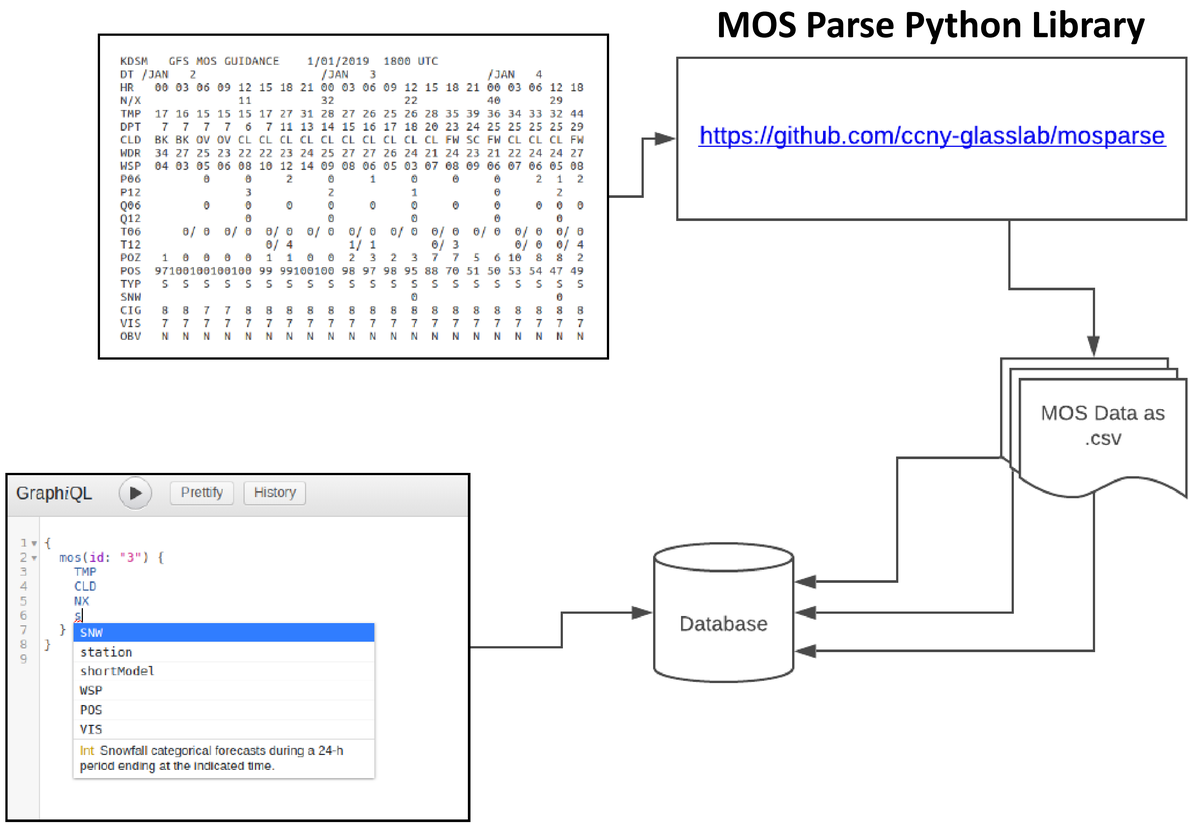
MOS Parse: Library for converting Big MOS Datasets to CSV Format

Oscar Lucero[1], Theo Schiminovich[1], Hannah Aizenman[2], Michael Grossberg[3]
[1]2019 High School Initiative in Remote Sensing of the Earth Systems Science & Engineering (HIRES) Scholar, [2]Department of Computer Science, The City College of New York and CUNY Graduate Center, [3]NOAA Center for Earth System Sciences and Remote Sensing Technologies and Department of Computer Science, The City College of New York

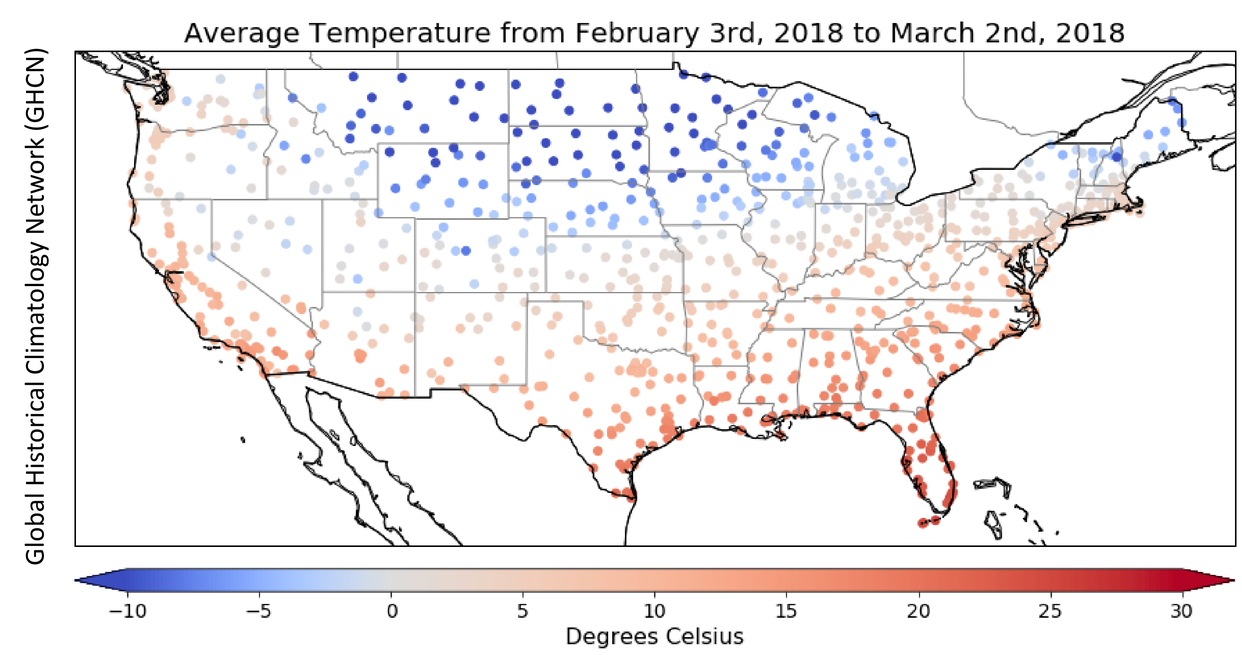
Motivation

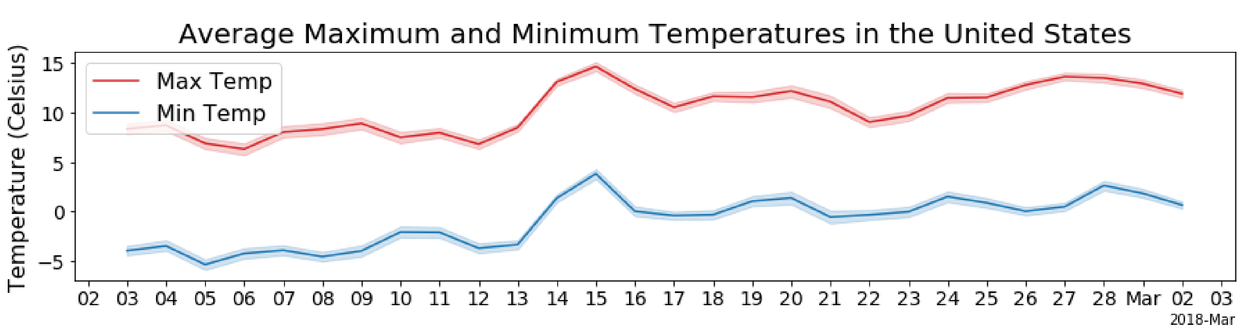
The National Weather Service provides Model Output Statistics, which is the source of most weather forecasts, in a text format that is easy for humans to read but difficult for programs to interpret. To facilitate large scale analysis, we reformatted model output to a format readable by nearly all data analysis programs. We then demonstrate using our library by comparing February 2018 continental United States GFS forecasts with ground truth observations from the Global Historical Climatology Network (GHCN).

MOS Parse Python Library

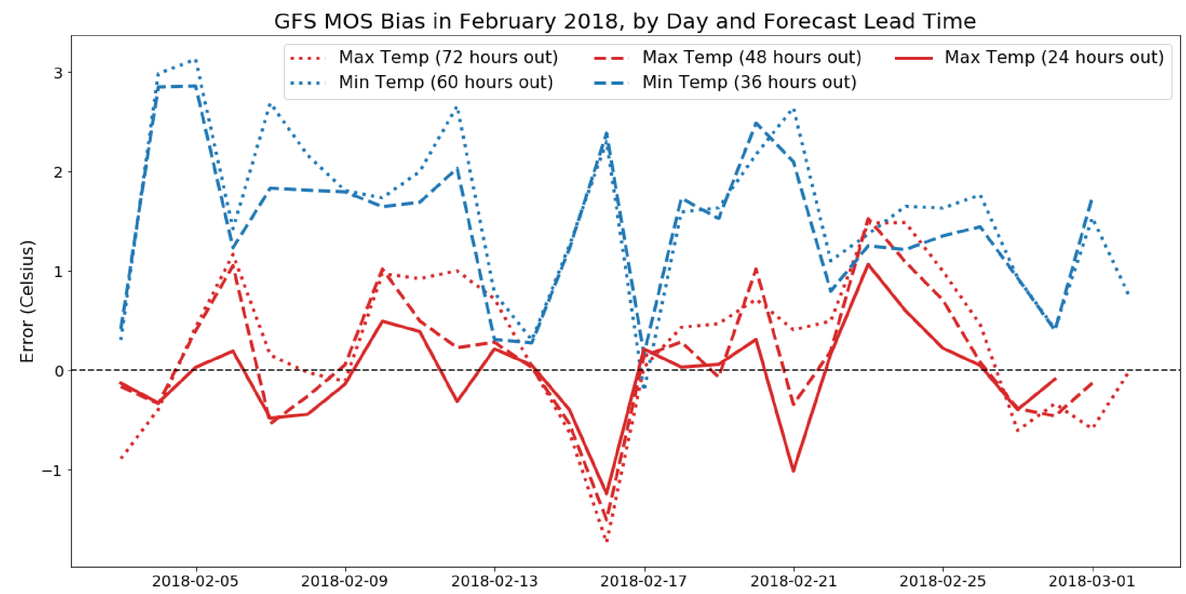


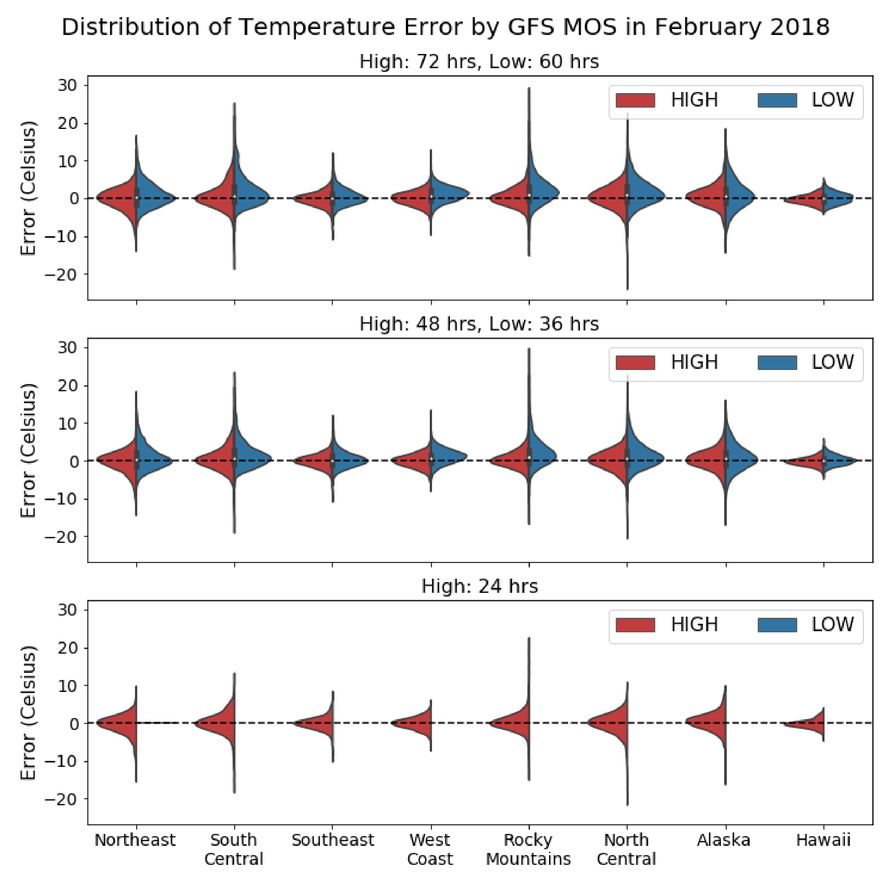
Stage of MOSParse Development	Python Libraries & Tools Used
Parse MOS station text messages into spreadsheets	Unlzw, Pandas, NumPy, Dask,
Test parsing and validate results	PyTest, Iowa Mesonet's MOS archives
Document code consistently	Sphinx, NumpyDoc,
Package code into an installable library	BNL NSLS-II's Scientific Python Cookie Cutter
Provide graphQL access (in progress)	Flask, SQLAlchemy, GeoAlchemy2, Graphene
Compare GHCN observations to MOS forecasts	Matplotlib, CartoPy, Geopandas, Seaborn, GeoPlot





Library Evaluation using MOS Bias Analysis





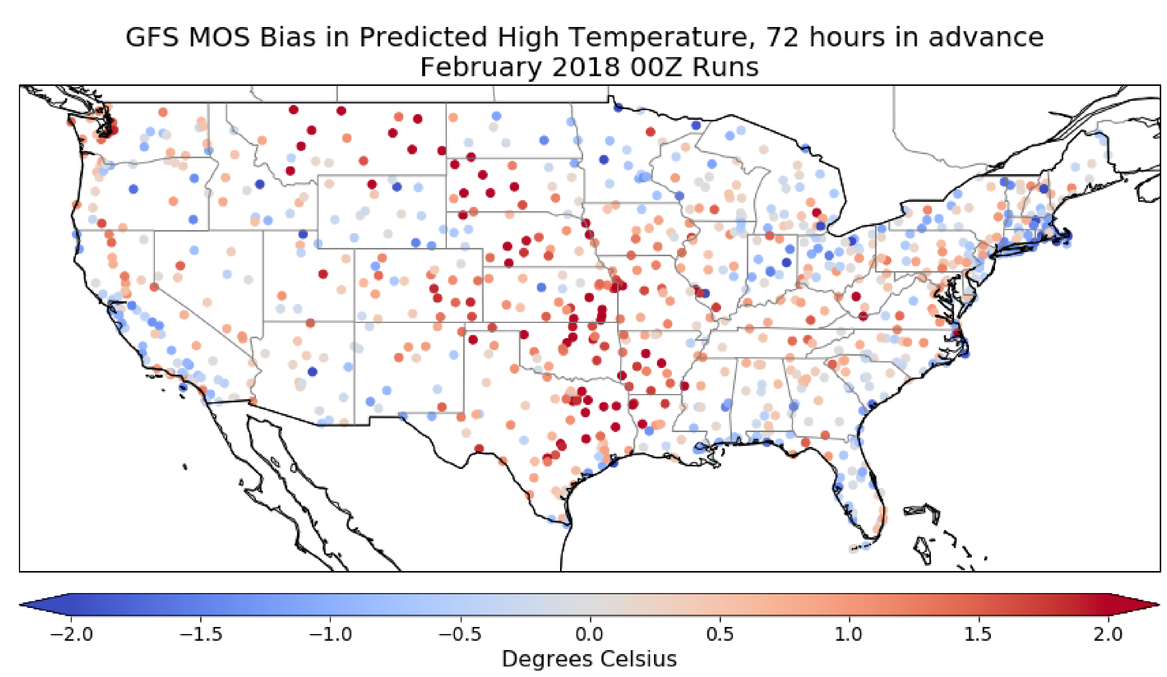
Key Takeaways

GFS often overestimates daily low temperatures. It does not have that tendency for high temperatures.

GFS typically gets more accurate as the forecast lead time decreases, but not by much.

GFS Bias is lower in Hawaii, the West Coast, and the Southeast than other locations, possibly because their temperatures are less variable.

The GFS underestimated high temperatures in New England, which had a warmer than normal February 2018.



Discussion

The GFS 72 hours ahead forecast tends to predict warmer high temperatures than observed in the middle of the country and lower high temperatures than observed in the coastal regions in this sample; further work could investigate if this trend holds true for a larger sample of the data.

American Museum of Natural History
The Pinkerton Foundation
CUNY CREST
CUNY CREST HIRES