# MACC Reanalysis NCDF Information

The NCDF files will be taken from /space/hall1/sitestore/eccc/aq/r1/alh002/NCDF/SPECIES/, which is a directory containing yearly species data for parameters Sulphur Dioxide, GEMS Ozone, the three Dust Aerosol bins, Sulphate Aerosol bin, Formaldehyde, Methane, Carbon Monoxide, Nitrogen Oxides, and Specific Humidity. Each of these species is separated by their own directory, containing a yearly NetCDF file separated by 6 hour timesteps for the 60 levels generated by the MACC Reanalysis. Each of these contains files for the years 2008 to 2012, with the exception of Sulphur Dioxide, which only contains data for 2009 to 2012, and October to December of 2008 found in the INCOMPLETE/ directory.

# multi\_fixed\_pv\_calc.py

# fixed\_monthlymean\_automated.py

The script can be called through python, and takes arguments --m (month integer, with 0 as January and 11 as December), and --p (ECMWF named parameter name, a table of all ECMWF parameters is available at [the ECMWF Parameter database](http://apps.ecmwf.int/codes/grib/param-db?&discipline=All&category=210)). This script is intended to be used with ECMWF species data over 5 years. It calculates the monthly mean for a specific species at a specific month, defined by the args, and outputs them to a FST file.

The natural logarithm of surface pressure files required for interpolating the data over a constant range of pressures is found at /space/hall1/sitestore/eccc/aq/r1/alh002/NCDF/LNSP/, as defined at lines [$X-$Y], and is separated by year spanning from 2008 to 2012.

### Script activity

The script is strictly imports and functions prior to line 270. On line 272, arguments are parsed. The two arguments that can be parsed from the command line are month integer (starting at 0 for January and ending at 11 for December), and the case-insensitive name of the parameter you wish to obtain the monthly mean of (as per ECMWF parameter names). Because there is only 4 available complete years (2009 to 2012) for Sulphur dioxide, if the parameter name is “so2”, a flag will be set to True (so2\_flag) which will change certain parameters in order to fit the years properly. Up until line 318, the script is setting up general variables. On line 318, if it doesn’t exist, the file will be created. The file is then opened through the rpnpy.librmn package for reading and writing. The variable directory containing the NetCDF files is acquired through line 328 and in 329, it is sorted in ascending order by year. The monthly mean array is then initialized through NumPy with zeros, to the length of the variable directory, and with three other dimensions for the vertical levels, latitudes, and longitudes respectively. In the for loop, each individual NCDF file is opened, and it’s respective years natural log of surface pressure (LNSP) file. In the “get\_var” function, the timesteps for the month integer are spliced from the year spanning array, and the variable is taken through the PyGeode object opened in the arguments. The same is done with the LNSP in the “get\_pressures” function. The shape of the parameter is then saved to variables “timelen, levlen, latlen, lonlen”. The latitudes dimension is then flipped, so the southernmost latitudes are assigned to index 0, and increasing in indexes is northbound. This is done to correspond with definition of the grid when writing to fst files using the rpnpy.librmn package. On line 352, the variable is then interpolated to the constant pressures defined in “levels.py”. Afterwards, the mean of the variable array is taken with respect to all of its 6 hour interval timesteps. It is then assigned to its year index in the “monthly\_mean” array.

This for loop continues until all years are spliced, latitudes flipped, vertically interpolated, averaged, and assigned. The “monthly\_mean” array is then averaged with respect to the years analyzed, leaving one array with 3 dimensions corresponding to 59 vertical levels, 161 latitudinal grid points, and 320 longitudinal grid points. The resolution of each gridpoint (which should be square) is taken in line 379, and the FST grid parameters are taken in lines 381 to 390, and encoded in line 392. In line 402 the grid is written to the FST file opened/created previously. Lines 406 to 419 are required to convert and change the partial grids because, as they’re written, they are encoded with some dissonance in their defined latitudes and longitudes in that they are not always separated by 1.125. Thus, their values are corrected and rewritten. In line 420, the vertical grid descriptor is written using the rpnpy.vgd package. The partial grid descriptors are then written to the fst file. The rest of the script writes the data records level by level, encoding the ip1 as a pressure level as it goes on onto the files. The ‘dateo’ key in the record is constantly 20120101, as it does not matter what date is defined due to this being a set of climatologies averaged for a single month over 4/5 years.

Directories