# MSDS Capstone Project Update #1

Overall, ran into multiple issues that set me back more than I would have liked. The good news is that all the issues I’ll describe below, all are fixed, and I am now able to start automating data processing. I’m hoping that I will have most of the data processing/cleaning done this upcoming week and can start working on the neural network.

The first issue was with viewing and extracting the data. This issue was that the UCAR/NCAR site has many different subset files of the NARR, so trying to figure out which one I needed took some time. The types were constant fields of NARR, original NCEP NARR 3-hourly files, regrouped 3-hourly files, NCEP’s NARR monthly means, NCEP’s NARR long term climatologies and Special NARR runs for 2004 NAME. Once the correct file was found (Original NCEP), it was a matter of trying to extract the data. I tried to do this entirely in Python by using the PyGrib library. However, when viewing the variables, there were some obvious ones missing or they were labeled as unknowns. It was then I remembered about Grid Analysis and Display System (GrADS) <http://cola.gmu.edu/grads/>. This program was introduced to me at College of DuPage when the meteorology department was switching to use it for all their weather model pages on their website. The scripting language of the program is fairly easy to use once you figure out all the built-in functionality.

As always when I try to install GrADS, I run into issues. The precompiled binaries were not running correctly on Linux Virtual machines. I attempted to use OpenGrads (<http://opengrads.org/>), a modified version of GrADS. I struggled with that initially as well. I was able to get GrADS to work on my Mac, however I wanted to get it to work on the VM so that I could be continuously running scripts even when I wasn’t around. It was on yet another iteration of installing CentOS on a VM that I finally got OpenGrADS to run correctly on my VM.

The next issue was getting the HDD where I store everything but the OS on my PC to mount on the Linux System. VirtualBox was mounting the drive correctly, but my Linux user account couldn’t read or write to it. After a couple hours of struggling, I finally figured out the very simple fix. VirtualBox mounts have a user group of vboxsf. Once the Linux user account was added to it, it was able to read and write to the drive.

With GrADS installed and the drive mounted, it was time to write the scripts. I remembered seeing on OpenGrADS website that they have a Python library that integrates with the software. I looked into it but decided that in order to progress onward for this project it wasn’t worth the hassle. Instead I am using the os package in Python to call command line scripts to run GrADS via a GrADS scripting language script.

I have created a GitHub repository (<https://github.com/sburgholzer/MSDS-Capstone-Project>) where my scripts and Jupyter Notebooks will be uploaded each week as evidence of progress. I ran into issues where my initial repository branches somehow got mixed up and I couldn’t upload to it. I created a new repository and copied everything over to it to get around this issue.

I didn’t get as much progress as I wished by now due to these issues, work schedule being mixed up from my usual work schedule. Family events also put a slow down on my progress, as well as being sick. Even though I didn’t get as much done as I wished, I am glad I figured out essentially all my scripts and will be able to start automating this process. My hope is by the end of next week to start the actual machine learning step.

In addition, in order to get the data, I have to request small batches from NCAR/UCAR and that takes time for their system to process. They also expire files, and my first batch expired before I was able to download it all, so I will have to do that again.