Improvement of the c48 GEOS-Chem v14.4.0 Off-Line CO Vertical Transport via the Integration of a Divergence-Goverened Re-gridding Scheme

Jace Alloway (1006940802 / alloway1)

A report submitted in conformity with the requirements of the summer undergraduate research project PHY479Y under the supervision of Professor Dylan Jones.

Abstract This work examines the issue of vertical transport reduction in the c48 off-line chemical transport model (CTM) GEOS-Chem. High-resolution, full-chemistry tracer models are generally computationally costly due to data storage and mathematical coupling of chemical tracer sources and sinks in the atmosphere. A reduction in resolution of the CTM GEOS-Chem from c360 to c48 assists in diminishing this computational cost, however it has been previously noted [1] that this reduction weakens the vertical motions between grid layers in the atmosphere. A divergence-governed re-gridding scheme was applied to input c360 wind vector components using climate data operators (CDO) in an attempt to improve the vertical motions of CO transport in the c48, 72-level rectilinear branch of GEOS-Chem version 14.4.0. By comparing the outputs and differences between the initial c48 inputs and the re-gridded c48 inputs, it was found that applying the re-gridding scheme generally increased the vertical transport either in concentration or along the pressure axis of the CO profile at specific columns. Challenges, improvements, and the continuation of the project are discussed.

I INTRODUCTION

A. Background

A atmospheric model is a simplified representation of an atmospheric system which enables the inference of that system[3] via real-world observations. These models are deterministic: only requiring a set of initial conditions (a "restart") for them to determine a unique final state, even if chaos is present. The models considered in this project were mathematical, that is, implementing mathematical equations to govern the complexity and intricacy of the system[3]. An "on-line" model is therefore a model which runs under the corrections of real-time observed data, while an "off-line" model runs via assimilated data from an on-line model.

Goddard Earth Observing System - Chemistry (GEOS-Chem) is an off-line mathematical chemical transport model (CTM) implementing meteorological data assimilated from the Global Modeling and Assimilation Office (GMAO) general circulation model (GCM) of the National Aeronautics and Space Administration (NASA). GEOS-Chem, developed by the Harvard School of Applied Sciences and Engineering[4], is a multi-faceted model which can be executed with various build parameters such as grid resolution, high-performance (GCHP), and chemical tracer content, making it a practical approach of modeling various components of the atmosphere.

B. GEOS-Chem Operation

Since atmospheric models become chaotic over long run periods, it is necessary to apply a "correction" over intervals of time, else the model tends to drift too far from the physical reality. This concept is directly applied to on-line GCMs, such as GEOS-Forward Processing (GEOS-FP), from which accurate and reliable data can be archived every 6 hours by measurement devices (satellites, buoys, aircraft, ground stations, etc). GEOS-FP archives data using a forward-timestep integration method, iterating over continuity equations until an optimal solution is obtained[3]. The class of archived data from GEOS-FP from GMAO implemented

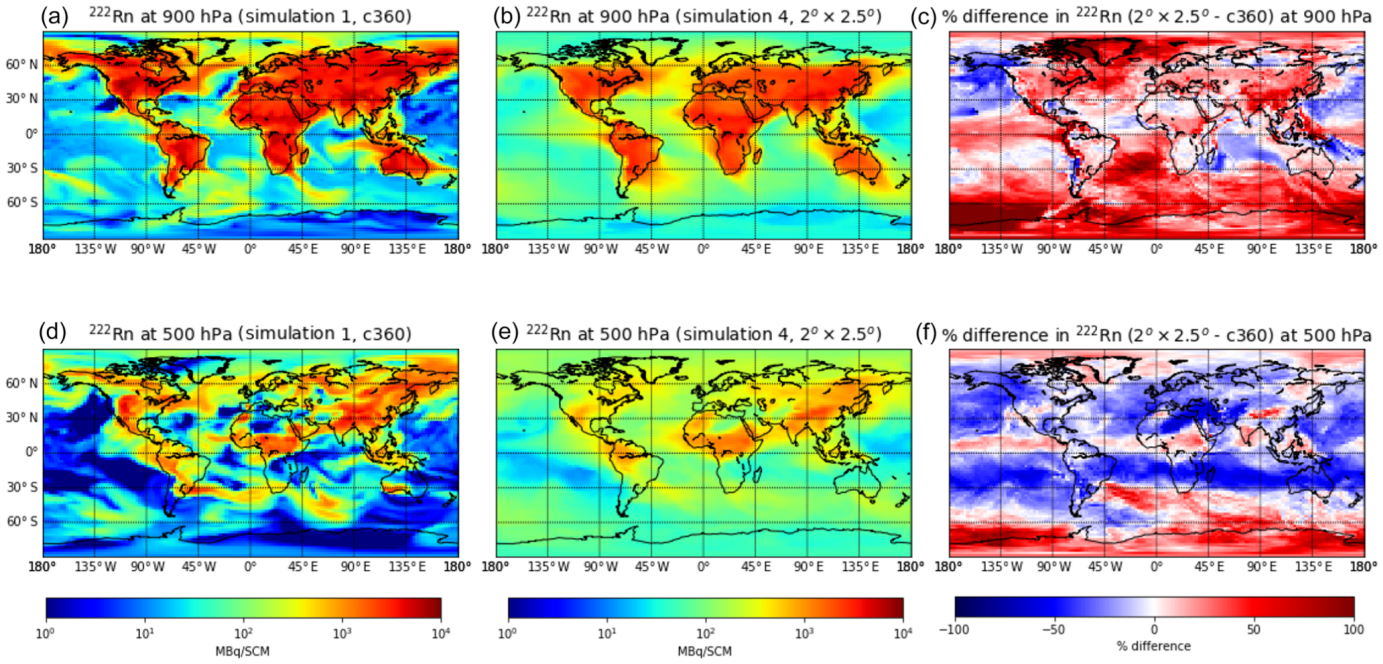

Figure 1: from Yu et al., 2018 [1]. ^{222}Rn *concentrations at 900 hPa ((a)-(c)) and 500 hPa ((d)-(f)) produced by GEOS-Chem at c360 and c48 resolutions. Figures (c) and (f) display the percentage error between the two resolutions.* ^{222}Rn *lifetime (∼ 4 days) is much shorter compared to that of the vertical mixing time in the troposphere (∼ 1 month) making for strong vertical gradients* [1] *and evidence of vertical transport.*

---

here to GEOS-Chem is the Modern-Era Retrospective analysis for Research and Applications volume 2 (MERRA-2) meteorological fields (MetFields). GEOS-Chem then implements the data and executes chemical evolution over the specified time period.

The only MERRA-2 MetFields run by GEOS-Chem in this project were the latitudinal and longitudinal components of the wind evolution (respectively named "Met_U" and "Met_V") and the carbon monoxide (CO) tracer concentration emissions, removals, and advections. CO was selected as the only chemical tracer since the lifetime of CO in the atmosphere (∼ 1-2 months) is significantly shorter than that of CO_2 (∼ 100 years) and CH_4 (∼ 9-12 years), making it an excellent tracer for atmospheric transport[5] as the total atmospheric concentration is not overwhelmed by emissions. This is why GEOS-Chem version 14.4.0 was used, since any later releases of the code concatenate all carbon-based emissions into one group, making it more difficult to isolate tracer transport.

GMAO provides MERRA-2 data in a variety of resolutions[4], hence GEOS-Chem can be compiled and executed at such grid resolutions as well. The only resolutions investigated for this project were the rectilinear (called "GCClassic") (lat × lon) $2 \times 2.5^\circ$ (c48) resolution for GEOS-Chem and the $0.25 \times 0.3125^\circ$ (c360) resolution for some MERRA-2 files. There are other executable versions of GEOS-Chem, such as GCHP which implements a cubical-spheroid grid to process fields, or other GCClassic resolutions such as $4 \times 5^\circ$ and $0.5 \times 0.625^\circ$ rectilinear grids, but these will not be investigated here.

C. Motivation

Running GCClassic at a c360 resolution is computationally costly and costly in terms of data storage due to the large number of data points and chemical tracer coupling in the conservation equations. One solution to improve this is to reduce the GCClassic grid resolution from c360 to c48.

It has previously been noted that a reduction in GEOS-Chem and archived data resolution (c360 to c48) affects the vertical transport of tracer

concentrations[1] (Figure 1). This vertical transport considers convective mass fluxes (CMF), barometric gradients, radiative heating, and Coriolis forces, for instance. There are a variety of explanations as to why this occurs. First is that, due to an increase in grid size in GCClassic, vertical eddy motions otherwise present in high-resolution wind fields are averaged out when the grid cell size increases. This includes vertical mixing depths[1] (Figure 2). This can also originate in the choice of spherical projection utilized onto a rectilinear grid, since a cylindrical projection compresses parallels latitudinally and Mercator projections stretch meridians longitudinally[2].

Furthermore, operating at lower resolutions require GEOS-Chem to contain a "pressure-fixer" operator, not previously included in the c360 executable. This is because, at lower resolutions, pressure gradients are averaged and vertical transport becomes chaotic, resulting in a violating of mass-conservation[2].

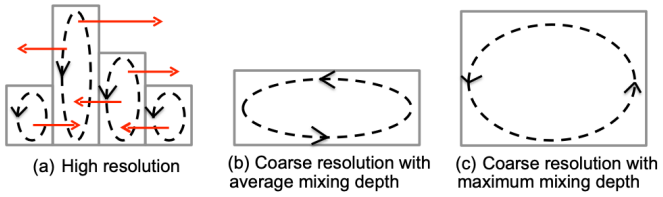

Figure 2: from Yu et al., 2018. Illustration of the affects of resolution reduction in regards to the vertical mixing depth of GEOS-Chem. (a) represents the high-resolution transport vertical eddies, (b) shows the result of increasing resolution but leaving the mixing depth between layers the same, and (c) represents a vertical profile with a coarse resolution and large mixing depth.

General re-gridding schemes utilize various weight-oriented methods to compute optimal interpolation or averaging between resolutions. GMAO computes a general regridding for assimilated data using Python packages `SciPy interp2d`, `xemsf` and `GCPy` [6],[7]. This can include bi-linear (averaging) methods, patch (least squares) interpolation, conservative (1st and 2nd order integration) weighting, or nearest-source (vector tip-tail) re-gridding[7]. Although it is unclear which operation is taken out on c360 MERRA-2 data to yield the c48 fields, the intention behind implementing a divergence-goverened re-gridding scheme is to attempt to obtain an increased vertical transport in CO tracer concentration. This is simply because taking a divergence retains gradient information about the sources and sinks of the vector field, which is how GEOS-Chem reads vertical motions between vertical layers. Some information would otherwise be lost if integrated or interpolated.

D. Review of Model Mathematics

As previously mentioned, the on-line GMAO model GEOS-FP writes archived data based on inputs every 6 hours, else the model may drift from actual observation. It is important to note how GEOS-FP governs the dynamics and chemistry of the tracers, namely CO, since an increase in tracer chemistry implies an increase in coupled conservation equations.

In general, a chemical tracer of concentration q (units mol/mol dry air) must obey mass conservation

$$\frac{Dq\rho}{Dt} \equiv \frac{\partial q\rho}{\partial t} + \nabla \cdot (\mathbf{v}q\rho) = A + L \qquad (1)$$

where ρ is the air density, $\mathbf{v}$ are the wind flows, A is the additions of tracer concentrations and L is the loss of tracer concentrations. Pure tracer advection has $A + L = 0$ [3], which is goverened by a "finite-volume" (FV) numerical scheme[2] to reduce numerical noise. Setting $A + L = 0$ and $q = 1$, we obtain the mass conservation equation for air.

When chemistry is introduced, chemical reactions in the atmosphere occur, introducing several couplings between species at various reaction rates[3]. Since only CO concentrations are considered here, the sources considered are antropogenic (accounting for nearly half), wildfire and agricultural emissions (biomass combustion, where CO is an freely produced unstable product), biogenics (oxidation from non-methane volatile organic compounds (NMVOC))[5]. The primary sink of CO is methane oxidation in the troposphere[5],[8]:

$$\mathrm{CH_4 + O_2 \longrightarrow CO_2 + H_2 + H_2O} \qquad (2)$$

In general, for a chemical reaction

$$\mathrm{A + B \longrightarrow C + D + \ldots} \qquad (3)$$

with reaction rate r_{AB} and chemical densities ρ_a, ρ_b with photolysis frequency (proportionality factor

between concentration and productions and losses) J_A, J_B, the conservation equation for a species i takes the form[3]

$$\left[\frac{\partial \rho_i}{\partial t}\right]_{\text{chem}} = \sum_{i \neq j} \left[J_j \rho_j + \sum_{k \neq i} r_{jk} \rho_j \rho_k \right] - \left[J_i + \sum_j r_{ij} \rho_j \right] \rho_i \quad (4)$$

where the j and k indices indicate other chemical species which can react with i to produce or remove concentration, summing over all individual processes. As previously mentioned, this introduces chemical couplings between species, making multi-species computations costly.

II ANALYSIS METHODOLOGY

A. Building and Running GEOS-Chem v14.4.0

The GEOS-Chem model utilized in this project was the c48 rectilinear grid (GCClassic) resolution of version 14.4.0 (May 2024[4]). Retrieved from Github, the model was compiled on an Intel server with a 2013 intel64 iFortran compiler. The model was built at c48 ($2 \times 2.5°$) resolution with 144 longitudal cells and 91 latitudinal cells with 72 vertical layers. The input data was selected to be the carbon model with tracer CO with data archived from GMAO GEOS-FP MERRA-2.

Once the scripts were compiled and the Harmonized Emissions Component (HEMCO) (to govern CO chemistry evolution) was built, the `HISTORY.rc` file was edited to write output files every (simulation) hour. That is, every 6 iterations of timestepping, which occurs every 10 minutes. Usually, hourly output writing is dense in data storage for long runtimes, however this particular model was specified to run from July 1 2018 to August 1 2018 (1 month summer), implying that hourly output files should be sufficient for analyzing CO evolution. A single-month summer execution of the model was chosen as the temporal parameter because summer is an optimal runtime to investigate carbon emissions (forest fires, surface heating, agricultural biogenics) from which CO was chosen, and a one-month run was deemed long enough to attempt to observe the effects of re-gridding in tracer evolution. This was configured in the `geoschem_config.yml` file in the GCClassic directory. To start the model, a restart file (initial conditions) was copied from a previous run in another directory, which was also changed in the GEOS configuration file.

Lastly, since the data directory was previously installed on the particular server utilized in another location, the `root_data_dir` was changed to read the input MetFields from the appropriate location. GEOS-Chem was run by executing `gcclassic > output.log &` in the linux command line, so the execution could be taken out in the background and write the output to a log file.

B. Re-gridding

The re-gridding process of the tracer was implemented by performing a re-gridding scheme on the U and V components of the winds (Met_U and Met_V files), located in the MERRA-2 A3dyn c360 directory. To assist in characterizing the magnitude of the differences, the initial build directory was copied to change the `HEMCO_Config.rc.gmao_metfields` file, the file which reads all of the MetField (data) input directories. In particular, the 'A3dyn' directory was changed to a seperate input directory where the re-gridded files were written.

GEOS-Chem inputs and outputs are written using Network Common Data Forms (NetCDF), which are most useful for storing large multi-faceted data arrays. Since NetCDF files are multi-dimensional, they can be read into various scripts and isolated along axes using `.sel` or `.selvar` operators. Values can be explicitly obtained from these arrays using `.values`. For Python scripts, data and be read and written using the `xarray` or `NetCDF4` packages.

The re-gridding process was implemented using a perl script to perform operations on the Met_U and Met_V wind files using Climate Data Operators (CDO). These files were first imported from the initial GEOS-Chem c360 high-resolution MERRA-2 'A3dyn' data archive from GMAO. The latitudinal and longitudinal components of the winds were isolated using the `.selvar` operator while the other dimensions ('DTRAIN', 'OMEGA',

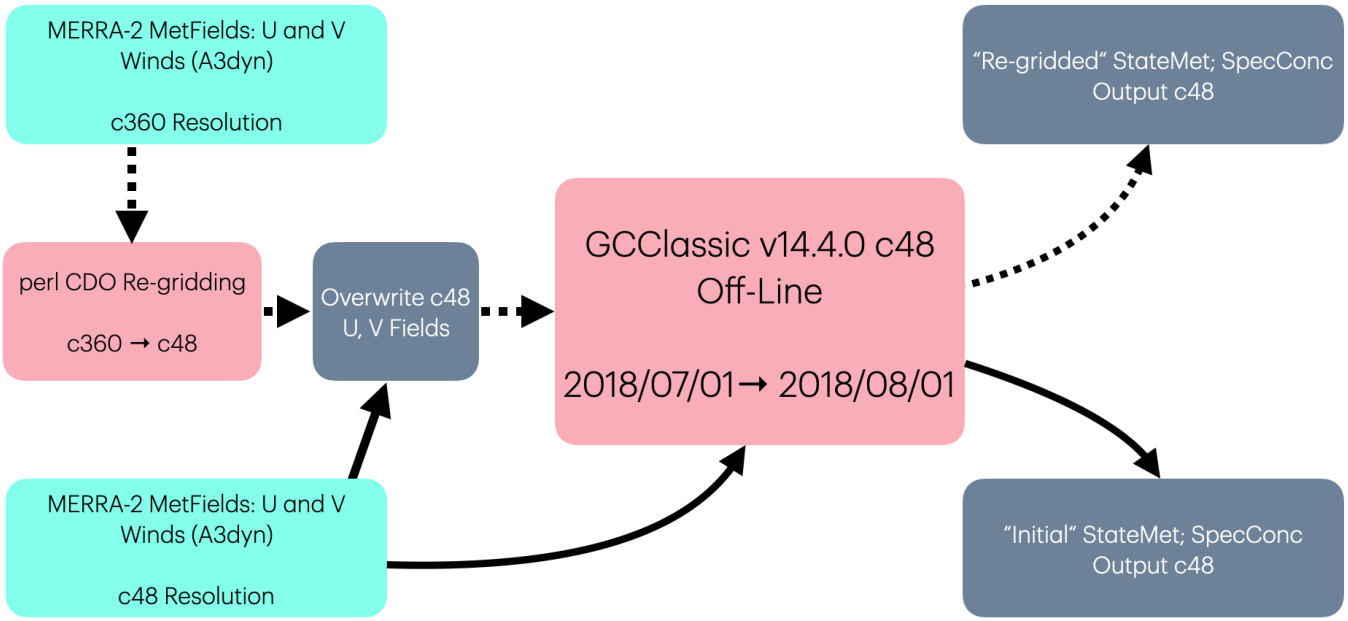

Figure 3: a schematic indicating the directory paths and order of re-gridding and running GEOS-Chem.

'RH') were left unchanged. Using the spectral method, the CDO script transformed the vector data from a Gaussian space (gridpoint) to a spectral space (`gp2sp`), from which a divergence operator was applied to the transformed array along each of the 72 layers (`uv2dv`). After returning to a Gaussian space, the resolution was then decreased from c360 to c48 using the `remapcon` operator on the scalar divergence field. The data was lastly transformed back into spectral weights and integrated into a vector field (`dv2uv`), from which the adjusted spectral weights were converted back into Gaussian space according to the GEOS-Chem c48 metrics (`sp2gp`). The new U and V fields were then written back to individual NetCDF formats separate from the initial A3dyn files in the c48 data directory.

To integrate the re-gridded wind files back into the GEOS-Chem c48 input, a Python script was written to overwrite the set of initial A3dyn Met_U and Met_V files in the c48 MERRA-2 directory, since there was some trouble in defining specific paths for the individual U and V files in the `HEMCO_Config.rc.gmao_metfields` file. The Python script implemented the NetCDF reading and writing with `xarray`, and this was quickly completed by setting the `.values` arguments of each U and V file equal to each other, then re-writing the files into an output directory read as the 'A3dyn' input by the copied GCClassic run directory, as mentioned. This process is represented using the schematic in Figure 3.

C. Analysis of Outputs

Results were analyzed using a Python script and `matplotlib.pyplot`. The expectation, as outlined, was to attempt to observe an increase in CO concentration at tropospheric pressure levels or to observe a 'shift' upward of CO concentration along the pressure spectrum.

Two approaches were taken. The first approach was to plot CO concentrations at low-level pressure values (600-900 hPa) across all 91×144 grid boxes, which is a cross-section. This was done to observe the CO sources and sinks at specific pressure levels and time values. Using `matplotlib.animate`, the CO concentration evolution was animated using the hourly timesteps to view locations of CO differences between the initial and re-gridded runs on a world-scale map. The data was overlayed on top of a coastline grid map with `mpl_toolkits.basemap`.

The second approach was to utilize locations (latitude, longitude) and timestamps noted from the cross-section animation where the re-gridded CO concentration appeared different than that of the initial run. This was done by inspection. Once some locations and timestamps of interest were determined, the carbon concentration profiles were

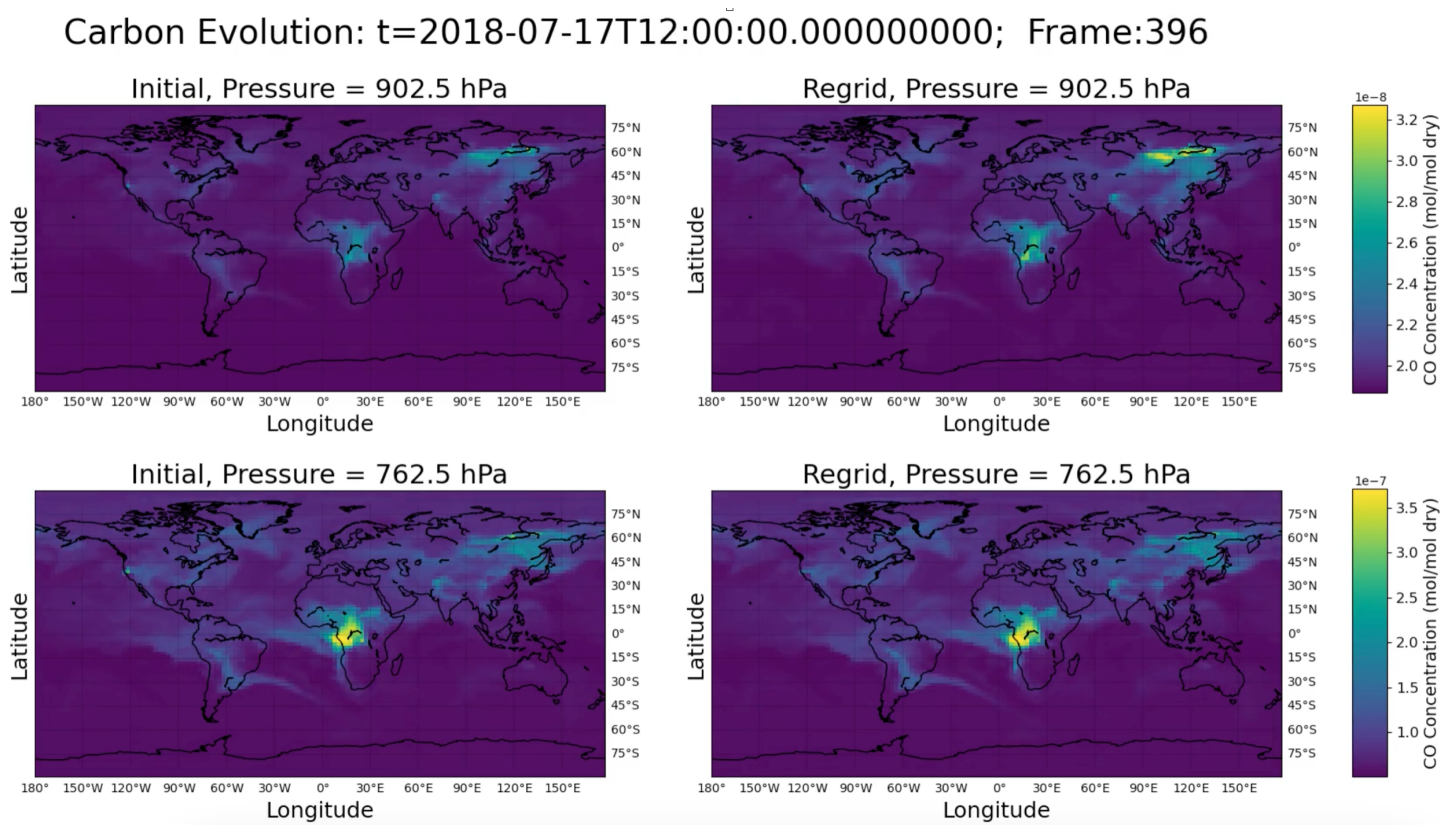

Figure 4: a selected timestamp in the CO evolution. The top row indicates initial and re-gridded concentration distributions at 902.5 hPa, while the second row indicates the distributions at 762.5 hPa. Differences can be observed at 902.5 hPa over Africa and Siberia, and at 762.5 hPa over North America (California), where the re-gridded distributions appears different than that of the initial distribution.

---

plotted in `subplots` along specific grid cells across all 1-72 levels.

Both initial and re-gridded CO profiles were plotted together to attempt to observe differences. They were plotted with an inverted y-axis so that the surface pressure (1000 hPa) was at the bottom, and the Stratospheric pressure was at the top ($\sim$ 1 hPa). Once this was done, specific sections of the profiles of interest were isolated with `ax.set_xlim(), ax.set_ylim()` to better observe the differences.

In conjunction with the previous set of plots, the differences in concentrations were also plotted, simply by taking the differences of CO initial and re-gridded values. The expectation would be to see a backward set of 'S' shapes, with a positive difference at low pressure levels (closer to Stratosphere) and a negative difference at higher pressure levels (closer to surface), since this would indicate an upward shift in vertical transport of CO concentration.

III DISCUSSION

A. Results

Using the CO animation previously described, 4 locations of interest (isolating specific values of latitude, longitude, and time) were selected. An instance of this process is shown in Figure 4, which displays significant concentration differences above Africa and Siberia (caused by forest fires) at 902.5 hPa and over some parts of North America at 762.5 hPa. The selected locations appear in Table 1 (Appendix 1). The profile over Siberian Russia is shown in Figure 5a, while the other profiles and differences are plotted in Appendix 1.

Observe that, in Figure 5a, the re-gridded profile of CO concentration is shifted upwards along

the pressure level axis. This indicates that the re-gridding performed on the c360 winds increases the amount of tracer propelled upwards. This can also be observed in Figures 5b-8 in Appendix 1, with a general increase in CO concentration at fixed pressure levels, or a shift upwards in pressure upwards between re-gridded wind outputs and the initial outputs. Furthermore, in Figures 5b-8, a general increase in CO concentration is observed at lower pressure levels (more vertical), indicated by the positive difference in concentration between the re-gridded and initial outputs.

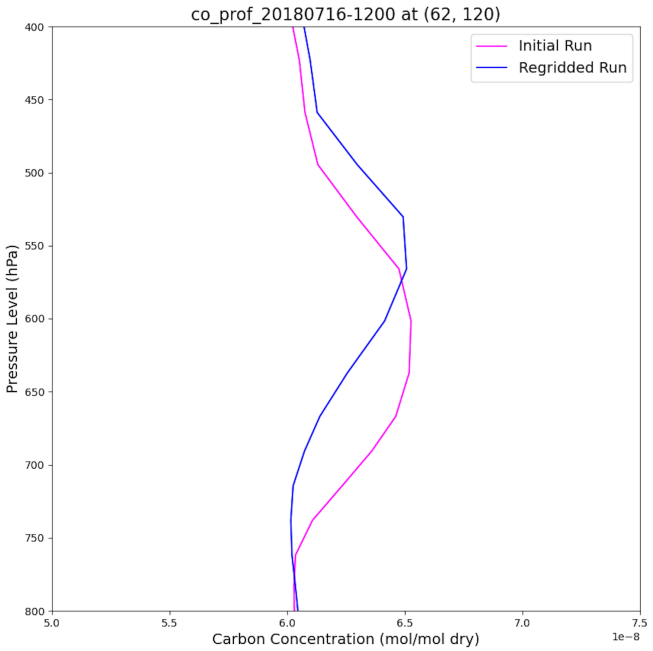

Figure 5a: CO profile taken over Siberian Russia (62 N, 120 E) at July 16 2018 1200h. The x-axis indicates CO concentration in mol/mol of dry air, while the y-axis indicates the pressure level.

These results imply that the CO vertical transport due to divergence re-gridding has improved, whether that be in concentration level or along the pressure axis.

B. Challenges and Improvements

The primary difficulty in achieving the results was building and running the appropriate version of GEOS-Chem. The intended installation was version 14.4.0 due to the presence of the isolated CO tracer, however the Github installation indication version 14.4.1. This difference established problems when attempting to perform a compile on the code installation, since the HEMCO integration was inconsistent with what was expected.

The second problem with installing the program was ensuring the code was compiling on the proper iFortran Intel core. At first, the cmake expected a different compiler than what was being used in the `~/.bashrc` file. This problem was solved by specifying an older 2013 compiler and compiling the code on a single server. The reason for this is unknown.

The third problem encountered in running GEOS-Chem was ensuring that HEMCO was reading the correct input MetFields and restart files from the correct locations. This was fixed by specifying the correct NetCDF environment in the `~/.bashrc` file and overwriting the old directories in the `HEMCO_Config.rc.gmao_metfields` file with the correct data location.

Considering the promising results of the increase in vertical transport of the CO tracer, the next step would be to run the model over a longer period of time with an increase in tracer content to better characterize the impact.

In terms of characterizing the impact on the CO tracer, the analysis could be improved by calculating a percentage difference in between vertical layers (cross-sections) such was done in Yu et al., 2018 (Figure 1). This could be applied via the equation

$$\% \text{ difference} = 200 \cdot \frac{|A - B|}{A + B} \tag{5}$$

for each concentration at each grid cell and level. The intention would be to compare the percentage differences between c48 initial, c360 runs and c48 re-gridded, c360 runs to attempt to notice a net difference in the latter comparison with that of the first. It would then be expected for the c48 re-gridded, c360 comparison to have a lower net percentage difference. In general, determining a metric to characterize differences is one important aspect of measuring the impact of the re-gridding. Lastly, it would be worth investigating more locations (other than the 4 mentioned here) on the globe where tracer vertical transport has a noticable difference between initial and re-gridded trials.

C. Conclusion

To conclude, GEOS-Chem is a mathematical CTM used to simulate archived assimilated data

from the GMAO GCM. The purpose of this project was to apply a divergence-goverened re-gridding scheme to advection winds in climate data to attempt to observe an increase in vertical transport of a CO tracer implemented in the c48 rectilinear model of GEOS-Chem with 72 vertical layers. This was taken out using CDOs in a perl script to transform data into spectral weights to take a vector divergence, which was then re-written into NetCDF format and re-directed into the GCClassic input.

It was found that, after applying the re-gridding scheme, vertical transport of the CO tracer was increased in two ways: the first being a general increase in CO concentration at pressure levels, and the second being a shift in CO concentration upwards vertically along the pressure axis.

The next step would be to design a script to perform the re-gridding more efficiently across a set of multiple input MERRA-2 files, and then to run GEOS-Chem with an increased tracer content over a longer period of time (1-2 years).

In terms of analysis, some improvements could be made in order to better characterize the impact of the re-gridding, such as determining percentage differences between c360 and c48 model outputs and examining more locations on the globe where tracer concentrations are increased.

Bibliography

[1] Yu, K., Keller, C.A., Jacob, D.J., Molod, A.M., Eastham, S.D. and Long, M.S., 2018. Errors and improvements in the use of archived meteorological data for chemical transport modeling: an analysis using GEOS-Chem v11-01 driven by GEOS-5 meteorology. Geoscientific Model Development, 11(1), pp.305-319.

[2] Stanevich, I., 2018. Variational Data Assimilation of Satellite Remote Sensing Observations for Improving Methane Simulations in Chemical Transport Models.

[3] Brasseur, G.P. and Jacob, D.J., 2017. Modeling of atmospheric chemistry. Cambridge University Press.

[4] GEOS-Chem Wiki: List of MERRA-2 Met Fields. https://wiki.seas.harvard.edu/geos-chem/index.php/List_of_MERRA-2_met_fields.

[5] Tailong, H., 2022. Mitigating Model Errors in Chemical Data Assimilation: Application of New Data Assimilation and Machine Learning Approaches

[6] Regridding, Earth System Modelling Framework. https://earthsystemmodeling.org/regrid/#regridding-methods.

[7] (n.d.). GMAO software infrastructure. GEOS. https://geos5.org/wiki/index.php?title=GMAO_Software_Infrastructure&mobileaction=toggle_view_desktop

[8] Y. Matsumura, J.B. Moffat, 1994. Partial Oxidation of Methane to Carbon-Monoxide and Hydrogen with Molecular-Oxygen and Nitrous-Oxide over Hydroxyapatite Catalysts. Journal of Catalysis, Volume 148, Issue 1, pp323-333. ISSN 0021-9517, https://doi.org/10.1006/jcat.1994.1213.

Appendix 1: Large Figures and Tables

Index	Latitude (deg)	Longitude (deg)	Timestamp (UTC)	Geographical location
1	0	20 E	2018-07-11 1700h	Africa - Rep of Congo
2	36 N	102.5 E	2018-07-14 2100h	China - Lanzhou
3	77.5 N	144 E	2018-07-14 2100h	Russia - Far Eastern
4	62 N	120 E	2018-07-16 1200h	Russia - Siberian

Table 1: list of selected locations analyzed for CO concentration profiles. Latitude and Longitude are indicated in degrees North, South, East or West. Timestamps shown appear in Coordinated Universal Time (UTC) generated by GEOS-Chem output. Approximate geographical locations are indicated in the last column.

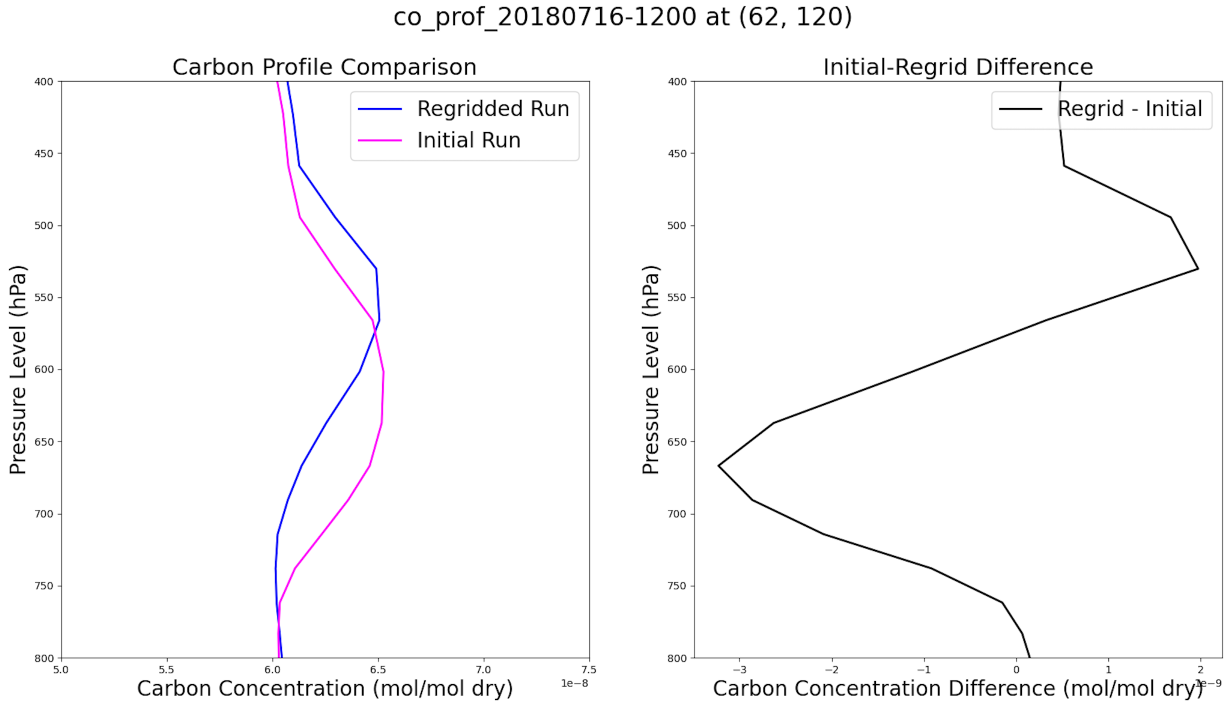

Figure 5b: CO profile over Siberian Russia (62 N, 120 E) at July 16 2018 1200h. The initial comparison is shown on the left, while the difference in chemical concentrations per atmospheric pressure level is shown on the right.

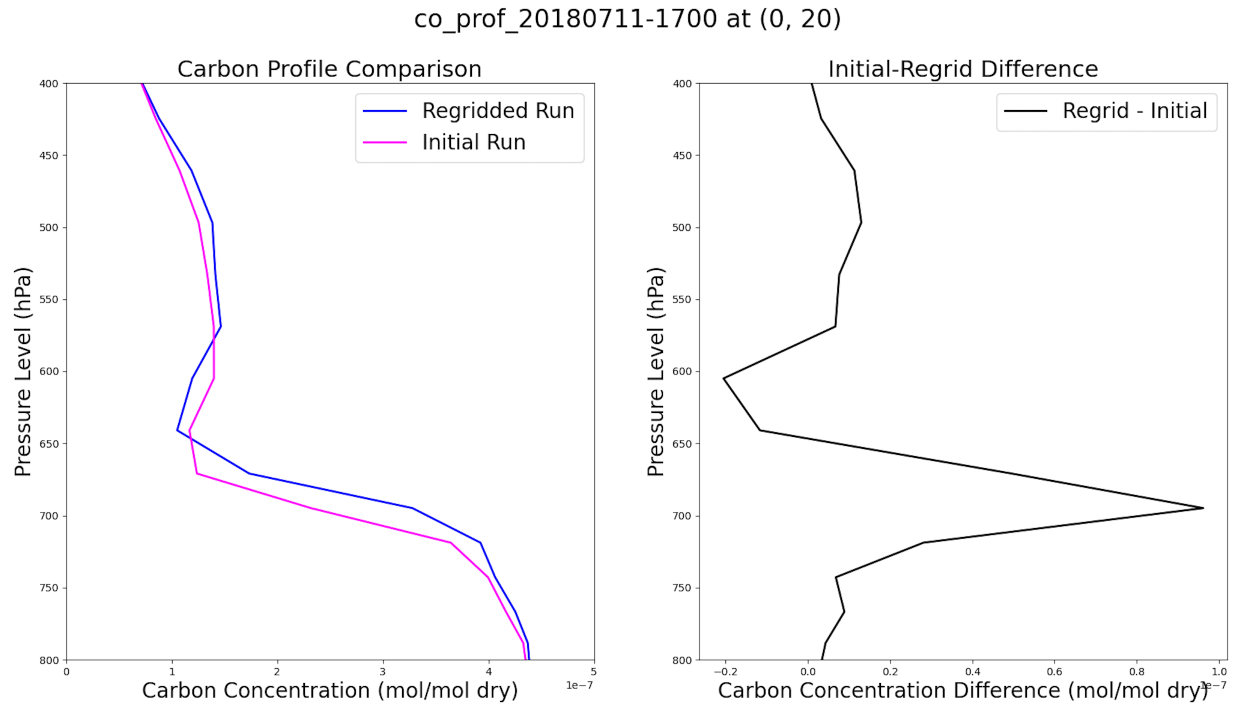

Figure 6: CO profile over Republic of Congo, Africa (0 N, 20 E) at July 11 2018 1700h. The initial comparison is shown on the left, while the difference in chemical concentrations per atmospheric pressure level is shown on the right.

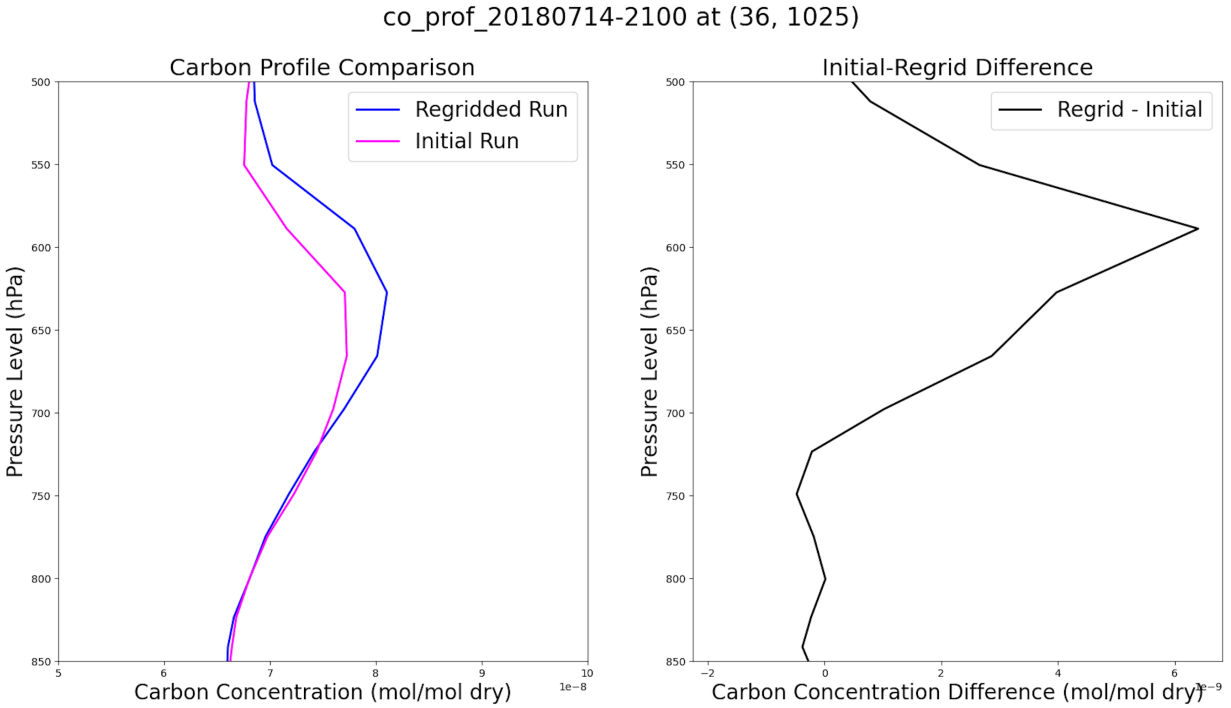

Figure 7: CO profile over Zanzhou, China (36 N, 102.5 E) at July 14 2018 2100h. The initial comparison is shown on the left, while the difference in chemical concentrations per atmospheric pressure level is shown on the right.

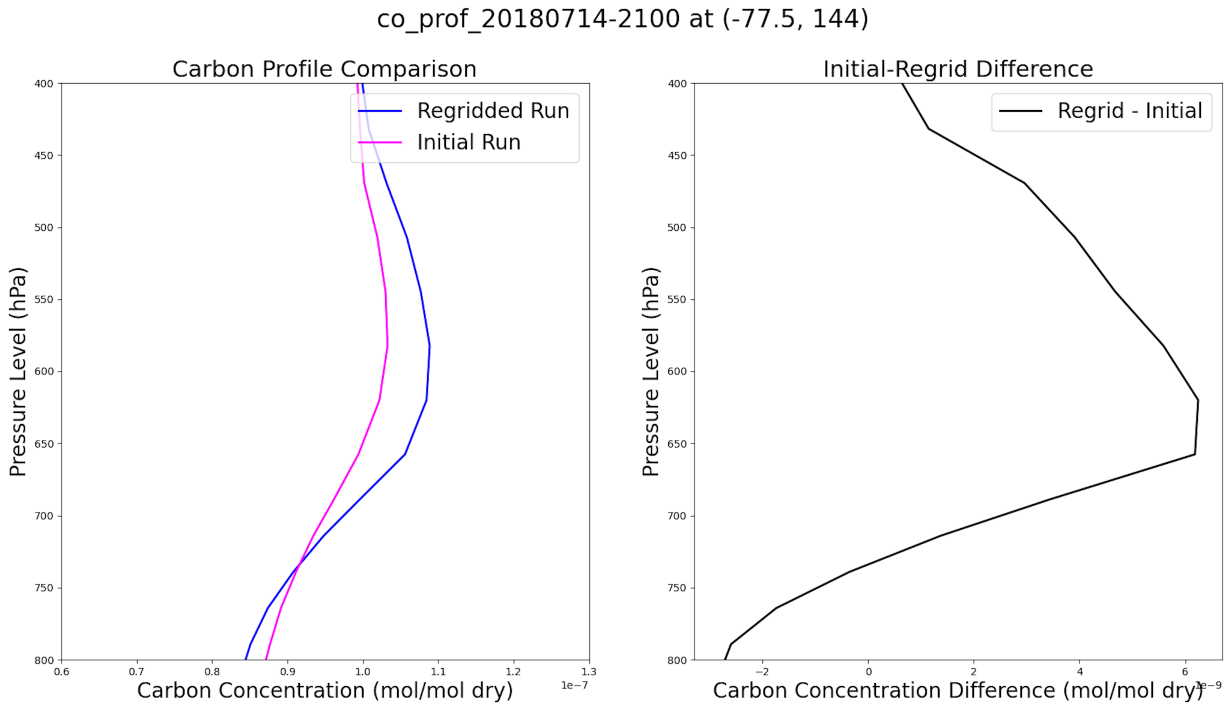

Figure 8: CO profile over Far Eastern Russia (77.5 N, 144 E) at July 14 2018 2100h. The initial comparison is shown on the left, while the difference in chemical concentrations per atmospheric pressure level is shown on the right.

Appendix 2: List of Acronyms and Definitions

- GEOS: Goddard Earth Observing System
- FP: Forward Processing
- GCHP: GEOS-Chem High Performance (cubical-spheroid grid)
- GCClassic: GEOS-Chem Classic (rectilinear grid)
- On-line: Modeling under real-time observations
- Off-line: Modeling under archived assimilated data
- CTM: Chemical Transport Model
- GCM: General Circulation Model
- NASA: National Aeronautics and Space Administration
- GMAO: Global Modeling and Assimilation Office
- MERRA-2: Modern-Era Retrospective analysis for Research and Applications volume 2
- HEMCO: Harmonized Emissions Component
- MetField: Meteorological Field
- FV: Finite-Volume

- NMVOC: Non-Methane Volatile Organic Compounds
- CDO: Climate Data Operator
- NetCDF: Network Common Data Form
- c360: $0.25 \times 0.3125^\circ$ resolution
- c48: $2 \times 2.5^\circ$ resolution
- UTC: Coordinated Universal Time
- CMF: Convective Mass Flux