**Part I: model selection**

1. Environmental drivers, fossil fuel, biomass burning, and oceanic fluxes

Environmental drivers are from MERRA2 meteorology. We use *wget* to download MERRA2 datasets. Wget script included.

Fossil fuels are from ODIAC.

Biomass burning are from GFED4 in HDF5 format.

Oceanic fluxes are from ECCO-Darwin.

all code used for downloading are included.

1. **X** matrix

Seven columns of intercepts

seven columns for each driver

one column for combined fossil fuel, biomass burning, and oceanic fluxes

Code included.

1. Pass X through h()

Code included.

1. BIC model selection procedure

Code included.

After model selection, we will have a range of drivers that will be used for the GIM.

**Part II: create initial guess**

With selected drivers h(X) and satellite observations y, we can readily estimate beta values for individual drivers; and the we use X\*beta0 to get an initial guess for surface fluxes.

**Part III: RML for decorrelation length and decorrelation time**

We estimate the decorrelation time/length for land and ocean, respectively.

**Part IV: GIM**

All the modified FORTRAN code associated with GEOS-Chem forward and inverse model are included. Anywhere that are modified are with commented with ‘zc’. Please search key word ‘zc’ to find lines been modified.

The observation operator needs us to have XCO2 data in daily netcdf format (each single file contains a single day).

Pro scrips are also included, mostly used to convert bpch files to netcdf or csv files.

Appendix: conditional realization used for Figure 3 in our ERL 2021.