BEHR Project Log Summary – Summer 2019 in Review

Jun 3-7 – Started work for the summer; Figured out how to deal w/ multi-dimensional variables; Figured out efficient extraction of variables of any dimension from the nested swaths data structure in Matlab files; Learned how BEHR satellites scan and retrieve OMI pixel data; Read OMI and BEHR documentation.

Jun 10-14 – Figured out code for aligning pixels by geolocation; Continuing to read BEHR Documentation and OMI paper; Listed out and categorized the research papers recommended by Chi and Ron; Purchased the Machine Learning course from Udemy; Enrolled in Edx online ML course from MIT.

Jun 17-21 – Started the online ML courses from Udemy and Edx; Set up my workspace in Python; Began lab safety training; Created and had BRC cluster account approved so that I have access to Savio; Looked at BRC website and Chi's job script for tips to write my own bash script; Made good progress on the Edx ML course.

Jun 24-28 – Continued work on the Edx ML course; Began to learn bash scripting through tutorials; Talked to Kevin and Chi extensively on the details of how to write and submit job scripts on Slurm, creating my workflow on the cluster, setting up my code within bash scripts, and creating virtual environments; Discussed w/ Chi about data outputs of my job script and directory usage in the cluster.

July 1-3 – Learned how to use the argparse library through tutorials and documentation; Continued chugging through a good chunk of the Edx ML course.

July 8-12 – Copied over WRF-Chem and OMI files for the year 2014 into my scratch directory using the data transfer node; Separated my Jupyter notebook code into Python files; Set up the virtual environment in my home user directory, and after activation, downloaded important packages; Continued working on Edx ML course.

July 15-19 – Decided to lower priority on my Edx ML course and focus on creating my HDF file first (the output goal for my OMI and WRF-Chem pixel collocation); Finished setting up my PyCharm workflow environment; Created a requirements.txt file for my virtual environment; Tried installing Sublime Text 3 on the cluster; Reviewed argparse documentation and usage; Planned my approach to making my job script.

July 22-25 – Deemed Sublime Text 3 unnecessary for installation on my cluster account; Referred to Chi's bash scripts; Talked w/ Chi on what my HDF file would entail but slightly confused and finding it hard to visualize; Laptop crashed and froze for a few days, so was unable to work.

July 30-Aug 2 – Backed up my laptop and re-imaged it in order to resolve my laptop's software and freezing issues; Reviewed my written notes w/ Chi and began creating this project log summary to evaluate the next steps for my future work; Continued working on my bash script for collocation of BEHR and OMI pixels using my prior Python functions.

Aug 5-9 – Still somewhat confused on visualizing output HDF files for BEHR and OMI pixel collocation; Wrote pseudocode to plan out my bash script code; Notified by Chi that I do not need to focus on pixel collocation anymore, but can instead focus on using the WRF-Chem files themselves as the training dataset for my ML work; Read some articles on feature selection techniques; Reviewed my prior projects on classifiers and models I created for Data100.

Next steps: Planning to create data visualizations and plots to explore how WRF-Chem inputs relate to NO2 profile outputs; Working in Jupyter notebook for testing, creating data visualizations in seaborn, and utilizing the sci-kit learn library to create regression models.

General approach to fall semester 2019: Creating a machine learning model to directly replace WRF-CHEM model runs by training on the WRF-CHEM dataset, as well as adding variables mentioned by Chi as relevant to the model NO2 profile output. Might be working w/ a grad student or Josh Laughner on applying ML to this dataset and can hopefully gain some valuable insight from creating this ML model and using feature selection techniques. Also, going to investigate the different ML approaches I can use from research papers recommended by Chi and Ron.