

MAST 467/667: Introduction to Polar Oceanography (Fall 2021)
Andreas Muenchow (muenchow@udel.edu)

Sept.-29, 2021

Workshop/Homework-3: Advanced work flow via Command Line Interface

Data: Ocean Melts Greenland (OMG) at <https://omg.jpl.nasa.gov/portal/>

Introduction. For our third workshop we will prepare a large unknown number of Arctic Ocean Water Profiles from Greenland for graphical display in a “tidy format” where each variable (Longitude, Latitude, Depth, Temperature, Salinity, Profile-ID) has its own column for each row of data record.

Goal. Acquire command line skills to automate data download and processing.

Expectations. Same as last week.

Assignment. Add latitude and longitude as a column for every ocean profile record that you downloaded from executing the hyperlinked cURL file in https://omg.jpl.nasa.gov/portal/browse/products/urn:omg:OMG_Ocean_AXCTD_L1

The following steps may aid to accomplish this:

1_Create a list of the filenames that you downloaded from NASA’s site, recall that these files have extensions “.edf” and directory listings are done at the command line via “ls -la” (Apple computers) or “dir”, so on Windows you do

```
dir *.edf > list.txt
```

which redirects the output from the dir *.edf command to a new file called list.txt

2_Extract the filenames starting with “OMG” from this list and write them to another

```
gawk “$5 ~ \”OMG\” {print \”rename \”, $5, NR\”.dat\”}” list.txt >list.cmd
```

[creating an executable via an executable]

3_Execute the file “list.cmd” that you just created by typing “list.cmd” which automatically renames all files in your current directory to 1.dat, 2.dat, ... 234.dat etc. (the gawk parameter “NR” in \$2 above is the line number).

4_Copy all .dat files from your current directory into your data directory and then delete all the .dat files in your current directory.

[files and directories]

5_Create a new file called “NASA.cmd” that will contain the following loop to extract latitude and longitude, say, from all the files 1.dat, 2.dat, etc.

Inside NASA.cmd:

```
for /L %%i IN (1,1,9) do (  
    gawk "$1 ~ \"Lat\" {print $3,$4,$3+$4/60}' ..\Data\%%i.dat  
)
```

6_Create a new file called “pos.awk” that will contain the following lines(s)

Inside pos.awk:

```
{ $1 ~ "Lat" {print $3,$4,$3+$4/60 }
```

[awk shell scripting]

7_Rewrite NASA.cmd to read

Inside NASA.cmd:

```
for /L %%i IN (1,1,9) do (  
    gawk -f pos.awk ..\Data\%%i.dat  
)
```

[edit text, create/execute files/scripts]

8_Do you see that the construction using pos.awk executes identically to what you executed in #5 without pos.awk? If so, then we can built ourselves a “tidy” file that strips latitude and longitude from the header and appends it to each data line in the OMG data files. You can accomplish this with the following more advanced gawk syntax:

Inside pos.awk between { ... }:

```
if ($1 ~ "Lat") { y = $3+$4/60 }  
else if ($1 ~ "Lon") { x = $3+$4/60 }  
else if (NF == 9) { print x,y,$3,$4,$6 }
```

[advanced gawk scripting]

9_The output to screen must be annoying by now, but could you not redirect the output from all the .awk and .cmd scripting to a data file called output.dat using the redirect operators “>” used already in #1 and #2 (that overwrites the existing file of that name) or “>>” that will append to the file name should it already exist. Recall that are looping over 3 files, but you easily change that to 100 or 200 or all the OMG files that you have. Try it – by changing “NASA.cmd”

10_We are now almost ready to switch to R or MatLab or any other software you may know of to plot the data as maps, profiles, sections, or anything we desire. I will focus on RStudio next week and strongly encourage you all to install it for next week.