**How to Download and Tabulate**

**The ERA5 Monthly Dataset**

February 21, 2020

**Data Overview**

The ERA5 gridded dataset is generated via “reanalysis”, which means it is an assimilation of “in-situ” weather observations and calculations performed by a weather model(s). The data is therefore somewhat synthetic, but presumably it is highly accurate in geographic areas that are rich in data, such as North America and Europe.

The data is global and is gridded at 0.25 degrees. The data is in hourly form, which means there are 24 observations per day at each grid point. This high level of detail results in a dataset that is over 5 petabytes, or 5 million gigabytes.

Fortunately, the data can be downloaded in user-specified subsets that are small enough to fit on a laptop or desktop. For example, you can download a single weather variable – as opposed to downloading the entire set of weather variables – and you can also the data in the form of monthly averages (as opposed to hourly observations).

**Downloading the Data**

A range of ERA5 data subsets can be downloaded from this URL:

<https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset>

For this exercise, I’m focusing on “ERA5-Land monthly averaged data from 1981 to present”, which can be downloaded here:

<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land-monthly-means?tab=form>

Via this URL, I selected “2m temperature”, which is the temperature 2 meters above the earth’s surface, and I selected every year from 1981 to 2019, and every month from January through December. Lastly, I selected “NetCDF” file format. The download took about 10 minutes, resulting in a 2 gigabyte data file. The default name for the data file was at least 50 characters long, so I renamed it “data\_era5\_monthly\_temperature.nc”.

**Extract a Subset of Data from the NetCDF File and Convert it to a CSV File**

I’ve provided R code that can open the NetCDF file and extract data from it, converting it to CSV format. To use the R code, set the following variables such that point to the relevant locations on your computer:

Location\_R\_Code <- "D:/JPW/data\_ERA5\_Monthly" # location of your R program

ERA5\_Data <- "D:/JPW/data\_ERA5\_Monthly/data\_era5\_monthly\_temperature.nc"

CSV\_Output\_File <<- "out\_era5.csv" # this output file is produced by this program

There are other variables that I’ve included in the R code which allow you to determine which “slice” of data you want to extract from the NetCDF file. For example, you can narrow the range of years to output to the CSV file, or you can focus on a rectangular area of the earth’s surface, as opposed to output data for the entire earth.

Note that while this dataset is supposed “Land” based, the data contains values over both land and sea. I’m not sure what to make of this. Perhaps this particular ERA5 data subset is mislabeled on the website?