Script4DataExtraction Notes

**Input Data:** Original data in NetCDF format.

**Output Data:** Data for a specific variable and spatial extent in NetCDF format.

**Description:** Extracts data for a specific variable and spatial extent and exports them to a new file for use in the Func4ModelComparison Script.

**Date:** June 2022

**Author:** Emma Perkins

Step 1: Import Data

* Imports necessary python packages (xarray and glob) for use in later steps.

Step 2: Load Data

* Loads in full data by combining all of the data from all of the specified files.
* Files can be specified individually, or, in the case that there are many files of interest with similar paths and similar names, multiple files can be specified at once using the glob module by substituting a ‘\*’ into the path and/or filename in all of the locations in which the paths and filenames differ. Doing so will retrieve all of the paths and files that fit the remaining part of the path / filename.

Step 3: Select Variable

* Selects a variable of interest by specifying the name of that particular variable exactly as it appears in the original data.

Step 4: Select Area of Interest

* Selects only the data for the specified area of interest by:
  + Determining the variable names for the spatial coordinates. **These should be changed/specified by the user for each run to match the data in that run.**
  + Renaming the spatial coordinates to be ‘lat’ and ‘lon’
  + Sorting the data by latitude
  + Determining the form of the longitudinal coordinates, either long1 (-180 to 180) or long3 (0 to 360). **These should be changed/specified by the user for each run to match the data in that run.**
  + If the longitudinal coordinates are not already in long1 form, convert the coordinates to long1 form.
  + Select the input area from westernmost to easternmost / left to right, and southernmost to northernmost / bottom to top. Note it is ok for the minimum longitude to be a larger number than the maximum longitude, in this case the script will select the data across 180. **These should be changed/specified by the user for each run to match the data in that run.**
  + Sort the data by longitude.

Step 5: Standardize Time Step

* Standardize the time step of the data if it is not already using the desired time step. Note, this section can be commented out if the data already use the desired time step.
  + **If the data are not using the desired time step, the user should specify the analysis\_time\_type (ex: 6H, 1D, 1M, 1Y, etc.) and the method, .mean(), or .sum() for converting the time step.**

Step 6: Export Data

* Writes the selected data to a new NetCDF file in a location specified by outpath and data\_name. **These should be changed/specified by the user for each run.**

**Submit script to casper job queue:**

#!/usr/bin/env python3

#PBS -N Script4DataExtraction

#PBS -A YourProjectKey

#PBS -l select=NumberOfNodes:ncpus=NumberofCPUs:mem=AmountOfMemoryGB

#PBS -l walltime=HH:MM:SS

#PBS -q casper

#PBS -j oe

export TMPDIR=/glade/scratch/YourUsername/temp

mkdir -p $TMPDIR

module load conda

conda activate npl

python /YourPathToScript/Script4DataExtraction.py