December 2020 Eclipse Festival Analysis

Kristina Collins, David Casente, Joanna Elia, Marius Mereckis, David Meshnick

Case Western Reserve University

PRESENTED AT:

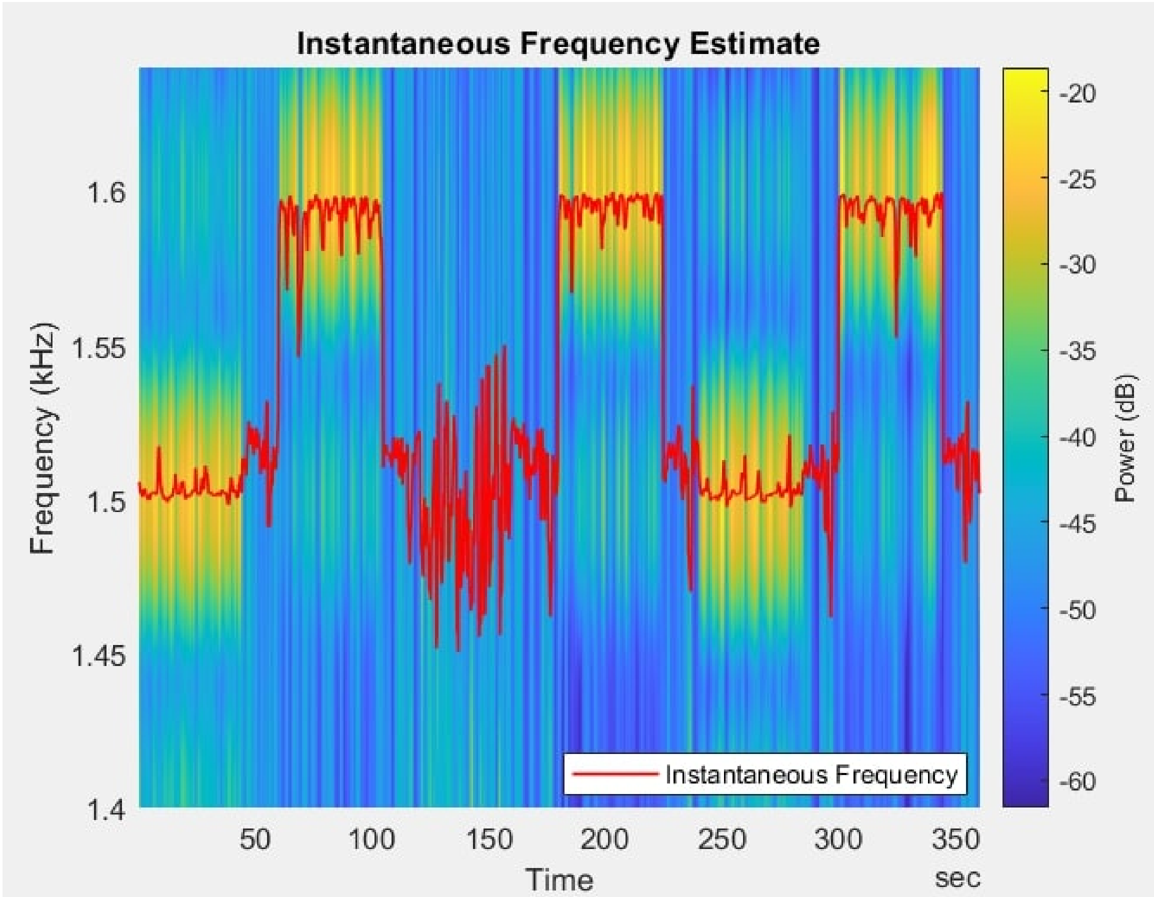
DIGITAL SIGNAL PROCESSING

There are various characteristics of the time standard station signals WWV and WWVH that can be analyzed using digital signal processing techniques to extract meaningful information. For one, the varying 440 Hz, 500 Hz, and 600 Hz minute tones can be used to actually distinguish between WWV and WWVH by comparing the tone currently being heard to the known tone being played by each of the corresponding stations at that minute. This analysis can then be used to construct a beacon mapping that visualizes the switching of the time standard station being heard by each of the participant stations as the solar eclipse travels along its path of totality. Furthemore, the IRIG time code embedded within the 100 Hz subcarrier can be extracted to determine very precise time, making the former script far more robust.

different methods of analyzing WWV and WWVH

Signal Component	Pros	Cons
Carrier	Continuous; easy to filter, monitor; can get frequency estimation over long periods	Interference from other time standard stations, local sources. (The typical ham shack has a lot of 10 MHz noise.)
Second Ticks	Time of flight estimation	Difficulty in estimating time delay; requires position and 1 pps signal from GPSDO
440 minute tone	Distinguish WWV, WWVH more easily than second ticks	Only available intermittently; requires >1s data to get <1 Hz resolution
Audio Sideband	No interference; can distinguish between WWV and WWVH	Not all minutes; only 900 ms/sec
Digital Time Code	Pseudorandom signal for a matched filter. Can be used to automatically identify the time at which data was recorded.	Complicated.

An estimation of minute tone instantaneous frequencies over a 6 minute interval



PARALLELIZATION

The current data wrangling process has two tasks that require a innefficient amount of time to complete: processing the audio data from its collection format into an analyzable format and data import/export. We are developing two seperate data wrangling processes to decrease the speed required for such tasks.

The first Data Processing pipeline that is in development uses matlab parallelization. The process takes the audio file, parses formatting variables from the metadata and the timestamp embedded in the file name, and converts the audio data into a timestamped table. In addition, the data processing script will also collate the audio files into the tables based on the recording location. From there the data is written to csv files, which can be imported by another matlab script for statistical analysis and visualization.

The specific technical implementation of this pipeline utilizes the parallelization functions built in to matlab to generate multiple workers to process the data at the same time. In addition, the process also runs expensive processing functions on CUDA cores in the GPU to further speed up the conversion process.

The second data processing pipeline utilizes hadoop and apache spark for parallelization as well as data i/o. The process will use matlab for converting the format of the data, but instead will use hadoop to create an apache spark database to hold the timestamped tables. From there the matlab scripts to conduct statistical analysis and visualization can grab the data from the database rather than importing it from a csv. This will not only speed up the time necessary for data wrangling, but also the time necessary for data import and export, which is another large time cost in the current data processing system.

Future Work: We plan on implementing the second pipeline sometime soon. In addition, we are extremely interested in not only developing documentation for reproducing our specific work, but also creating general documentation for how to implement parallelization into a data wrangling process in a high performance cluster computing environment. This could be extremely useful for future work in audio data or even for other professors looking to speed up their data processing within their research.

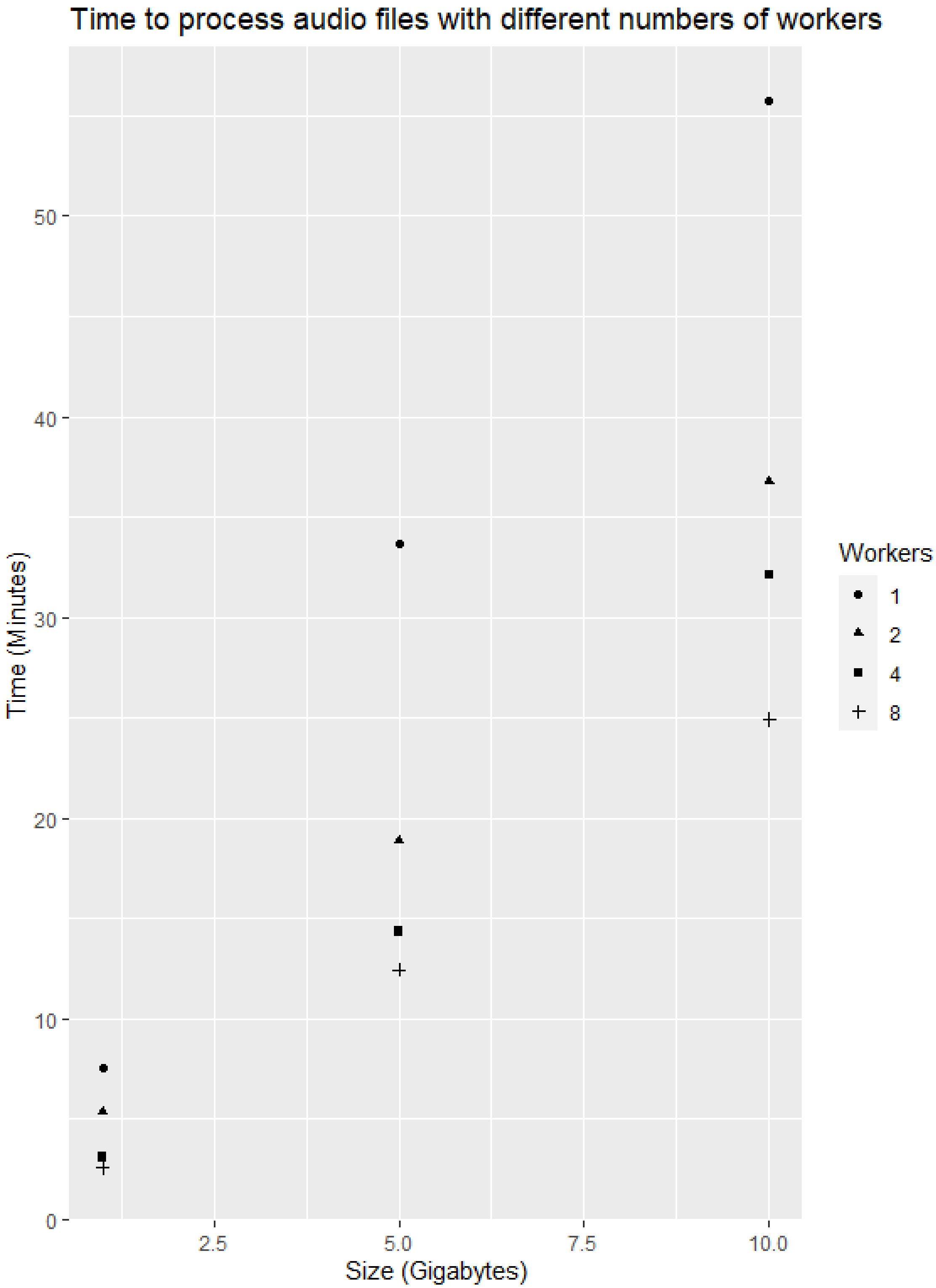


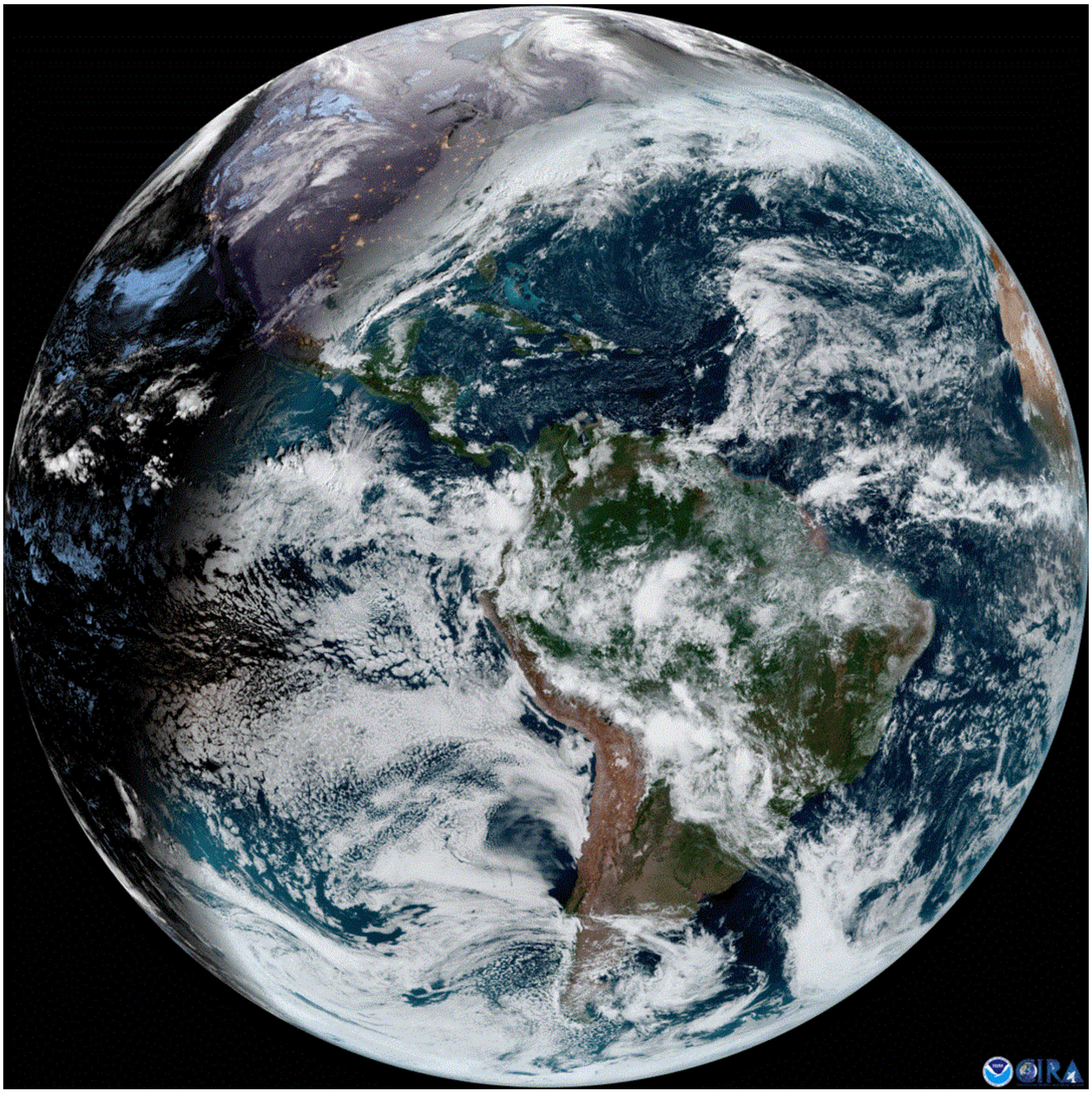
Figure 1: Breakdown of decrease in time for parallelization implementation. This data was collected on the first pipeline described above.

2020 ECLIPSE FESTIVALS

For complete information on the 2020 eclipse festivals, please refer to these links:

https://hamsci.org/june-2020-eclipse-festival-frequency-measurement (https://hamsci.org/june-2020-eclipse-festival-frequency-measurement)

https://hamsci.org/december-2020-eclipse-festival-frequency-measurement (https://hamsci.org/december-2020-eclipse-festival-frequency-measurement)

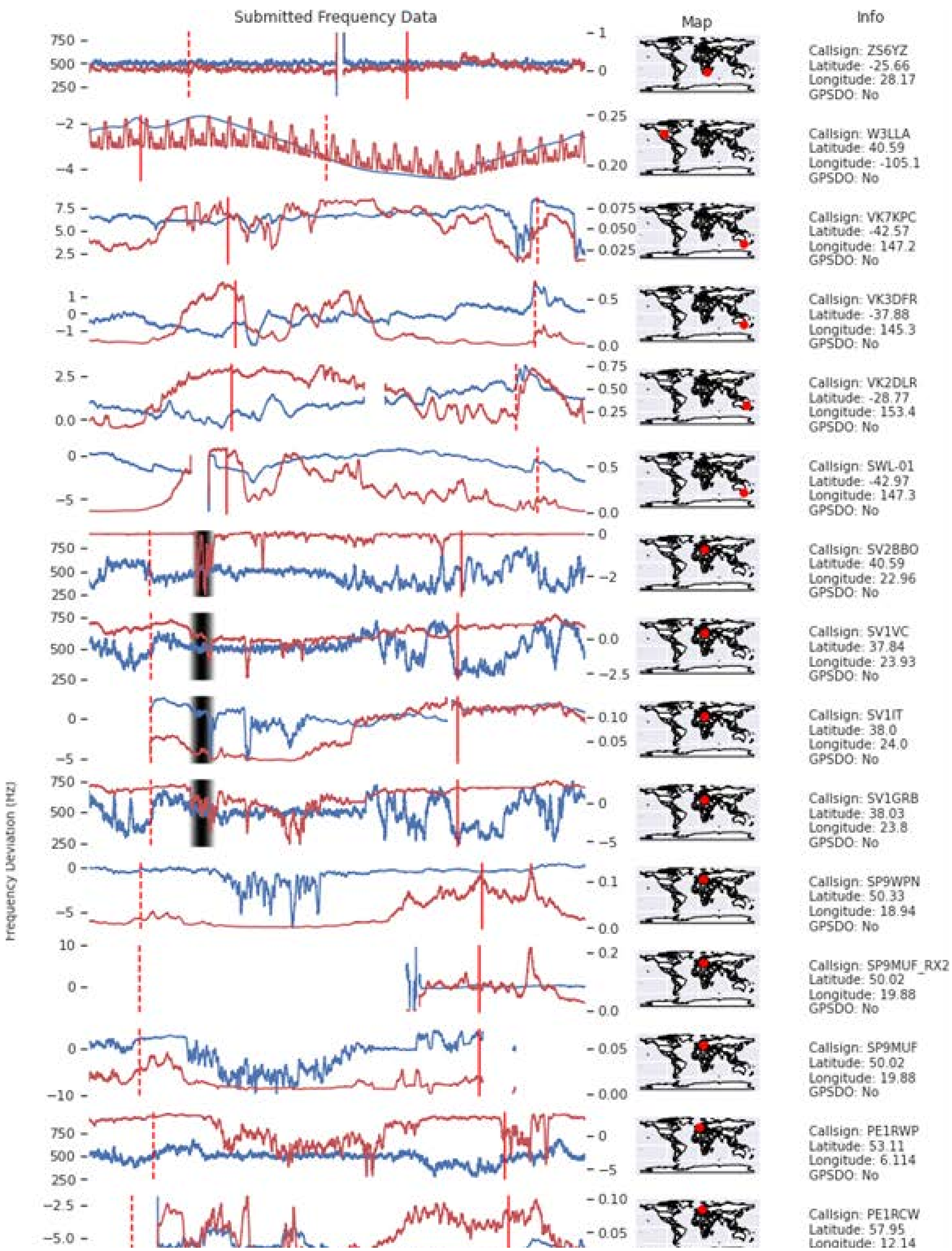


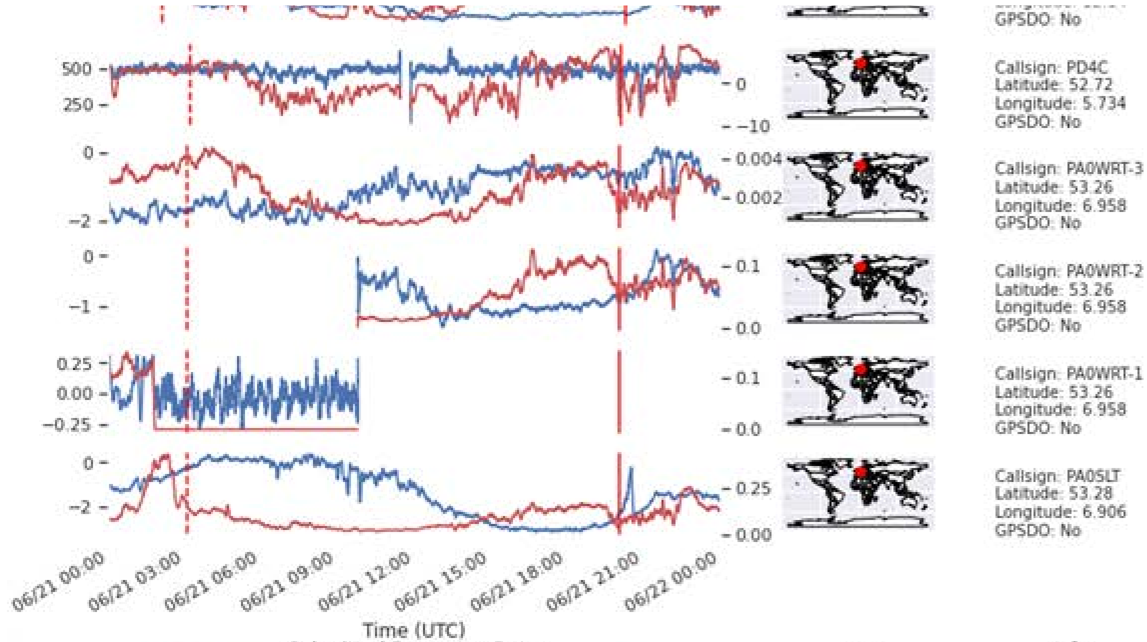
DATA VISUALIZATION

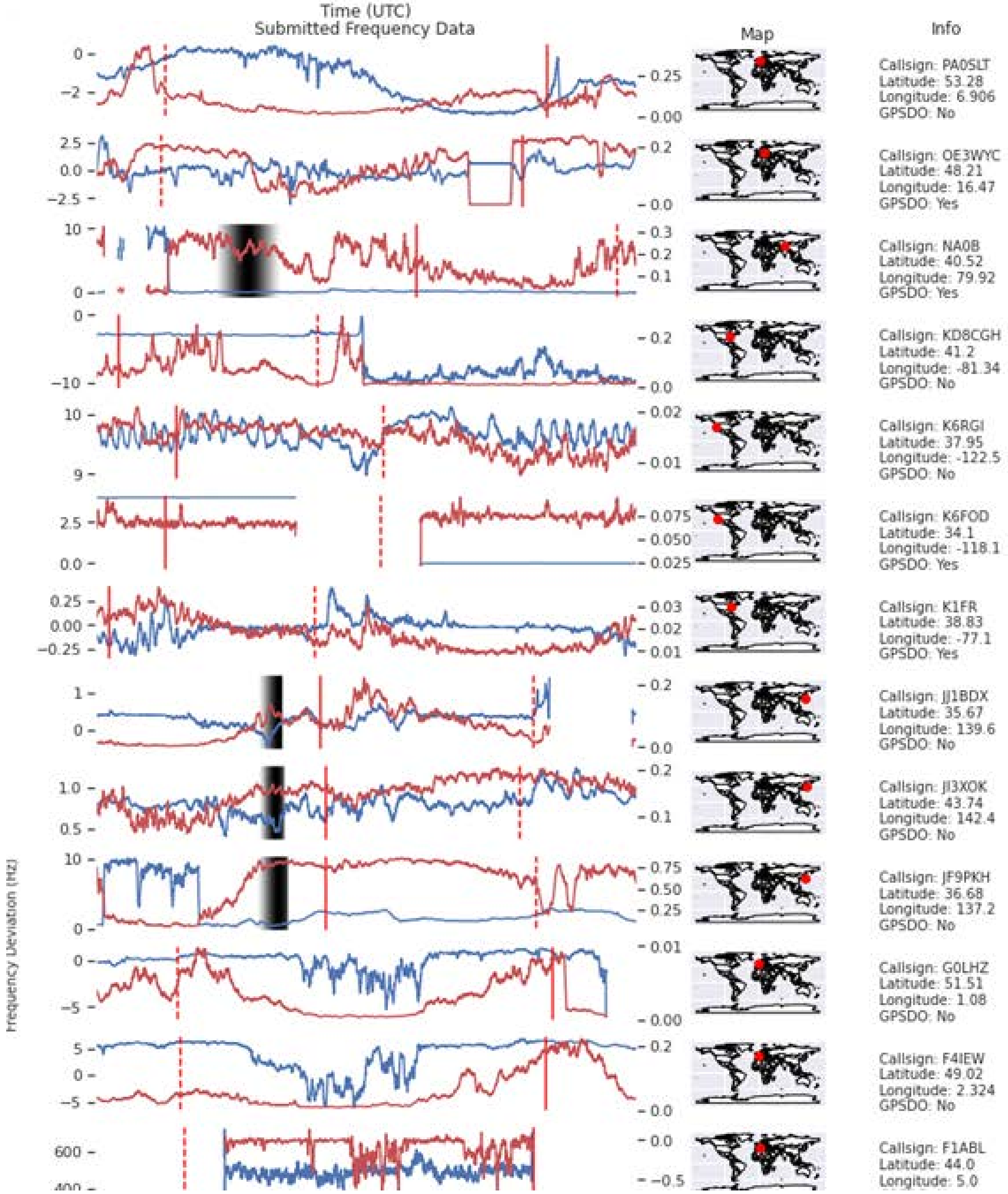
I created a shell script that extracts the metadata from the raw ".wav" files and places them in a ".csv" file. I was able to modify a preexisting git project to produce a GIF of the eclipse obscuration on a map for the December Eclipse.

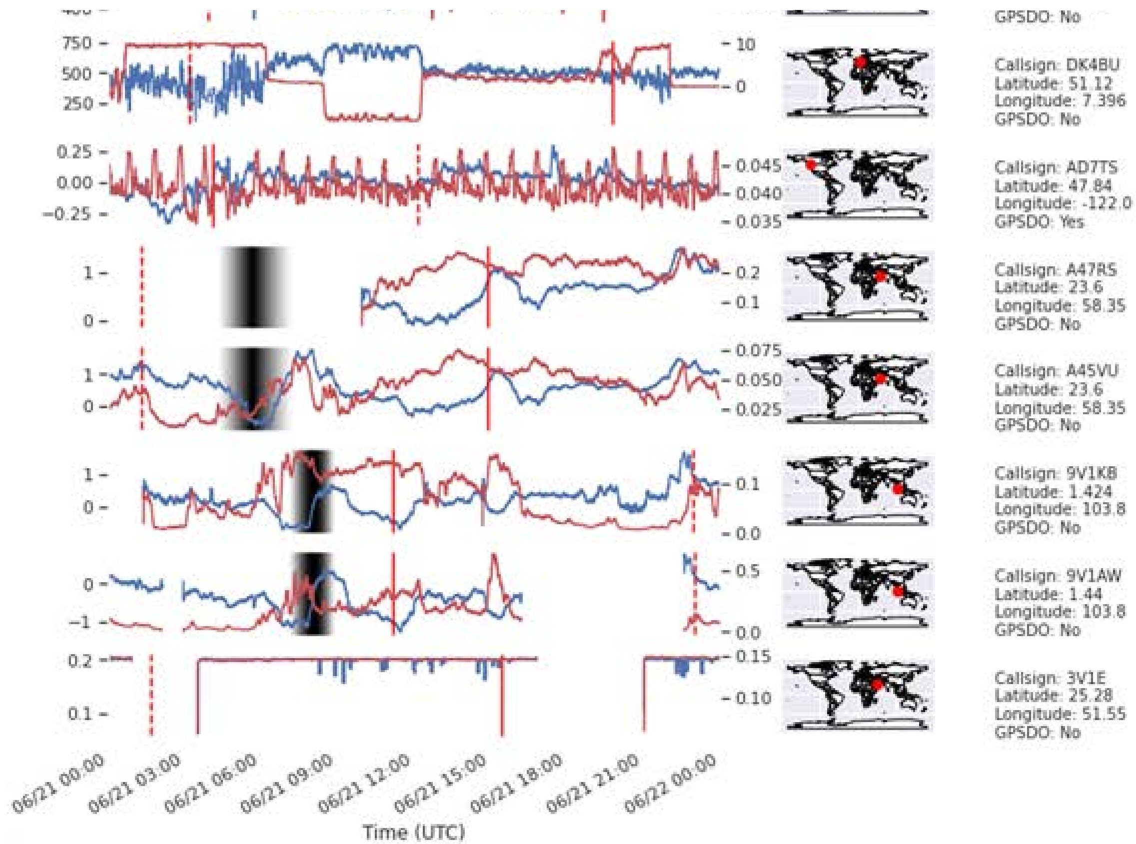
Visualization here. (https://youtu.be/uMwLXWKLnYo)

Using the June 2020 data set, I was able to plot the Frequency Error and Amplitude overlaid with shading indicating eclipse obscuration




GPSDO: No
Callsign: PD4C
Latitude: 52.72
Longitude: 5.734
GPSDO: No
Callsign: PA0WRT-3
Latitude: 53.26
Longitude: 6.958
GPSDO: No
Callsign: PA0WRT-2
Latitude: 53.26
Longitude: 6.958
GPSDO: No
Callsign: PA0WRT-1
Latitude: 53.26
Longitude: 6.958
GPSDO: No
Callsign: PA0SLT
Latitude: 53.28
Longitude: 6.906
GPSDO: No
06/21 00:00
06/21 03:00
06/21 06:00
06/21 09:00
06/21 12:00
06/21 15:00
06/21 18:00
06/21 21:00
06/22 00:00
Time (UTC)


Submitted Frequency Data
Map
Info
Frequency Deviation (Hz)
Callsign: PA0SLT
Latitude: 53.28
Longitude: 6.906
GPSDO: No
Callsign: OE3WYC
Latitude: 48.21
Longitude: 16.47
GPSDO: Yes
Callsign: NA0B
Latitude: 40.52
Longitude: 79.92
GPSDO: Yes
Callsign: KD8CGH
Latitude: 41.2
Longitude: -81.34
GPSDO: No
Callsign: K6RGI
Latitude: 37.95
Longitude: -122.5
GPSDO: No
Callsign: K6FOD
Latitude: 34.1
Longitude: -118.1
GPSDO: Yes
Callsign: K1FR
Latitude: 38.83
Longitude: -77.1
GPSDO: Yes
Callsign: JJ1BDX
Latitude: 35.67
Longitude: 139.6
GPSDO: No
Callsign: JI3XOK
Latitude: 43.74
Longitude: 142.4
GPSDO: No
Callsign: JF9PKH
Latitude: 36.68
Longitude: 137.2
GPSDO: No
Callsign: G0LHZ
Latitude: 51.51
Longitude: 1.08
GPSDO: No
Callsign: F4IEW
Latitude: 49.02
Longitude: 2.324
GPSDO: No
Callsign: F1ABL
Latitude: 44.0
Longitude: 5.0


GPSDO: No
Callsign: DK4BU
Latitude: 51.12
Longitude: 7.396
GPSDO: No
Callsign: AD7TS
Latitude: 47.84
Longitude: -122.0
GPSDO: Yes
Callsign: A47RS
Latitude: 23.6
Longitude: 58.35
GPSDO: No
Callsign: A45VU
Latitude: 23.6
Longitude: 58.35
GPSDO: No
Callsign: 9V1KB
Latitude: 1.424
Longitude: 103.8
GPSDO: No
Callsign: 9V1AW
Latitude: 1.44
Longitude: 103.8
GPSDO: No
Callsign: 3V1E
Latitude: 25.28
Longitude: 51.55
GPSDO: No
06/21 00:00
06/21 03:00
06/21 06:00
06/21 09:00
06/21 12:00
06/21 15:00
06/21 18:00
06/21 21:00
06/22 00:00
Time (UTC)

MACHINE LEARNING

So far, I have done preliminary research on artificial neural networks, working with time series data and how to integrate such a data format into an artificial neural network. It is important to understand what exact data science question I am tackling, what data is used as an input like geophysical data and a sliding time window, and what data is being predicted like the change in ionosphere height. I have begun to construct an artificial neural network but it does not work at this time.

Now that some of the June 2020 solar eclipse data has been processed, I will be able to begin training the artificial neural network and be able to start parsing the geophysical features to identify the significant and the noise. I hope to be able to begin exploring the differences between transfer learning vs. a custom architecture and make a choice of which to use.

ABSTRACT

A crowdsourced science experiment called the December 2020 Eclipse Festival of Frequency Measurement was carried out for the total solar eclipse across South America on December 14, 2020. Over 80 stations around the world recorded WAV files of 10 MHz time standard stations. We have undertaken to process and visualize this data, and identify geophysical features within it. This poster will summarize our work to date.