  dobsons/day * - 1 1
* DTOMS2_O3_COL -                                                     DTOMS2 -                  - -  dobsons/day * - 1 1
)))+TOMS_SBUV_O3+

# --- Linear stratospheric chemistry fields ---
# These fields will only be read if the +LinStratChem+ toggle is activated.

(((+LinStratChem+

# --- Stratospheric Bry data from the CCM model  ---
* GEOSCCM_Br_DAY      $ROOT/STRAT/v2015-01/Bry/GEOSCCM_Bry.2007$MM.day.nc   BR     2007/$MM/1/0 C xyz pptv * - 60 1
* GEOSCCM_Br_NIGHT    $ROOT/STRAT/v2015-01/Bry/GEOSCCM_Bry.2007$MM.night.nc BR     2007/$MM/1/0 C xyz pptv * - 60 1
... etc...


#--- GMI chemistry: prod/loss rates (for strato-/mesosphere) ---
* GMI_LOSS_A3O2       $ROOT/GMI/v2015-02/gmi.clim.A3O2.geos5.2x25.nc        loss    2000/$MM/1/0 C xyz s-1   A3O2     - 1  1
* GMI_PROD_A3O2       $ROOT/GMI/v2015-02/gmi.clim.A3O2.geos5.2x25.nc        prod    2000/$MM/1/0 C xyz v/v/s A3O2     - 1  1
... etc ...
)))+LinStratChem+
```

The ***Non-Emissions Data*** section is similar to ***Base Emissions Data*** section, with a couple of differences:

- The switches +UValbedo+ and +TOMS_SBUV_O3+ will be automatically set if FAST-JX is turned on.
- The switch +LinStratChem+ will be automatically set if stratospheric chemistry is turned on

This prevents errors caused by inconsistent input between `input.geos` and `HEMCO_Config.rc`.

HEMCO: Summary

- HEMCO now puts the complexity of combining emissions into a configuration file: no more hardwiring!
- HEMCO config files are different for each simulation
  - GCST will provide HEMCO config files sample run directories
  - More on this in a few slides!
- HEMCO's linked list can also track diagnostics
  - Emissions diagnostics are now tracked via HEMCO
  - But eventually all G-C diagnostics will be stored in the HEMCO linked list structure. (Coming in the next version …)
- Kudos to Christoph for coming up with all of this...

The standard G-C (fullchem) emissions configuration

Aerosols

AEROCOM volc SO2
Bond et al BC/OC
SOA emissions
Cooke et al BC/OC

Anthro

EDGAR v4.2
GEIA NH3
RETRO VOC
Y & L biofuels
BRAVO (Mexico)
CAC (Canada)
EMEP (Europe)
MIX (Asia)
NEI2011 (USA)
Xiao et al C2H6
MASAGE agric. NH3
HTAP
Streeets
GEIA/Piccot
EDGAR v3

Aircraft & Ship

AEIC aircraft
ARCTAS ship SO2
ICOADS ship CO
EMEP ship CO, NO, SO2
Corbett ship SO2
HTAP ship
NEI2011 ship
EDGAR v3 ship
EDGAR v4.2 ship

Biomass

GFED4 (monthly)
GFED3 (daily/3hourly)
FINN
QFED
GFED2
Duncan et al

Bromine

Liang et al VSL

HEMCO Extensions

Ocean exchange
DEAD dust model
MEGAN 2.1 + Guenther
Lightning NO
Hudman et al soil NO
PARANOX
GINOUX dust model

Future

RCP scenarios

Bromine

Liang et al VSL

Non-Emissions Data

GMI strat prod/loss
Strat Bry from CCM
Timezone offsets
TOMS/SBUV O3 col's
UV albedoes

Legend:

Data set is currently used in GC benchmarks

Data set is optional

Data set is either obsolete or not used for a specific reason

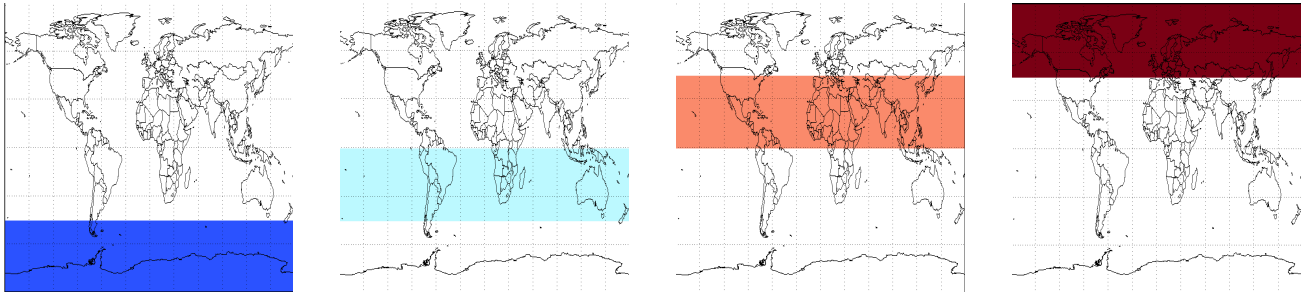
The standard G-C (fullchem) emissions configuration

- For more information about available emissions for use with GEOS-Chem via HEMCO, please see this wiki page:
  - `wiki.geos-chem.org/HEMCO_Data_Directories`

- If your system doesn't already have these emissions inventories, then you can download them with our hemco_data_download package.
  - `wiki.geos-chem.org/HEMCO_Data_Directories#Downloading_the_HEMCO_Data_Directories`

COARDS-compliant netCDF

- NetCDF is the preferred file format for atmospheric and climate communities
- NetCDF is a self-describing file format
  - Data
  - Metadata (info about data)
- NetCDF is useful for running in HPC environments
  - Data in a netCDF file can be compressed
  - Each CPU can read a subset of a netCDF file
  - Sequential binary format (e.g. bpch) not allowed in HPC



All CPU's can see the entire "world". In a parallel loop, each CPU only works on part of the "world" at a time.

Master CPU (is always active)

Beginning of OpenMP parallel loop

CPU 0 **CPU 1** **CPU 2** **CPU 3**

End of OpenMP parallel loop

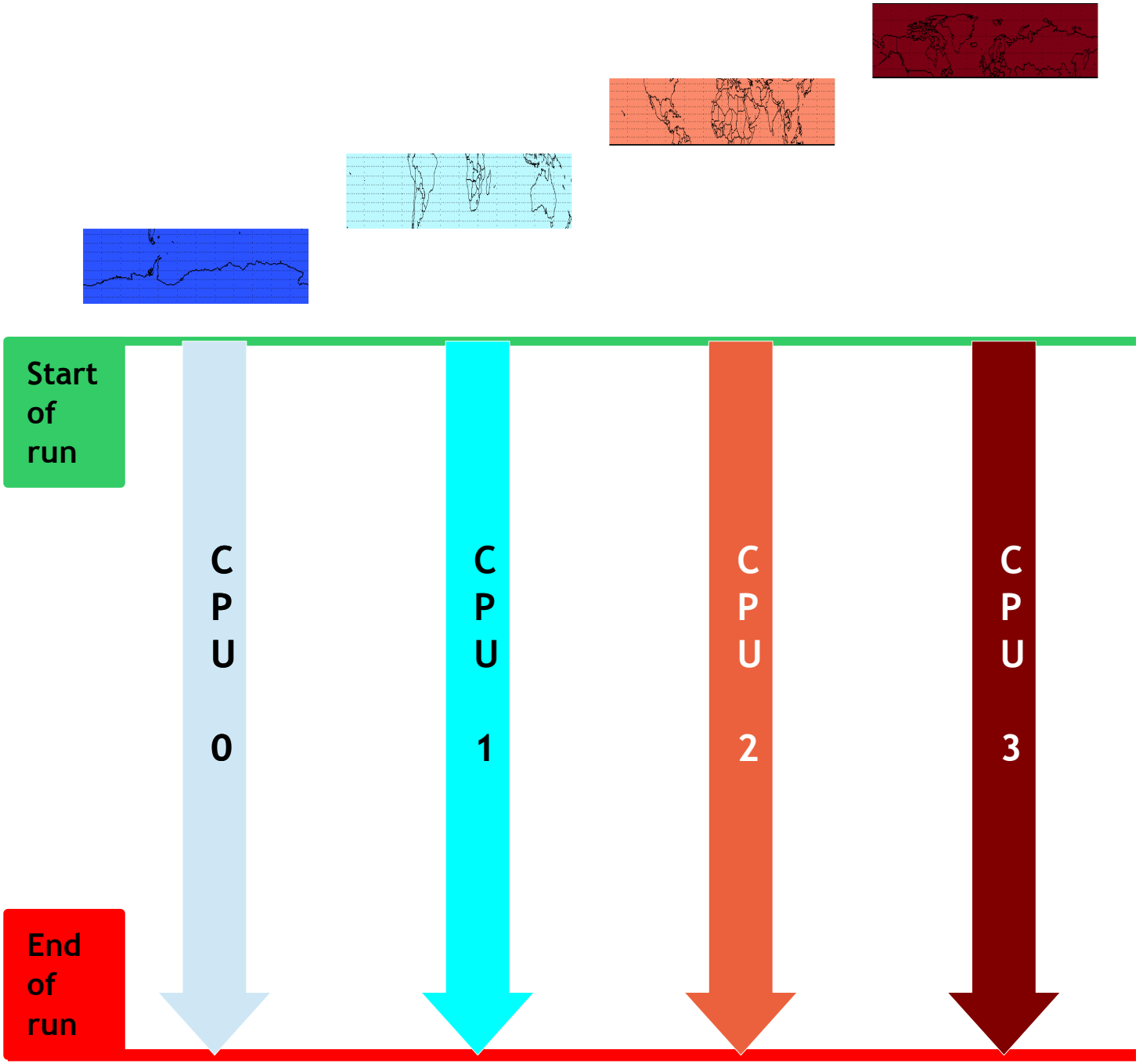
Current G-C "Classic" (with OpenMP)

The G-C simulation starts running on a single "master" CPU.

Extra CPUs are invoked to help share the work every time a parallel DO loop is encountered.

Extra CPUs go back to being idle at the end of the parallel loop.

Binary file I/O is done on the master CPU (outside of any parallelized loops).



Each CPU only can "see" a small part of the "world".

GEOS-Chem HPC with ESMF/MPI

Arrays are decomposed in lon & lat such that each CPU only gets a small piece of the "world" to work on. This can be a lon/lat range or a single column.

Each CPU runs G-C in parallel from start to finish using its assigned geographic domain.

Binary file I/O is not allowed. Each CPU has to read only the chunk of the file corresponding to the part of the "world" it has been assigned.

COARDS-compliant netCDF

- **C**ooperative **O**cean/**A**tmosphere **R**esearch **D**ata **S**ervice (COARDS) conventions faciliate data interchange
  - `ferret.wrc.noaa.gov/noaa_coop/coop_cdf_profile.html`
  - COARDS defines standards for naming dimensions, variables, etc.
  - COARDS defines standards for the metadata

- Also see our wiki page for more information
  - `wiki.geos-chem.org/Preparing_data_files_for_use_with_HEMCO`

Components of a COARDS-netCDF file

- Dimensions
  - Specify the # of grid boxes along each axis (X, Y, Z, T)
- Coordinate vectors
  - Define the values along each axis (X, Y, Z, T)
- Variables
  - Store emissions, concentrations, prod/loss, or other data
- Attributes
  - Contain information about:
    - A particular data variable (aka “variable attributes”)
    - The file itself (aka “global attributes”)

COARDS-netCDF dimensions

Dimension	Description
`time`	Specifies the number of points along the T axis.
`lev`	Specifies the number of points along the Z axis.
`lat`	Specifies the number of points along the Y axis.
`lon`	Specifies the number of points along the X axis.

- `time` must always be specified
- `time` may be set to “UNLIMITED” (allows you to append more time points after the file is created)
- `lev` may be omitted if the file contains data on a single level
- Recommended ordering: `time, (lev), lat, lon`

COARDS-netCDF coordinate vectors

Dimension	Description
`time`	Specifies the timestamps along the T axis.
`lev`	Specifies the vertical level along the Z axis.
`lat`	Specifies the latitudes along the Y axis.
`lon`	Specifies the longitudes along the X axis.

- COARDS coordinate vectors **must** have the same name as the dimension used to define them
- Axis values must be either monotonoically increasing (preferred) or monotonically decreasing
  - e.g. longitude must be either -180 .. 180 or 0 .. 360, depending on where the starting point of the grid is.

COARDS-netCDF coordinate vector: time

```
dimensions:
        time = UNLIMITED ; // (12 currently)
variables:
        float time(time) ;
                time:long_name = "time" ;
                time:units = "hours since 1985-01-01 00:00:00" ;
                time:calendar = "standard" ;
                time:axis = "T";
```

Output from "ncdump -c"

- Attributes (required, recommended)
  - `long_name`: Use "time" or "Time"
  - `units`: Use days, hours, or minutes since a reference time
    - Recommended: set reference time to 1st timestamp in the file
  - `calendar`: Use "standard" or "gregorian".
  - `axis`: Use "T". (Plotting programs may look for this.)

COARDS-netCDF coordinate vector: lev

```
dimensions:
        lev = 72 ;
 variables:
        int lev(lev) ;
                lev:long_name = "GEOS-Chem levels" ;
                lev:units = "level" ;
                lev:positive = "up" ;
                lev:axis = "Z" ;
```

Output from "ncdump -c"

- Attributes (required, recommended)
  - `long_name`: For GEOS-Chem data, use `GEOS-Chem levels`
    - This allows HEMCO to vertically regrid data between grids
  - `units`: Use `level`, `sigma_level`, etc.
  - `positive`: Use "up" (most G-C data goes up from the surface)
  - `axis`: Use "Z" (Plotting programs may look for this.)

COARDS-netCDF coordinate vector: lat

```
dimensions:
        lat = 181 ;
variables:
        float lat(lat) ;
                lat:long_name = "Latitude" ;
                lat:units = "degrees_north" ;
                lat:axis = "Y" ;
```

Output from “ncdump -c”

- Attributes (required, recommended)
  - `long_name`: Use “latitude” or “Latitude”
  - `units`: Use “degrees_north”
  - `axis`: Use “Y”. (Plotting programs may look for this.)

COARDS-netCDF coordinate vector: lon

```
dimensions:
        lon = 360 ;
variables:
        float lon(lon) ;
                lon:long_name = "Longitude" ;
                lon:units = "degrees_east" ;
                lon:axis = "X" ;
```

Output from “ncdump -c”

- Attributes (required, recommended)
  - `long_name`: Use “longitude” or “Longitude”
  - `units`: Use “degrees_east”.
  - `axis`: Use “X” (plotting programs may look for this.)

COARDS-netCDF data arrays

```
dimensions:
        time = UNLIMITED ; // (12 currently)
        lev = 72 ;
        lat = 181 ;
        lon = 360 ;
float CO(time, lev, lat, lon) ;
                CO:long_name = "CO" ;
                CO:units = "kg/m2/s" ;
                CO:add_offset = 0.f ;
                CO:scale_factor = 1.f ;
                CO:_FillValue = 1.e+15f ;
                CO:missing_value = 1.e+15f ;
                CO:gamap_category = "ANTHSRCE" ;
```

Output from "ncdump -c"

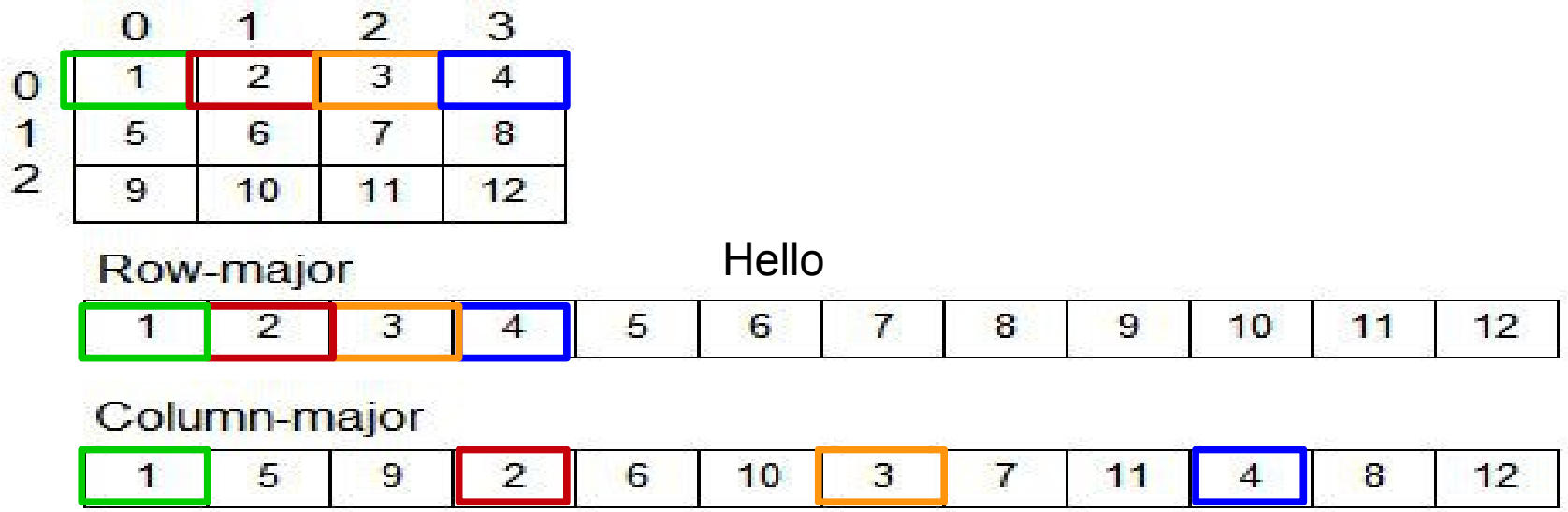
- Attributes (required, recommended)
  - `long_name`: Specify the name of the data
  - `units`: Use "kg/m2/s" for emissions; "1" for unitless, etc.
  - `add_offset, scale_factor`: Use 0 and 1 respectively
  - `missing_value, _FillValue`: Denotes missing data
  - `gamap_category`: Add the GAMAP category name (if applicable)

COARDS data arrays: data ordering

- If you use the **`ncdump -c`** command on a netCDF file, you will get output similar to what I showed in the previous slides.
- You may have noticed that the order of the data array from the previous slide displayed as **time, lev, lat, lon**.
- This is opposite to how GEOS-Chem stores data.
- So what's going on here???

COARDS data arrays: data ordering

- NetCDF (and ncdump) use in **row-major order**
  - Rows get stored first in memory, then columns
- GEOS-Chem (and Fortran) use **column-major order**
  - Columns get stored first in memory, then rows
- The reverse ordering is used to access the same data!!!



Creating COARDS netCDF files

- GAMAP routine BPCH2COARDS can convert a binary punch file to COARDS-netCDF format
- You can further refine the file by manually editing the variable names and attributes with:
  - NetCDF Operators (NCO)
  - Climate Data Operators (CDO)
- See our wiki page for a tutorial:
  - `wiki.geos-chem.org/Preparing_data_files_for_use_with_HEMCO`

NetCDF software

- For viewing netCDF files:
  - ncview
  - panoply
  - ncdump (useful for printing file headers)
- For manipulating netCDF files:
  - NetCDF Operators (NCO)
  - Climate Data Operators (CDO)
  - NCAR Command Language (NCL)
  - Compress netCDF files with nccopy -dN (where N=1-9)
- See our wiki page for a tutorial:
  - `wiki.geos-chem.org/Preparing_data_files_for_use_with_HEMCO`

Extra Slides

Flexible precision in GEOS-Chem

We have added a new module (`Headers/precision_mod.F`) that defines a parameter named `fp`.

We now declare most real variables in GEOS-Chem with the `fp` parameter, such as:

```
USE PRECISION_MOD
REAL(fp)               :: A
REAL(fp), ALLOCATABLE  :: ARRAY(:,:)
REAL(fp), POINTER      :: PTR(:,:)
```

etc.

If you compile GEOS-Chem with the optional `PRECISION` variable, such as:

```
make -j4 MET=geosfp GRID=4x5 PRECISION=4 ...
```

Then the fp parameter will be set to `REAL*4` precision:

```
! Use 4-byte floating point precision
INTEGER, PARAMETER, PUBLIC :: fp = KIND( 0.0 )
```

And all of the variables that we defined above will be declared using 4-byte precision.

If you omit `PRECISION` from the compilation command, then by default all variables will be declared using 8-byte precision (which is the status quo).

Flexible precision in GEOS-Chem

- Why do we need flexible precision?
  - We can halve the amount of memory needed for data arrays
- Using 4-byte precision is **much** more efficient for HPC
  - Parallelization in an HPC environment is much faster
  - NASA GEOS-5 GCM can now run at 7km horizontal resolution
    - So we need to reduce the memory footprint as much as possible
- 4-byte precision would also benefit nested-grid sims
  - Currently, we can only use the ¼ nested grids in the “large memory” queue
  - Only a 1 or 2 nested grid sims can currently run simultaneously