# GEOS-CHEM documentation: Emory - Liu group

J.H. Belle

---

We use GEOS-CHEM to obtain additional information on chemical composition, optical properties, and vertical distributions of aerosols in the atmosphere that is not available from most satellites or ground monitors. This information is most frequently used to 'scale' vertically integrated AOD values obtained from satellites to ground-measured $PM_{2.5}$. *This guide covers how to download additional data, run simulations and interpret and use results from GEOS-CHEM 10.1 on our cluster.*

**What is GEOS-CHEM**
GEOS-CHEM is "a global three-dimensional model of tropospheric chemistry driven by assimilated meteorological observations from the Goddard Earth Observing System (GEOS) of the NASA Global Modeling Assimilation Office (GMAO)" - `http://acmg.seas.harvard.edu/geos/doc/man/`. This means that it primarily depends on meteorological observations, or 'met fields' supplied by GMAO. The 10.1 version set up on our cluster is currently capable of using meteorological data from the following products: GEOS-5 and GEOS-FP. Each product has a different spatial resolution and temporal coverage period, which restricts the duration and resolution of simulations run using each of the meteorological products. I.e. A simulation run using GEOS-5 data has a maximum spatial resolution of 0.5x0.666 degrees and can only be run through June of 2013 because this is the finest resolution of the meteorological data and GEOS-5 only has temporal coverage from 2004-2013. A simulation run using GEOS-FP data has a maximum spatial resolution of 0.25x0.3125 degrees and can cover the years from 2013-present. (Additional information: `http://wiki.seas.harvard.edu/geos-chem/index.php/Overview_of_GMAO_met_data_products`)

We will assume that the the source code for GEOS-CHEM 10.1 and netcdf libraries has been downloaded, installed, and pre-compiled, and additionally that the run and shared data directories have been set up appropriately. Please note that a compiled version of GEOS-CHEM only works for a specific grid (effectively the spatial resolution of the output, must be one of: 4x5, 2x2.5, 0.5x0.666 or 0.25x0.3125), chemical mechanism, and meteorological product (one of: GEOS-4, GEOS-5, GEOS-FP, MERRA, or MERRA2) combination.

**NOTE:** *If you need to run a simulation with a different grid and/or meteorological product from the one initially set up for you, see J.H. Belle (jessica.hartmann.belle@emory.edu). Do not attempt to recompile source code, set up run directories yourself, or share a run directory with another person, as there is a possibility that the resulting runs could overwrite someone else's output, we may run out of space to store files, or that multiple simultaneous accesses will result in corrupted input or output files. However, once a run directory with a geos executable included in it has been set up for you, you can make as many copies as necessary if you change the folder name each time. Additionally note that our version of the source code currently differs substantially from that in the main git repository, and the original v10-01 source code will not compile and run correctly on our system. Details on the alterations made are listed in section 10.1 of this document.*

---

**Table of Contents**

# 1 Setting up your environment

GEOS-CHEM is a large, computational and memory-intensive program to run and requires libraries located in a place on the hard drives for the cluster that is not, by default, included in list of places to look for programs a user needs to run. This means that unless you make changes to your environment, or the default values for certain variables, and the resources available, GEOS-CHEM will not be able to run, either because it is missing libraries, or because it has run out of resources. If your default shell for the cluster is Bash, these options can be set in a file, located in your home directory (`/home/[username]`), called '.profile' that can then be sourced from within your job script. This means that any commands placed in a .profile file will be run as soon as you start a job. To obtain this file, either copy it from Jess or create a file in your home directory with the appropriate options set.

To copy Jess's .profile to your own home directory, enter the following two commands into your terminal. These will return you to your home directory if you have left it, and then copy the file over:

```
cd
cp /home/jhbelle/.profile .profile
```

The resulting file should include the following code:

```
# .profile

# User specific aliases and functions
ROOT_LIBRARY_DIR=/liu_group/remotesensing1/Jess/GEOS_CHEM_10_1/geos-chem-libraries/opt/ifort/nc4
GC_BIN=$ROOT_LIBRARY_DIR/bin
GC_INCLUDE=$ROOT_LIBRARY_DIR/include
GC_LIB=$ROOT_LIBRARY_DIR/lib
export GC_BIN
export GC_INCLUDE
export GC_LIB

# Max out machine limits
ulimit -s unlimited              # stacksize
ulimit -c unlimited              # coredumpsize
ulimit -m unlimited              # memoryuse
ulimit -v unlimited              # vmemoryuse

# Increase a specific stack size value to an arbitrarily large number
export KMP_STACKSIZE=500000000

# Source the ifort variable file - without this the geos executable won't run
source /usr/local/intel/composer_xe_2015.3.187/bin/ifortvars.sh intel64

# Set default number of threads for OpenMP - GEOS-CHEM runs most efficiently over 16 threads,
but 12 is the limit on our cluster
export OMP_NUM_THREADS=12
```

The first set of options export the variables `GC_BIN`, `GC_INCLUDE`, and `GC_LIB` to your environment. These variables need to be available so that GEOS-CHEM can find and use the netcdf libraries it needs to read and write inputs and outputs. The second and third sets of options set larger than default limits for different types of memory use. Setting these limits higher will prevent GEOS-CHEM from crashing because it believes it has run out of space to store information in as it runs. The fourth option sets the number of threads GEOS-CHEM will use in parallel. We have found that GEOS-CHEM runs the fastest when parallelized over 16 threads, and speed increases are not typically seen above this number, but we use 12 on our cluster because this provides more efficient resource use when we have multiple simulations running at a time.

## 2 The `ExtData` directory

Static data inputs, grid definitions, emissions inventories, and Met field files are stored in the `ExtData` directory located at `/liu_group/remotesensing3/GEOS_CHEM_10_1_MetFields/` on the cluster. This directory has already been set up with the structure outlined below and should not be altered, as any changes to the structure of this directory will typically cause GEOS-CHEM to crash. In a few cases, symbolic links have been used to obtain this structure and make downloads from the ftp directories slightly more straightforward. Note that the online documentation on ExtData suggests that the GEOS_NATIVE directory should be nested within CHEM_INPUTS, but the system is actually expecting to be able to access GEOS_NATIVE from ExtData, not from ExtData/CHEM_INPUTS.

- ExtData
  - CHEM_INPUTS
  - GEOS_0.25x0.3125_CH
    - GEOS_FP
      - 2012
      - 2013
      - 2014
      - 2015
  - GEOS_0.25x0.3125_NA
  - GEOS_0.5x0.666_CH
  - GEOS_0.5x0.666_NA
  - GEOS_2x2.5
  - GEOS_4x4.5
  - GEOS_NATIVE
    - FastJ_201204
    - Linoz_200910
    - MODIS_LAI_201204
    - Olson_Land_Map_201203
    - TOMAS_201402
    - UCX_201403
  - HEMCO

The actual structure of our `ExtData` directory, ignoring the symbolic links, is outlined in below (this is what you see when you view `ExtData` from WinSCP). Our actual directory structure closely resembles that of the dalhousie ftp directory (discussed below), which stores met fields in folders located directly under `ExtData` and labeled: `GEOS_[GRID]_[NEST].d` or `GEOS_[GRID].d`.

- ExtData
  - CHEM_INPUTS
    - HEMCO
    - FastJ_201204
    - Linoz_200910
    - MODIS_LAI_201204
    - modis_surf_201210
    - Olson_Land_Map_201203

        * RRTMG_201411
        * TOMAS_201402
        * UCX_201403
    - GEOS_0.25x0.3125_CH
    - GEOS_0.25x0.3125_CH.d
        * GEOS_FP
            · 2012
            · 2013
            · 2014
            · 2015
    - GEOS_0.25x0.3125_NA
    - GEOS_0.25x0.3125_NA.d
    - GEOS_0.5x0.666_CH
    - GEOS_0.5x0.666_CH.d
    - GEOS_0.5x0.666_NA
    - GEOS_0.5x0.666_NA.d
    - GEOS_2x2.5
    - GEOS_2x2.5.d
    - GEOS_4x4.5
    - GEOS_4x4.5.d

## 2.1 Copying ExtData to a new location

To copy ExtData to a new location or drive, you need to make the copy and then set up all the symbolic links within the directory. The script below, run from /aqua, was used to copy ExtData on 08/11/2016 from /liu_group/climatechange2 to /aqua and took a few hours to run.

```
!#/bin/bash

## This script copies ExtData from the old location to a new one
## It also reestablishes the symbolic linkages that would have been lost in the transfer
## NOTE: Didn't set up linkages for GEOS_5 and 2x2.5 or 4x5 or GEOS_FP for 4x5 models
since we are done with or not using these

# Copy directory over
cp -r /liu_group/climatechange2/ExtData .
# Reestablish symbolic linkages within the directory
# For each broken link need to delete old broken one and make a new one
# Linkage from ExtData/CHEM_INPUTS/HEMCO to ExtData/HEMCO
cd ExtData
rm HEMCO
ln -s CHEM_INPUTS/HEMCO HEMCO
# Set up links within GEOS_NATIVE back to the actual folder locations within CHEM_INPUTS
cd GEOS_NATIVE
rm *
ln -s ../CHEM_INPUTS/FastJ_201204 FastJ_201204
ln -s ../CHEM_INPUTS/Linoz_200910 Linoz_200910
```

```
ln -s ../CHEM_INPUTS/MODIS_LAI_201204 MODIS_LAI_201204
ln -s ../CHEM_INPUTS/Olson_Land_Map_201203
ln -s ../CHEM_INPUTS/TOMAS_201402 TOMAS_201402
ln -s ../CHEM_INPUTS/UCX_201403 UCX_201403
# Linkage from ExtData/GEOS_0.5x0.666_NA.d/GEOS_5 to ExtData/GEOS_0.5x0.666_NA/GEOS_5
cd ../GEOS_0.5x0.666_NA
rm GEOS_5
ln -s ../GEOS_0.5x0.666_NA.d/GEOS_5 GEOS_5
# Linkage from ExtData/GEOS_0.5x0.666_CH.d/GEOS_5 to ExtData/GEOS_0.5x0.666_NA/GEOS_5
cd ../GEOS_0.5x0.666_CH
rm GEOS_5
ln -s ../GEOS_0.5x0.666_CH.d/GEOS_5 GEOS_5
# Linkage from ExtData/GEOS_0.25x0.3125_NA.d/GEOS_FP to ExtData/GEOS_0.25x0.3125_NA/GEOS_FP
cd ../GEOS_0.25x0.3125_NA
rm GEOS_FP
ln -s ../GEOS_0.25x0.3125_NA.d/GEOS_FP GEOS_FP
# Linkage from ExtData/GEOS_0.25x0.3125_CH.d/GEOS_FP to ExtData/GEOS_0.25x0.3125_CH/GEOS_FP
cd ../GEOS_0.25x0.3125_CH
rm GEOS_FP
ln -s ../GEOS_0.25x0.3125_CH.d/GEOS_FP GEOS_FP
# Linkage from ExtData/GEOS_2x2.5.d/GEOS_FP to ExtData/GEOS_2x2.5/GEOS_FP
cd ../GEOS_2x2.5
rm GEOS_FP
ln -s ../GEOS_2x2.5.d/GEOS_FP GEOS_FP
```

# 3 Downloading new met fields

Met fields can be downloaded from the Harvard (`ftp://ftp.as.harvard.edu/gcgrid/`) and/or Dalhousie (`ftp://rain.ucis.dal.ca/ctm/`) ftp directories. Met fields may or may not be formatted as compressed files (end in .nc.gz instead of .nc), but will typically not be compressed. Compressed files can be extracted with gunzip from the bash command line (Ex. `gunzip -r /liu_group/remotesensing3/GEOS_CHEM_10_01_MetFields/ExtData/`), or you can ask GEOS-CHEM to extract on the fly in the input.geos file used to specify the simulation run, although the second option has not yet been tested on our system with 10.1. Directories will occasionally be taken down without warning, so check data availability using a web browser, by entering the urls above, before attempting a download. Also, met field files will occasionally be missing, corrupted, or improperly extracted on one or both of the ftp directories and will cause GEOS-CHEM to fail to read the file and crash during a simulation. The GEOS-CHEM help team is typically notified and the affected files are replaced quickly, but it is recommended that you regularly save restart files during a simulation. A check can be performed to identify and replace any files that have been edited or reuploaded in the ftp directories since the last download, by using the flag -N with wget.

You will need to download met field files from one of the two ftp directories to the appropriate place in the `ExtData` directory located on our cluster at `/liu\_group/remotesensing3/GEOS\_CHEM\_10\_01\_MetFields/`. To download files from an ftp directory to the cluster, you will need to submit a bash script to the cluster submission engine. To do this, create a file, ending in .sh, similar to the one of the examples below, but with the ftp directory replaced with that of the specific files you require for your simulation.

*Example 1:* A script to download all met fields for july 2013 into the appropriate folder in ExtData. The flag -nd in the wget command specifies that all files are to be dumped into the current working directory, while the flag –tries=45 means that the system will make up to 45 attempts to retrieve a file if the connection is lost partway through. This file can be submitted to the cluster from any directory, using the command `qsub example1.sh`

```
#!/bin/bash
cd /liu_group/climatechange2/ExtData/GEOS_0.25x0.3125_NA.d/GEOS_FP/2013/07
wget -nd --tries=45 ftp://rain.ucis.dal.ca/ctm/GEOS_0.25x0.3125_NA.d/GEOS_FP/2013/07
```

*Example 2:* A script to download all met fields for the years 2003 and 2004 into the appropriate folders in ExtData. The flag `-r` in the wget command tells wget to search directories under `ftp://rain.ucis.dal.ca/ctm/GEOS_2x2.5.d/GEOS_5/2003` and `ftp://rain.ucis.dal.ca/ctm/GEOS_2x2.5.d/GEOS_5/2004` on the ftp site recursively for files to download. The flag `-nH` tells wget not to include the ftp address, `rain.ucis.dal.ca`, as part of the directory structure on our system, while the –cut-dirs=1 flag tells wget to additionally cut `/ctm` out of the directory pathway on our system. Together, both options mean that, if this script is submitted from `ExtData`, wget will place files in the appropriate directories on our system, where `GEOS_2x2.5.d` is present in the `ExtData` directory. This script must be run from `/liu_group/climatechange2/ExtData`, and can be submitted from this location using the command `qsub -cwd Example2.sh` if your current working directory is set to `/liu_group/climatechange2/ExtData`.

```
#!/bin/bash
wget -r -nH --tries=45 --cut-dirs=1 "ftp://rain.ucis.dal.ca/ctm/GEOS_2x2.5.d/GEOS_5/200[34]"
```

### 3.1 Data storage requirements

The majority of the hard drive space taken up by GEOS-CHEM is done by the time-varying met fields. The amount of storage space required for these files varies based on the grid resolution, meteorological data product, chemistry mechanism (we usually use the tropospheric-only full chemistry mechanism), and type of output requested for the simulation. Table 1 outlines input and output storage requirements for uncompressed files.

| Resolution | Nested grid | Met. product | # Input files | Inputs storage size | # Output files | Output storage size | Total space |
|---|---|---|---|---|---|---|---|
| 4x5 | - | GEOS-FP | 6 | 90 Mb | 1 | - | - |
| 2x2.5 | - | GEOS-5 | 3 | 330 Mb | | | 1 Tb |
| | - | GEOS-FP | 6 | 350 Mb | 1 | 405 Mb | 400 Gb |
| 0.5 x 0.666 | NA | GEOS-5 | 3 | 465 Mb | 1 | 540 Mb | 1.7 Tb |
| | CH | GEOS-5 | 4 | 425 Mb | 1 | 375 Mb | 1.2 Tb |
| 0.25 x 0.3125 | NA | GEOS-FP | 6 | 565 Mb | | | |
| | CH | GEOS-FP | 6 | 560 Mb | | | |

Table 1: Storage requirements for met field inputs and GEOS-CHEM outputs. Input and output size estimates are for 1 day's worth of simulation. Output sizes are estimated assuming you output instantaneous estimates every 180 minutes (3 hours). 24-hour average output files are far smaller.

Since we need to work with GEOS-CHEM outputs on the cluster, we need to account for the space to store the full time series of outputs when planning runs. The global model results don't necessarily need to be stored on the cluster and can be archived to Big4s, since we would typically prefer the nested outputs. We have 5 Tb of space for outputs available on /gc_runs, and 4.2 Tb of space for inputs and boundary condition files for the nested runs on /liu_group/climatechange2. Because 4.2 Tb is not enough space to store all met field files for all runs simultaneously, you will need to periodically delete old met fields and download new ones as your simulation runs.

## 4 Setting up a simulation

To set up a simulation, you will need to prepare an input file in the run directory for the simulation you are running, typically called input.geos. This file describes the exact simulation you would like to run, as well as where the program can find all of the met fields and other inputs it needs to actually run the simulation. You will also need to prepare the emissions configuration file, called HEMCO_Config.rc, which describes the emissions inventories you can use in a simulation and allows you to turn each one on or off. A number of example input files for different types of runs are located in the Unit Testing directory in remotesensing1 at `/liu_group/remotesensing1/Jess/GEOS_CHEM_10_1/UT/runs`, in the main testing directory at `/liu_group/remotesensing1/Jess/GEOS_CHEM_10_1/RunDirTesting`, and in the production run directories located at `/liu_group/climatechange2/GCRunDirs` and `/gc_runs`.

## 4.1 Input.geos

Each menu of the input file contains slightly different information, and each is covered in detail here: http://acmg.seas.harvard.edu/geos/doc/man/. We typically run a "Full Chemistry" simulation, or a "NOx-Ox-Hydrocarbon-aerosol" simulation, with UCX (stratospheric chemistry), RRTMG, and SOAs turned off. The full chemistry simulation we typically run is nearly identical to the default, but there a few things you may need to alter in the original input.geos file found in your run directory.

In the simulation menu, set the start and end dates for the simulation and check the file paths. To run a simulation from 07/01/2013 to 8/01/2013, you would need to alter the first two lines in the menu to read:

```
Start YYYYMMDD, HHMMSS  : 20130701 000000
End   YYYYMMDD, HHMMSS  : 20130801 000000
```

In this menu, also check to make sure that Make new restart file? is set to True (T), and that the Global offsets I0, J0 are appropriate for your simulation. For global simulations these should be set to 0 0, but for nested simulations (resolution higher than 2x2.5), you will need to use the values specified here: http://wiki.seas.harvard.edu/geos-chem/index.php/Setting_up_GEOS-Chem_nested_grid_simulations

In the tracer menu, the Type of simulation should be set to 3, and the Number of Tracers should be set to 66 for the standard tropospheric chemistry simulation that we typically run. We additionally use molecular weights for NITs (Nitrates in sea salt) and SO4s (Sulfates in sea salt) that correspond to salts instead of using the weights for NIT or SO4 alone. Since this change wasn't officially implemented until GEOS-CHEM version 11, the default input file in your run directory is unlikely to have been updated with the new weights. The molecular weights for tracers 28 and 33 should be set to 31.4 g/mol.

We usually export instantaneous values of tracer concentrations and optical properties, using the ND49 diagnostic. Your ND49 diagnostic menu should look like the below for global simulations, you can export additional tracers, but at a minimum, include those ;isted below in any output you save out from a production run. If you are running a nested simulation you would alter the I and J region definition to correspond to your simulation grid.

```
------------------------+----------------------------------------------------------
%%% ND49 MENU %%%       :
Turn on ND49 diagnostic : T
Inst 3-D timeser. file  : tsYYYYMMDD.bpch
Tracers to include      : 27 28 31 32 33 34 35 36 37 38 39 40 41 42 43
153 155 157 158 159 160 161 166 167 168 169 170 171 172
Frequency [min]         : 180
IMIN, IMAX of region    :  1  144
JMIN, JMAX of region    :  1  91
LMIN, LMAX of region    :  1  37
------------------------------------------------------------------------------------
```

## 4.2 HEMCO_Config.rc

The HEMCO configuration file is where emissions inventories for the simulation are specified. You will need to make a few changes to your HEMCO_Config.rc files before starting a simulation for the first time, and when you restart a simulation for the first time.

## 4.3 Restart files

Each time start or restart a simulation you will need to use a set of restart files, which contain starting values for the simulation. To start a simulation for the first time, you only need 1 restart file containing initial array values for the simulation you plan to run on the first day of the first month of the simulation, although the year can be different from the year in which you plan to start the simulation. Official restart files obtained from the Unit Tester are typically named initial_trac_rst.[metdatasource]_[gridsize]_[chemistrymechanism], while restart files generated on our system will typically be titled trac_rst.[metdatasource]_[gridsize]_[nestedgridregion]_[chemistrymechanism].YYYYMMDDHHMM. Most of the restart files obtained from the unit tester are for July 1st, but the timestamps can be checked in gamap. When starting a simulation for the first time, you will also need to set HEMCO_RESTART to false on line 168 of your HEMCO_Config.rc file in your run directory. This will force the HEMCO emissions component to initialize itself from scratch, using default values,

instead of trying to draw them from a restart file which doesn't exist. HEMCO restart files must be for the same year, month, and day as the simulation start.

To restart a simulation that's already been running, you will need 3 separate restart files.

- `trac_rst.[metdatasource]_[gridsize]_[nestedgridregion]_[chemistrymechanism].YYYYMMDDHHMM` - tracer concentrations. GEOS-CHEM cannot start without this file.
- `HEMCO_restart.YYYYMMDDHHMM` - initial values for HEMCO emission component
- `spec_rst.[metdatasource]_[gridsize]_[nestedgridregion]_[chemistrymechanism].YYYYMMDDHH` - initial values for chemical species used in by the chemistry solver. Without this file the chemistry will reinitialize to default values on restart, negating the previous run and possibly introducing 0 concentration errors into the chemistry.

When restarting a previously run simulation, you will also need to switch the HEMCO_RESTART flag to true in line 168 of HEMCO_Config.rc. Use of the species restart file should already be turned on in input.geos, as its the same flag that tells GEOS-CHEM to output the species restart files.

## 4.4 Regridding restart files for nested simulations

## 4.5 Non-standard simulation runs

If you need to run non-standard chemistry simulations you will need to request a fresh compiliation of the source code and run directory. In the case of the RRTMG and GEOS-CHEM HP models, additional testing will likely be required prior to setting up production runs. Some of the non-standard runs you may need to use at some point for our work:

- SOAs - Secondary organic aerosols, and additionally, secondary volatile organic compounds (SVOCs), can be added to the model as tracers if desired. This simulation takes slightly longer to run and requires fresh run directories and compilations of the source code. However, we've had some issues with the output from this model, where concentrations of organic and black carbon decrease in the boundary layer relative to the simulation that doesn't include the additional tracers.
- RRTMG - Radiative transfer model used to estimate radiances for either solar or climate change applications (radiative forcings). This is an add-on to GEOS-CHEM that must be coupled to the main model. It takes about double the resources and time as the main model to run a simulation with RRTMG on.
- Offline Aerosols - A lighter-weight (less resource intensive) version of the GEOS-CHEM model that only models aerosol concentrations. Requires output from a previous GEOS-CHEM full chemistry run, typically the 8.03 version.
- RCPs/Future scenarios - GEOS-Chem is capable of future projections, but we would typically use ensemble output from a standard model.
- GEOS-CHEM HP - A version of GEOS-CHEM designed to work with HPC systems that allow for parallelization across multiple nodes on a cluster. This is still in beta testing at Harvard and probably won't be included in an official GEOS-CHEM release for some time, but interfaces with ESMF and is less grid-dependent in how it runs than the standard model.
- HEMCO only - Run Geos-Chem with emissions and aerosols turned on, but transport, convection, and chemistry off to obtain emissions information as output on the Geos-Chem grids.
- Using GEOS-CHEM with MERRA, MERRA-2, GEOS-4, GEOS-3, GEOS-STRAT, or GEOS-1 Met fields - MERRA-2 is the only one we might realistically need to use, and would require recompilation of source code, fresh run directories, and some changes to `ExtData` to run. Additionally, only a small fraction of the met fields have bene posted on the ftp directories. GEOS-1, GEOS-3 and GEOS-STRAT met fields are not compatible with GEOS-CHEM 10.1 and cover time periods prior to the launch of most of the satellites we use.

# 5 Submitting a simulation run to the cluster

## 5.1 Estimating run times and memory usage

## 5.2 Creating a shell script

## 5.3 Submitting a run to UGE

# 6 Monitoring your simulation

The UGE system produces an output and an error file for each submission. However, output files are always empty, and error messages reported in the error files from UGE are taken from ifort, and usually not GEOS-CHEM directly. So, all the UGE error file will tell you is whether or not GEOS-CHEM crashed. In the event that GEOS-CHEM has crashed your UGE error file with either report that GEOS-CHEM exited with a clean error, or that the system hit a segmentation fault. Clean errors are usually reported as: `Code stopped from Do_Err_Out.`, and typically result from issues with the input files such as erroneous file paths, missing restart or boundary condition files, or failing to schedule output on the last day of the simulation. All of these are issues that GEOS-CHEM checks for prior to starting the simulation.

A segmentation fault error looks like the below, and results from the simulation crashing partway through.

```
forrtl: severe (174): SIGSEGV, segmentation fault occurred
Image              PC                Routine            Line        Source
libintlc.so.5      00002B1166445961  Unknown               Unknown  Unknown
libintlc.so.5      00002B11664440B7  Unknown               Unknown  Unknown
libifcoremt.so.5   00002B1164F9BF52  Unknown               Unknown  Unknown
libifcoremt.so.5   00002B1164F9BDA6  Unknown               Unknown  Unknown
libifcoremt.so.5   00002B1164F058D6  Unknown               Unknown  Unknown
libifcoremt.so.5   00002B1164F16CA8  Unknown               Unknown  Unknown
libpthread.so.0    00000035CAC0F7E0  Unknown               Unknown  Unknown
geos               000000000048EC36  Unknown               Unknown  Unknown
geos               0000000000AF3846  Unknown               Unknown  Unknown
geos               000000000063071E  Unknown               Unknown  Unknown
geos               000000000066B0A1  Unknown               Unknown  Unknown
geos               0000000000411EFE  Unknown               Unknown  Unknown
libc.so.6          00000035CA41ED5D  Unknown               Unknown  Unknown
geos               0000000000411E09  Unknown               Unknown  Unknown
```

We've compiled all geos executables with debugging and traceback turned off, so the lines, modules/functions, and source files in the original fortan code that were in use when the system crashed will be reported as 'Unknown'. This was done because leaving debugging on slows down the simulations and unless there are bugs in the source code that need to be found the additional output isn't needed.

However, GEOS-CHEM also writes a number of log files that can be used to determine the actual error that caused the crash. The last line of `log` will typically report the specific error causing the simulation to stop and will give more information about what happened and when. Many are self-explanatory or include instructions, but additionally refer to the wiki for common error messages and known solutions. The correct section is usually easiest to find by copying the error message into google search and adding geos-chem to the search query. In the event that your error is not included in the wiki, ask someone else in the group or email the GEOS-CHEM support team about it.

Even if your simulation runs through without crashing, the log files contain useful information and any non-fatal warnings that could have affected the output will be reported in these. GEOS-CHEM produces 3 log files, a log, a HEMCO.log, and an smv2.log.

## 6.1 log

This is the main log file for GEOS-CHEM itself. Some example log errors and their solutions are below:

```
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 !! In Ncop_Rd, cannot open:  /aqua/ExtData/GEOS_2x2.5/GEOS_FP/2015/12/GEOSFP.20151212.A1.2x25.nc
 !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
```

This message indicates that the meteorological field, `/aqua/ExtData/GEOS_2x2.5/GEOS_FP/2015/12/GEOSFP.20151212.A1.2x25.nc`, could not be opened. Check the file to ensure that it is of an appropriate size and compare the origin dates between our directory and the ftp directory it was downloaded from. If they differ, replace the file. If they do not differ, check a copy of the file from the remote directory in hdfview and if it cannot be opened, notify the Geos-Chem support team that the file needs to be replaced.

### 6.2 HEMCO.log

This is the log file produced by the HEMCO emissions component.

### 6.3 smv2.log

This is the log file produced by the SMVGEAR chemistry solver.

## 7 Using GAMAP

GAMAP is a set of IDL routines that are designed to work with GEOS-CHEM outputs and perform operations specific to GEOS-CHEM, such as extracting data or regridding and clipping restart files.

### 7.1 Setting up GAMAP

To set up GAMAP, you may find it easier to follow the instructions located here: acmg.seas.harvard.edu/gamap/doc/. They will ask you to download the source code and instruct you in how to set up your home directory so you can use the additional IDL routines included with it. These routines must be accessed from within IDL.

### 7.2 Converting .bpch files to other formats

### 7.3 Visualizing and extracting data

### 7.4 Comparing outputs

A difference plot for 24hour average SO4 concentrations on 07/30/2013 can be created with the command: `CTM_PLOTDIFF, 'IJ-AVG-$','filename1','filename2','TRACER=27,TAU0=250488.50` This command produces a 4-panel plot where the 1st two panels have tracer concentration maps from each of the files, the 3rd panel shows the absolute differences between the two, and the 4th, the percentage differences between the two. Note that we have had problems matching TAU0 values for 24h average diagnostic outputs (ND50) from nested model runs.

## 8 Tidying up - Storing outputs and getting rid of old met fields

After your simulation has completed, you will need to to delete the met fields you downloaded for the simulation and clean and store the outputs it generated. The data storage estimates in section 3.1 are based off of ND49 outputs only, and do not include the storage requirements for the trac_avg or restart files, which can double storage requirements. Since the trac_avg files are rarely used and are primarily archived for diagnostic purposes, it is recommended that you periodically move trac_avg files generated by your simulations down from the cluster to one of the Big4s and delete the originals from the cluster.

You will also need to periodically delete met field files for years you have already simulated and download missing years. We have space to store a few years worth of met field files for each simulation, so for longer simulations you will likely need to work in batches, moving forward a few years at a time, deleting the old met fields, downloading new ones, and restarting the simulation runs from the point you left off.

# 9 Example use case for output

# 10 Additional resources

The community of users and developers for the GEOS-CHEM program is fairly large and active, so there are plenty of other places you can turn to for additional information and more extensive documentation (this documentation was also written by people who know a lot more than I do).

This community maintains:

- A wiki: `http://wiki.seas.harvard.edu/geos-chem/index.php/Main_Page`
- A user-guide: `http://acmg.seas.harvard.edu/geos/doc/man/`
- A narrative description of the model: `http://acmg.seas.harvard.edu/geos/geos_chem_narrative.html`
- A support team - you can email them with questions not answered on the wiki or by going through the user-guide: `geos-chem-support@as.harvard.edu`
- Ftp servers with met fields, emissions, and input data: `ftp://ftp.as.harvard.edu/gcgrid/data/` or `ftp://rain.ucis.dal.ca/ctm/`. These directories are described at: `http://wiki.seas.harvard.edu/geos-chem/index.php/Downloading_GEOS-Chem_source_code_and_data#Shared_data_directory_archives`
- Source code for the GEOS-CHEM program: `https://bitbucket.org/gcst/geos-chem`

## 10.1 Differences between our source code and that in the main repository

As part of getting GEOS-CHEM v10-01 running on our system and producing consistent and mostly correct output we had to make a number of changes to our source code that were not scheduled to be included in the main code base until v11 was released. Code was versioned along the way and all changes should be trackable through git as well, using the command `git diff` and the commit ids listed below.

### 10.1.1 Updated molecular weights for SALA, SALC, NITs and SO4s to the newer weight for sea salt (from 36 g/mole to 31.4 g/mole)

While molecular weights are specified in the input.geos files in the simulation run directories, they are also specified within the source code for the dry deposition module in GeosCore/drydep_mod.F. Previously, the source code weights for sea salt (SALA and SALC) and NIT and SO4 in sea salt (NITs and SO4s) had been set to 36 g/mole, while the input.geos files used a molecular weight for these compounds of 31.4 g/mole for SALA and SALC, 96 g/mole for SO4s, and 62 g/mol for NITs. We updated the source code in the dry deposition module to reflect the change in the molecular weight of seasalt from 36 g/mole to 31.4 g/mole, and have been continually updating our input.geos files to use 31.4 g/mole for NITs and SO4s. To fix the dry deposition source code, the lines reading `XMW(NUMDEP) = 36e-3_f8` at lines 4498, 4556, 5075, and 5089 in GeosCore/drydep_mod.F, were altered to read `XMW(NUMDEP) = 314e-4_f8`. This change was committed to the git repository on 05/06/2016, with the commit ID: bb4dfd790557a66c6efe2e6c336786ad91c794b9.

### 10.1.2 Fixed uninitialized variable error from ifort

Two obsolete variables, TIME and TSPMIDC, weren't initialized properly in GeosCore/physproc.F and caused errors during compilation. Our version of ifort catches this as an error even though the affected variables are not used later on in the code, so we had to fix it in order to use ifort's debugging capabilities for other issues. The line `TIME = 0.0_fp` was added into GeosCore/chemistry_mod.F, at line 1152. This is a shortcut to initialize the variable TIME in physproc.F Within physproc.F,

we deleted the line `TSPMIDC = MOD(TSPMIDC + CHEMINTV, SCDAY)`. This change was committed to the git repository on 05/06/2016, with the commit ID: bb4dfd790557a66c6efe2e6c336786ad91c794b9.

### 10.1.3 Fixed bug in FINN

There was some improper indexing in the FINN extension module from HEMCO that caused the scaling factors for OC and BC in FINN to be applied incorrectly and an array out of bounds error to occur. This was fixed by replacing our copy of HEMCO/Extensions/hcox_finn_mod.F90 with a newer version provided by Christoph Keller at NASA, who was part of the team that developed HEMCO. Unfortunately, fully implementing this fix would have required significant additional time spent debugging and developing on our part and multiple additional changes to the original codebase, so we ultimately decided to switch to GFED biomass burning emissions. The codebase was reverted to that of bb4dfd790557a66c6efe2e6c336786ad91c794b9, on 05/09/2016 and the alterations were never committed.

### 10.1.4 Fixed bug causing high tracer concentrations in wet deposition in models using GEOS-FP or MERRA met fields

A constraint on the variable alpha to be less than 1 in the cloud convection module was missing in the section of code that handles aerosol tracer washout. The constraint code used to calculate washout of other tracers in different sections of the same code was added to line 2640 in GeosCore/convection_mod.F

```
! Restrict ALPHA to be less than 1
! (>1 is unphysical)  (jhb, 10-May-2016) This restriction was missing in this section,
! but present in others
IF ( ALPHA > 1e+0_fp ) THEN
    ALPHA = 1e+0_fp
ENDIF
```

This change was committed to git on 05/11/2016 with commit ID: .

### 10.1.5 Fixed equation for amount of aerosols lost to washout when using GEOS-FP or MERRA met fields

The amount of aerosols lost to washout in a grid box had been calculated as, at line 2656 or GeosCore/convection_mod.F:

```
                    ! Amount of aerosol lost to washout in grid box
                    WETLOSS = Q(K,IC) * BMASS(K) / TCVV *
    &                         WASHFRAC - GAINED
```

The equation was altered to be consistent with that used for other cases within GeosCore/convection_mod.F. The new equation, at line 2656, reads:

```
                    ! Amount of aerosol lost to washout in grid box
                    WETLOSS = ( Q(K,IC) * BMASS(K) / TCVV + GAINED ) *
    &                         WASHFRAC - GAINED
```

This change was committed to git on 05/12/2016 with commit ID: .

### 10.1.6 Fixed bug affecting mass conservation during cloud convection in models using GEOS-FP or MERRA met fields

Also within GeosCore/convection_mod.F, in the DO_MERRA_CONVECTION routine, the numerator required to calculate weighted average tracer mixing ratios below the cloud base was calculated outside of the appropriate time loop and mass was not conserved properly. To fix this issue, we commented out lines 2189 and 2198 containing the old declaration and calculations, added a line at 2199 to calculate the variable DELP (`DELP(K) = PEDGE(K) - PEDGE(K+1)`) for each layer below the cloud, and added the block of code below to line 2244 within the time step loop.

```
! Fix to QB_NUM calculation to conserve mass
QB_NUM = 0e+0_fp
DO K = 1, CLDBASE-1
   QB_NUM = QB_NUM + Q(K,IC) * DELP(K)
```

```
ENDDO
```

Because QB_NUM had been defined as an array variable at the beginning of the routine, we also commented out line 2083, defining it as an array, and added the variable to the list of scalars in line 2078 within DO_MERRA_CONVECTION. We additionally defined DELP as an array at line 2084 using, `REAL(fp)::DELP(LLPAR)` This change was committed to the git repository on 05/12/2016 with commit ID: .

### 10.1.7 Negative STT tracer warnings after chemistry in 0.5x0.666 models

Were getting large numbers of these warnings (`http://wiki.seas.harvard.edu/geos-chem/index.php/Nested_Model_Working_Group#Bug_fixes_for_the_0.25_x_0.3125_nested_CH_simulation`) in the nested models for Ozone in the upper stratosphere, near the tropopause. We eventually isolated the problem as occuring during chemistry or photolysis and decided the actual process that was causing negative tracer concentrations wasn't worth fixing. So, we simply removed the lines printing warnings about it to the log files, but left the line zeroing out the concentration values intact. Lines 285-287 in GeosCore/tracer_mod.F were deleted on 5/20/2016. This change was committed to the git repository on 05/23/2016 with commit ID: .

## 10.2 Lightning NOx redistribution for 2x2.5 and 0.25x0.3125 NA simulation missing after Oct. 2014

Pulled monthly lightning flash rates from Geos Chem and sent to Lee Murray. He replied with an updated copy of hcox_lightnox_mod.F90 and netcdf files containing the updated redistribution for the 2x2.5 and 0.25x0.3125 NA models. There is a copy of these located on T at `T:\eohprojs\CDC_climatechange\GC_LXUpdate`. hcox_lightnox_mod.F90 was incorporated into the source code and the source code recompiled for the affected simulations on 10/7/2016. The lightning redistribution files were replaced within ExtData at `~/ExtData/CHEM_INPUTS/HEMCO/LIGHTNOX/v2014-07`. Hemco_Config.rc files for the two affected simulations were edited to point to the new files by editing the lightnox section to read as the code below, and the simulation was restarted in Oct. 2014.

```
#==============================================================================
# --- LIGHTNOX emissions (Extension 103) ---
#
# Local redistribution factors: These are model-specific and hence need to be
# redefined for different model/resolution simulations!!!
#==============================================================================
103 LIGHTNOX_OTDLIS $ROOT/LIGHTNOX/v2014-07/OTD-LIS-Local-Redist.CTH.v5.geosfp.$RES.v20161007.nc OTD $YYYY/1
```

### 10.2.1 Oustanding bugs to be aware of

A few outstanding bugs remain in the 10.01 source code that were not addressed by us because the impact was minimal, the issue had been outstanding in previous versions, and/or we weren't interested in the affected output. Some that future users may wish to be aware of include:

1. ND21 diagnostic output - index mismatching between aerosol_mod, diag3_mod, and gamap_mod. This diagnostic outputs total AOD values to the trac_avg files. We typically rely on species-specific AOD values reported from ND49, but note that ND21 output is expected to be incorrect.
2. Hydrophobic aersol properties - incorrect units used for BC/OC radii.
3. Dust size bins - dust was treated as primarily accumulation, rather than course, mode.
4. Emissions mask overlap at 30N over N. Africa - results in zero emissions over a small part of Northern Africa
5. Accounting of moisture in tracer unit conversions - mix up in humidity of air during transport.
6. FINN extension to HEMCO - broken. Scaling factors for OC/BC applied incorrectly.

# 11 Using HEMCO independently

The emissions component for GEOS-CHEM, HEMCO, can be used independently to obtain emissions information. An independent copy of the HEMCO utility can be compiled along with Geos-Chem and used to obtain emissions information over any defined grid and area. More information about how to run HEMCO on its own can be obtained from christoph kellers github account for the software at: `https://github.com/christophkeller`, and from the article about the software, "HEMCO v1.0: a versatile, ESMF-compliant component for calculating emissions in atmospheric models": `http://www.geosci-model-dev.net/7/1409/2014/gmd-7-1409-2014.html` However, note that the input files for stand-alone HEMCO are different from those used for Geos-Chem and not as well documented, so if you wish to obtain emissions information, it may be easier to simply run Geos-Chem with all modules excepting emissions and aerosols turned off.