

Integration & Validation of a Standardized Library & File Format for PyDARNMusic & DARNtids: Migrating From Legacy Pickle Files to HDF5 & Implementing a Comprehensive Testing Suite

Nicholas J. Guerra¹, James P. Fox¹, Thomas J. Pisano¹, Michael Molzen¹, Nathaniel A. Frissell¹, Paul M. Jackowitz¹, Joseph B. H. Baker², and J. Michael Ruohoniemi³
¹University of Scranton, ²Virginia Tech, ³Virginia Polytechnic Institute and State University

Introduction

- DARNtids (<https://github.com/w2naf/DARNtids>) is a python-based software package that detects and characterizes MSTID activity in SuperDARN observations. It was originally developed to produce the results for Frissell et al. (2016). An architecture for the library has been mapped.
- DARNtids first detects the levels of MSTID activity in the SuperDARN data and then automatically runs the PyDARNMusic (<https://github.com/hamSCI/pydarnmusic>) analysis software on event periods with significant MSTID activity.
- The Multiple Signal Classification (MUSIC) algorithm in PyDARNMusic calculates MSTID period, wavelength, speed, and propagation direction.

Conversion of Pickle Files to HDF5

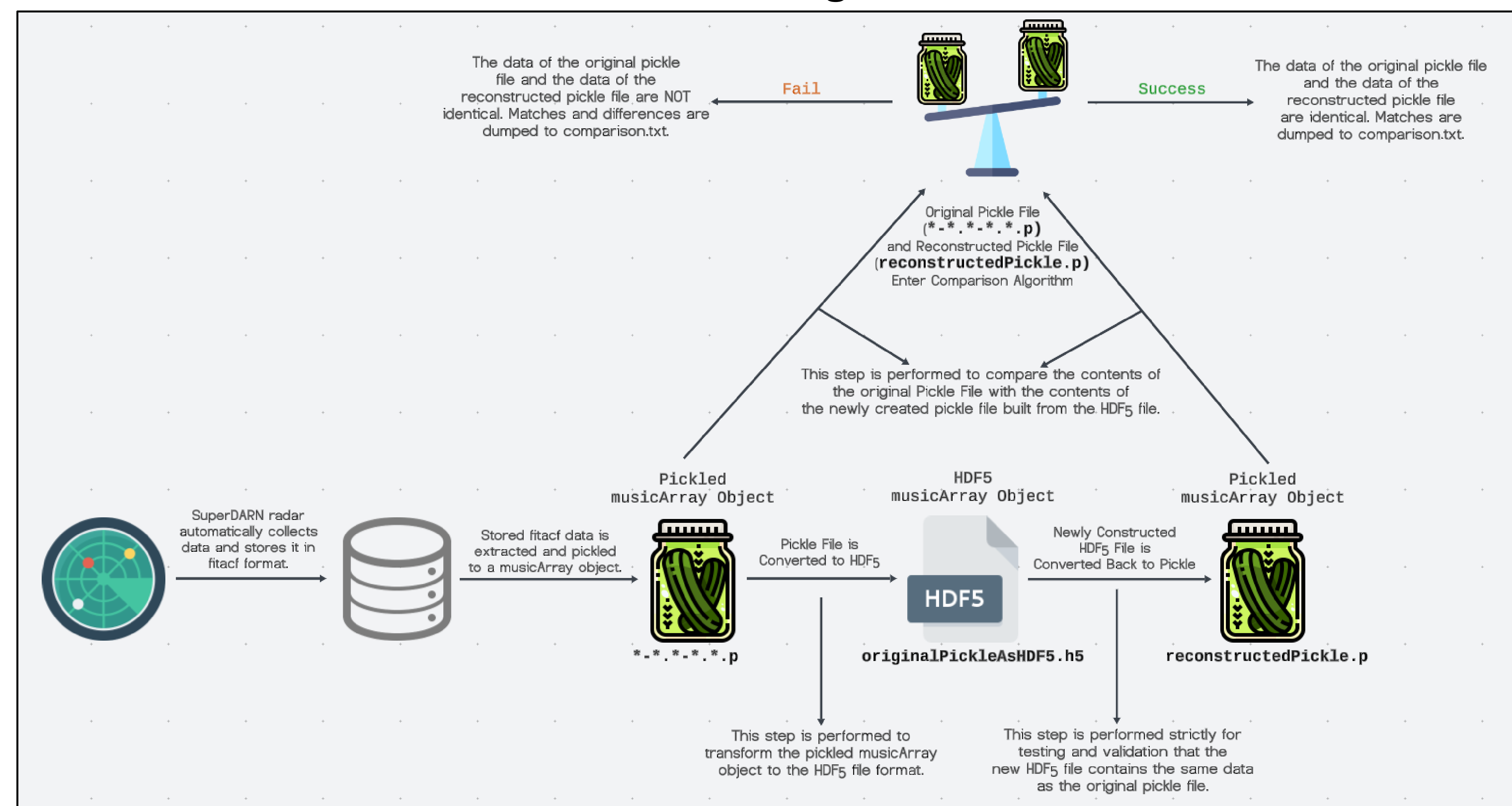
- Both PyDARNMUSIC and DARNtids use pickle files for storing Python objects as byte streams.
- Pickle files are highly undesirable for our use case because they are fragile, i.e., the pickle files generated by Dr. Frissell during his PhD cannot be easily opened, used, or modified on modern systems. Additionally, pickle files are not optimal for storage because they a.) are specific to Python and lack portability, b.) save massive, serialized values to disk, c.) are slower than alternative options, and d.) are insecure.
- HDF5 was created for representing and managing massive, highly complex datasets and is composed of a data model, file format, and software library accompanied by a series of features for optimizing time and space complexity.

- HDF5 has several benefits over Pickle, including A.) portability and language independence, B.) optimization for storing large, complex datasets efficiently using compression, and C.) structure that organizes data hierarchically into groups and datasets.

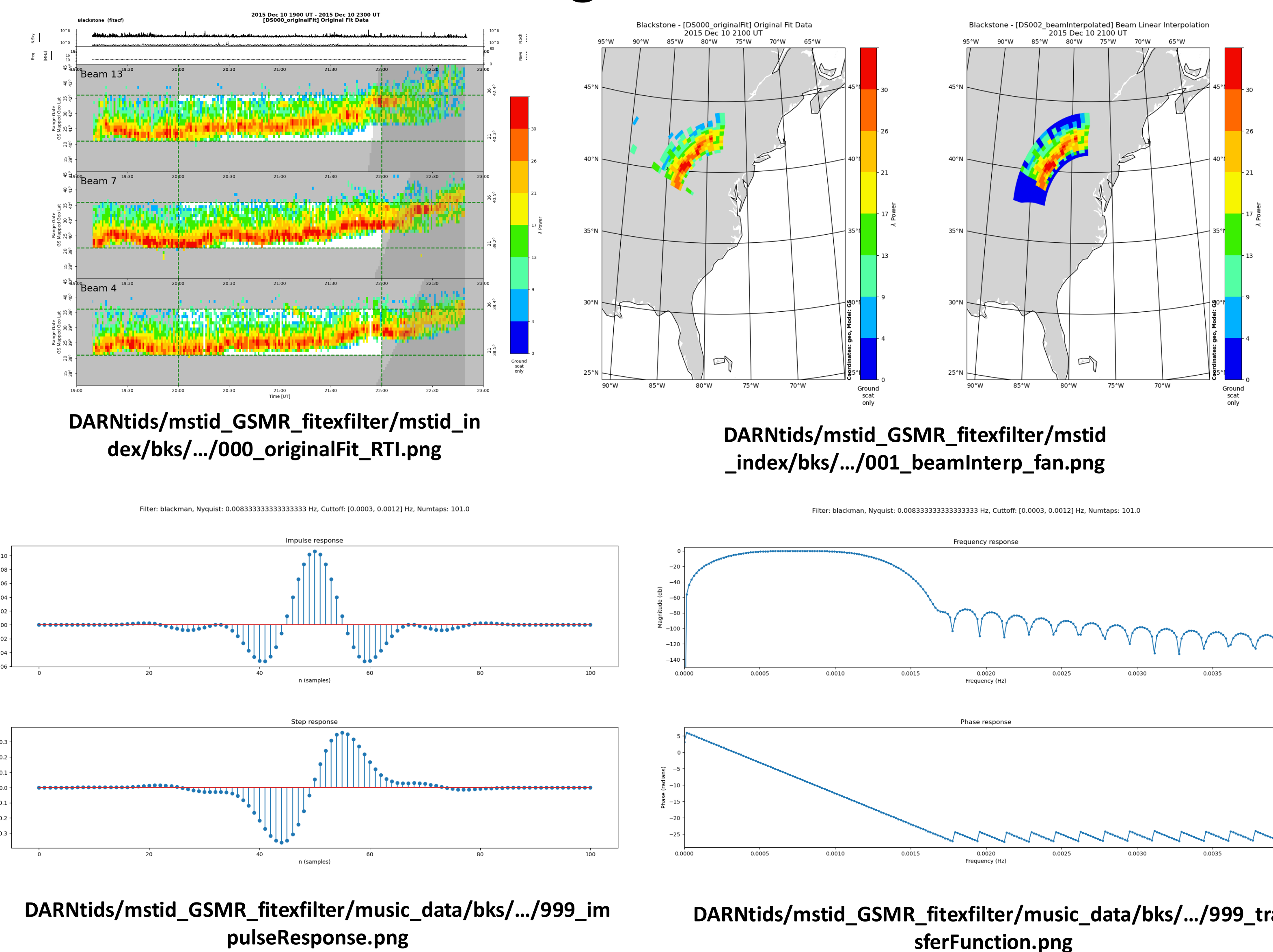
Summary & Future Work

- Currently, the Pickle to HDF5 Conversion Algorithm, HDF5 to Reconstructed Pickle File Algorithm, and Original Pickle File and Reconstructed Pickle File Comparison Algorithm is complete.
- At this point, I am replacing the dependencies on Pickle files using the Pickle to HDF5 Conversion Algorithm, deploying the changes to the Github repositories, and writing a comprehensive Thesis report.

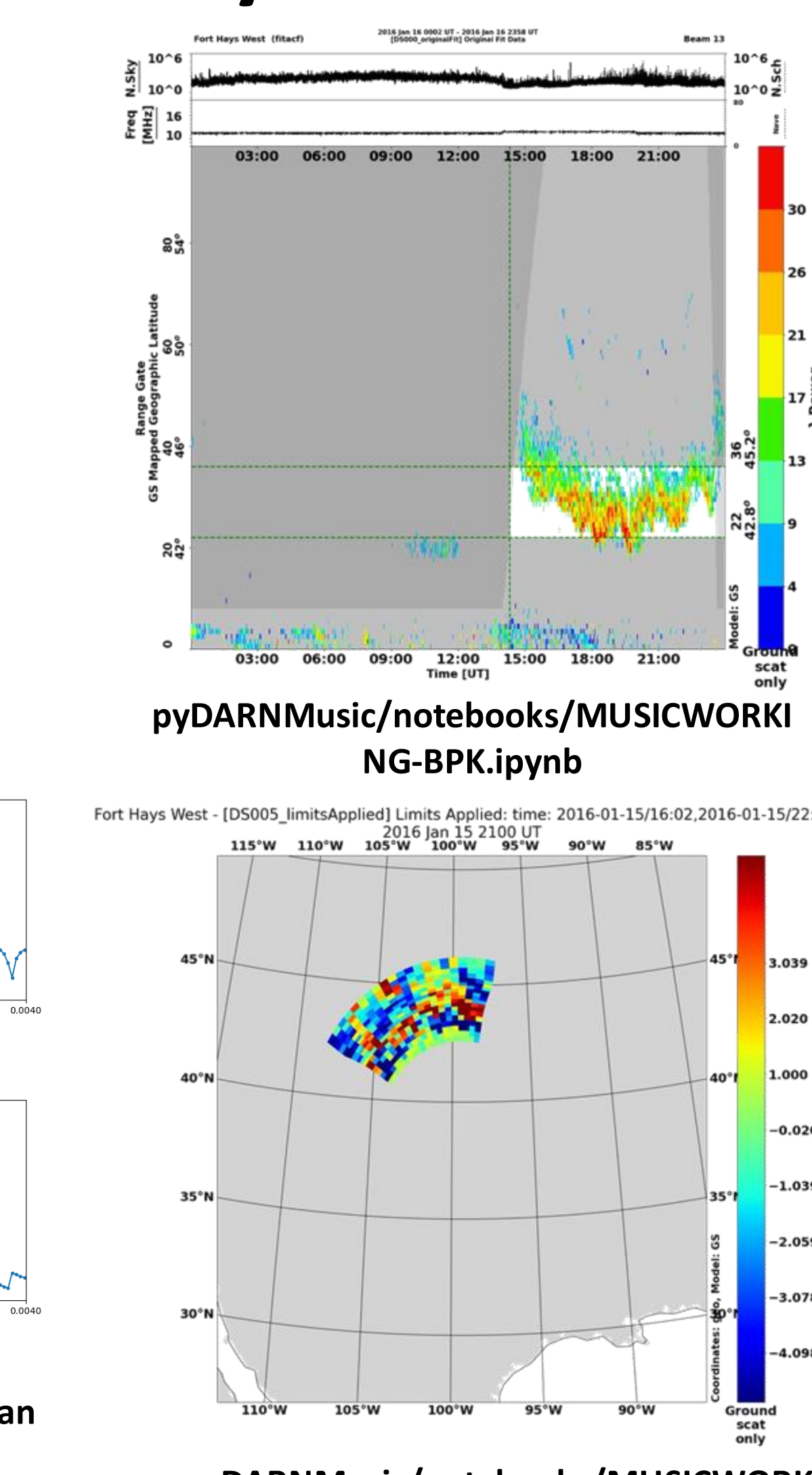
Data Flow Diagram



Plotting in DARNtids

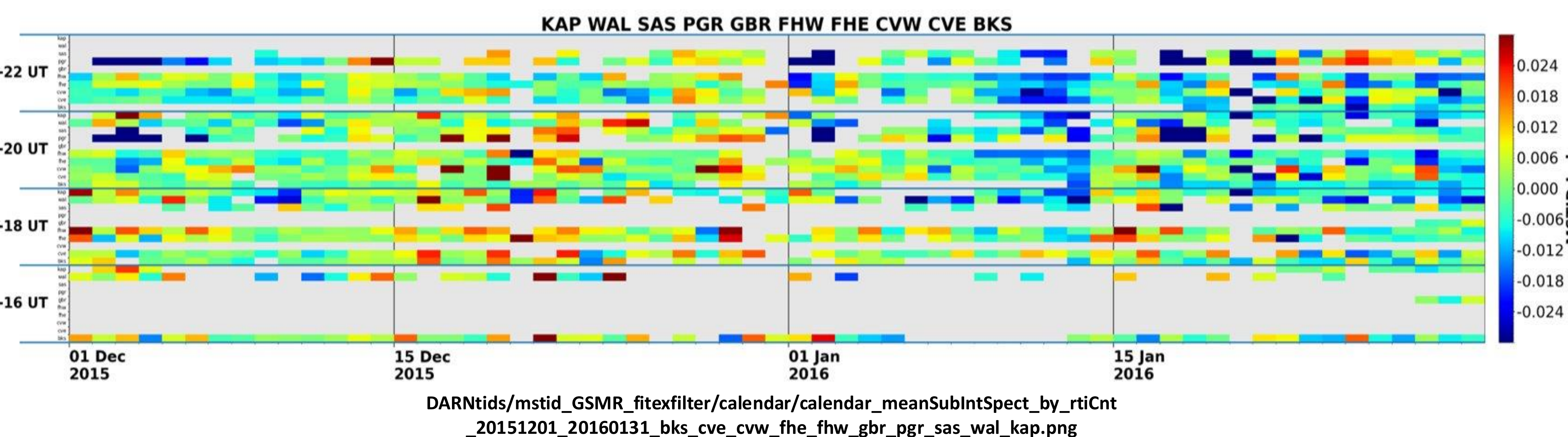


Plotting in PyDARNMusic



“Calendar” Plots

- “Calendar” plots are a time-series visualization summarizing the results of a DARNtids run.
- This calendar plot shows the MSTID index from 1 December 2015 – 31 January 2016 for the Kapuskasing, Wallops Island, Saskatoon, Prince George, Goose Bay, Fort Hays West, Fort Hays East, Christmas Valley West, Christmas Valley East, and Blackstone North American SuperDARN radars.



References

Frissell, N. A., J. B. H. Baker, J. M. Ruohoniemi, R. A. Greenwald, A. J. Gerrard, E. S. Miller, and M. L. West (2016), Sources and characteristics of medium-scale traveling ionospheric disturbances observed by high-frequency radars in the North American sector, *J. Geophys. Res. Space Physics*, 121, 3722–3739, doi:10.1002/2015JA022168.

Acknowledgments

This work is supported by NASA Grant 80NSSC23K0848 and NSF Grant AGS-2045755. We acknowledge the use of SuperDARN data. SuperDARN is a network of radars funded by national scientific funding agencies of Australia, Canada, China, France, Italy, Japan, Norway, South Africa, the United Kingdom, and the United States of America. Plotting was done using DARNtids, PyDARN, and PyDARNio, as well as open-source tools including python, NumPy, and matplotlib.