

Running order of the ChesWx Preprocess

1. Setting up the domain:

1.1 File name = setup_domain_grid.R

Pre required file = None

output file = bbox_interp.shp, bbox_stns.shp, bbox_dem.cdo

2. Downloading the Database:

2.1. File Name: download_ahccd.py

Pre required file = None

output file = prcp_ahccd_'ymd'_'ymd'.csv, ahccd_stns.csv

2.2. File Name: download_cpc_hrly_prdp.py

Pre required file = None

output file = precip.hour.'year'.nc

2.3. File Name: download_ushcn.py

Pre required file = None

output file = prcp_ushcn_'ymd'_'ymd'.csv, ushcn_stns.csv

Special Note: CPC file is last updated on 2002 at the destination website. So, No need to run the file.ahccd,ushcn are updated in the 2021 website. Please update it using obiso_env conda environment.

2.4. File Name: download_ghcnd.py

Pre required file = None

Output file = obs_all_'ymd'_'ymd'.nc

Special Note:

GHCND file is going to take around 2 hours to download all the files to year 2020. make sure you have enough space and stable internet connection. Make sure you have set the option `download_updates=False` for preprocessing. But if you want to update the observation data set, you can set it to true for download new files and update it. Make sure you have the `stn_nums`, `tobs_prcp`, `tobs_tmax`, `tobs_tmin` updated. I use the ftp file server to have it manually downloaded from the <ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/> link.

2.5. File Name : `download_nldas2.py`

Pre required file: None

Output file = `nldas2` folder containing the year data

Special Note:

`nldas` is from 1979 to 2021. Make sure it is downloaded in monthly aggregated netcdf file. To download it, you need to create a `.netrc` and `.urs_cookies` file in the `$HOME` directory. You need to have a earth-data login Information in your home directory. Details of the procedure can be found here (https://disc.gsfc.nasa.gov/data-access#mac_linux_wget). Its going to take 1 hour~2 hour depending on the computer speed. Also, activate conda environment `python2` for this script to run. After running the `download_nldas2`, don't forget to run "`merge_gridded_hourly.py`" script to merge all the output in a single file.

3. Data Setup:

3.1 file name: `merge_gridded_hourly.py`

Pre required file: 2.2 and 2.5

output file: `merged_cpc_hrly_prcp.nc`, `merged_nldas2_prcp.nc`

3.2 file name: `calc_obs_cnts.py`

Pre required file: 2.4

output file: `obs_all_'ymd'_'ymd'.nc`

4. Homogenize :

4.1 File Name: estimate_tobs.py

Pre required file: 3.1 and 3.2

output file: tobs_estimates.nc

Special Note: It will take a while to execute (~2 hours)

4.2 File Name: adjust_time_of_obs.py

Pre required file: 3.2 and 4.1

output file: 'prcp_tobs_adj_'ymd'_'ymd'.nc

Special Note: It will take a while to execute (~2 hours)

4.3 File Name: homog_prdp_ref.R

Pre required file: 2.1 and 2.3

Output file: prcp_homog_ref_'ymd'_'ymd'.nc

Special Note: We try to create the temporal range (starting year to end year) higher than 60 years, but unable to create it due to some error. So far, we are using starting year as 1951 and ending year to 2021.

4.4. File Name: homog_prdp.R

Pre required file: 4.2 and 4.3

Output file: prcp_homog_'ymd'_'ymd'.nc

4.5. File Name: calc_obs_cnts_homog.py

Pre required file: 4.4

output file: prcp_homog_'ymd'_'ymd'.nc

4.6 File Name: set_time_of_obs.py

Pre required file: 4.4,4.1 and 2.4

Output file: obs_all_'ymd'_'ymd'.nc

4.7 File Name: set_duplicate_stations.py

Pre required file: 4.4

Output file: prcp_homog_'ymd'_'ymd'.nc

4.8 File Name: set_previous_next_obs.py

Pre required file: 4.4

Output file: prcp_homog_'ymd'_'ymd'.nc