

Gabriel Moraga

Laboratory for Atmospheric and Space Physics
 Dr. Greg Lucas¹, Maxine Hartnett¹, Matthew Bourque¹, Dr. Wendy Caronde²
¹ Laboratory for Atmospheric and Space Physics, Boulder, Colorado, United States
² Masar Technologies, Westminster, Colorado, United States
 Email: gamo6782@colorado.edu

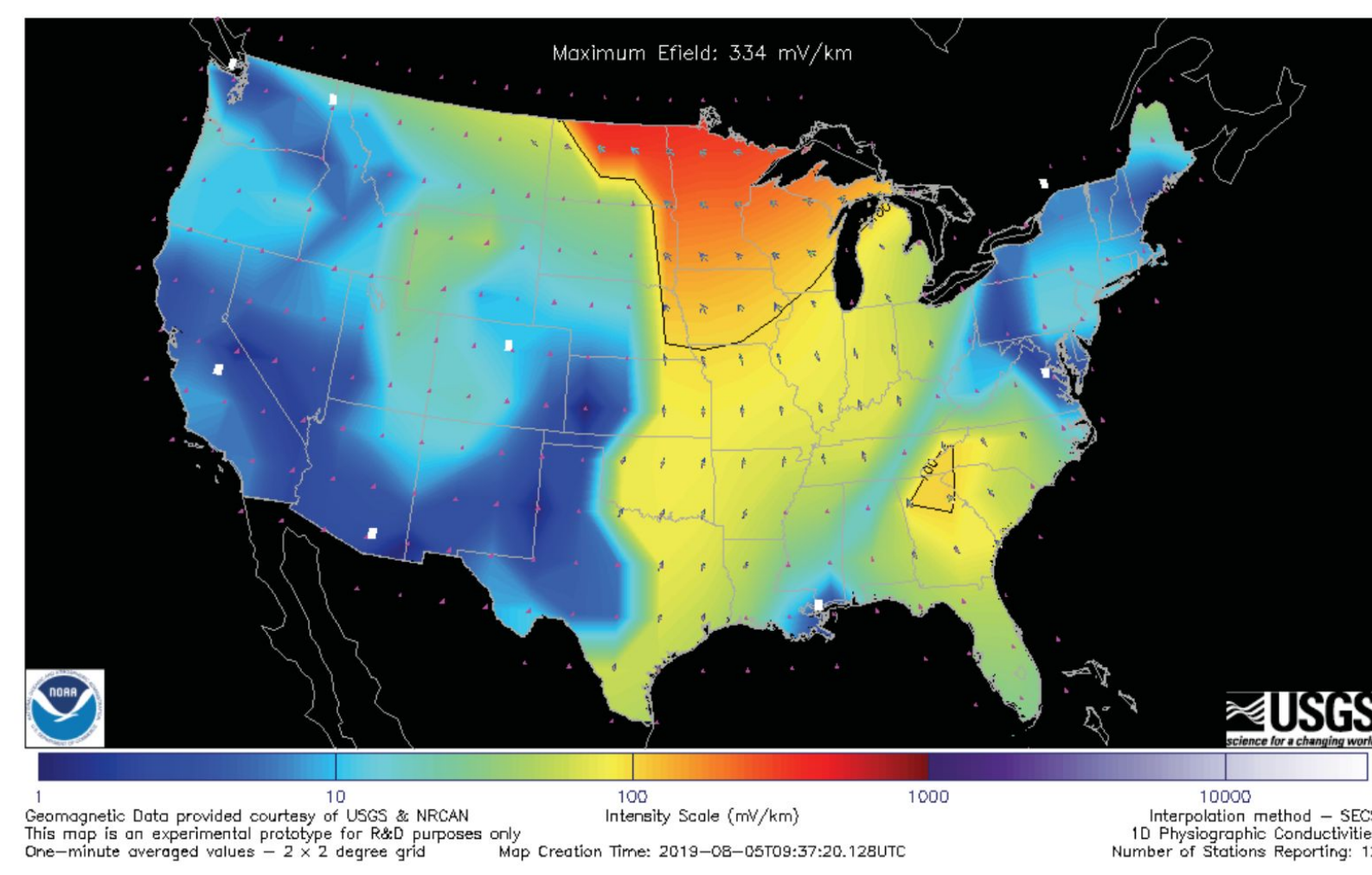
Abstract

Space weather, and its effects on Earth, is a trending topic. Geoelectric fields are influenced by geomagnetic field variation caused by Earth's surrounding space weather environment. Geoelectric hazards from space weather can critically affect infrastructure, and the aftermath is economically costly. It is essential to enhance geoelectric hazard modeling by implementing a combination of research and machine learning. This will determine where and when geomagnetically induced currents (GICs) may impact critical infrastructure.

Organizations such as NOAA, USGS, NASA, NSF and international partners (NRCAN) have contributed to an operational geoelectric field map data product. Currently, ground based stations across North America measure real-time surface geomagnetic field data, which is driven by natural-time ionosphere conductivities within the thermosphere. The current geoelectric field maps use the Spherical Elementary Current System (SECS) interpolation method, based on physical constraints of the magnetic field. Currently there is no machine learning method in place to fill missing data and flag outliers in real-time.

Our team created an algorithm in Python, using real-time data, which interpolates missing values and detects anomalies outside a standard deviation of 3 sigma. With the Seasonal Auto-Regressive Integrated Moving Average (SARIMA) anomaly detection model as the base, the newly created model will fill data gaps and flag anomalies. We generate more accurate geoelectric field maps using machine-learning algorithms that identify and replace missing data within the real-time data streams. The final machine learning model will be run in the CU's Space Weather Technology Research and Education Center's (TREC) cloud-based Testbed environment.

Geoelectrical Data and Space Weather



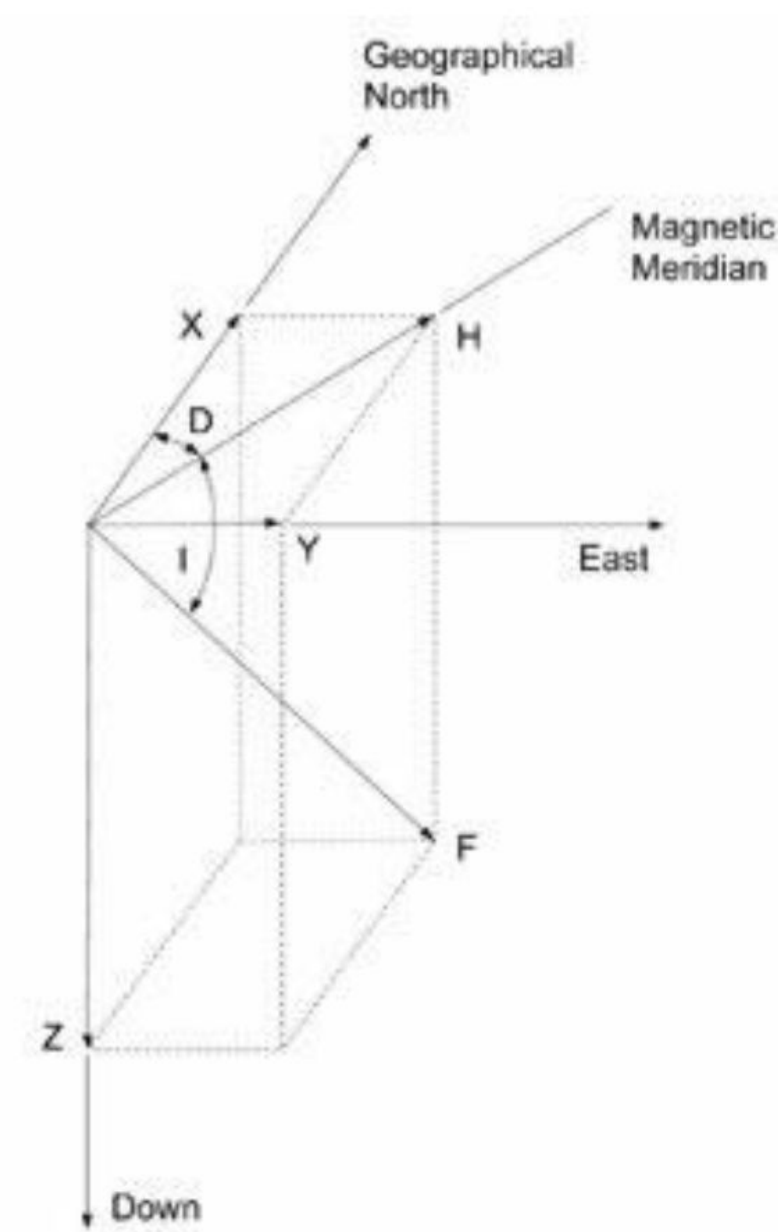
Historically space weather events have affected Earth. The Carrington Event, and the 1989 Quebec blackout that lasted 9 hours, take the popular vote. Humanity has built a world in which power grids are essential. Current research is trying to solidify a solution that will determine where and when geomagnetic induced currents (GIC's) may impact a society's infrastructure. There is a growing need to accurately predict and measure the geoelectric field data.

What you are looking at above is an early geoelectric field map, generated in 2019. It continues to evolve and be used for industry assessment and mitigation of geoelectric hazards. Upgrades in interpolation consisted of bilinear and nearest neighbor methods.

Zeiss Jena O10B Theodolite for making absolute measurements at Boulder magnetic observatory

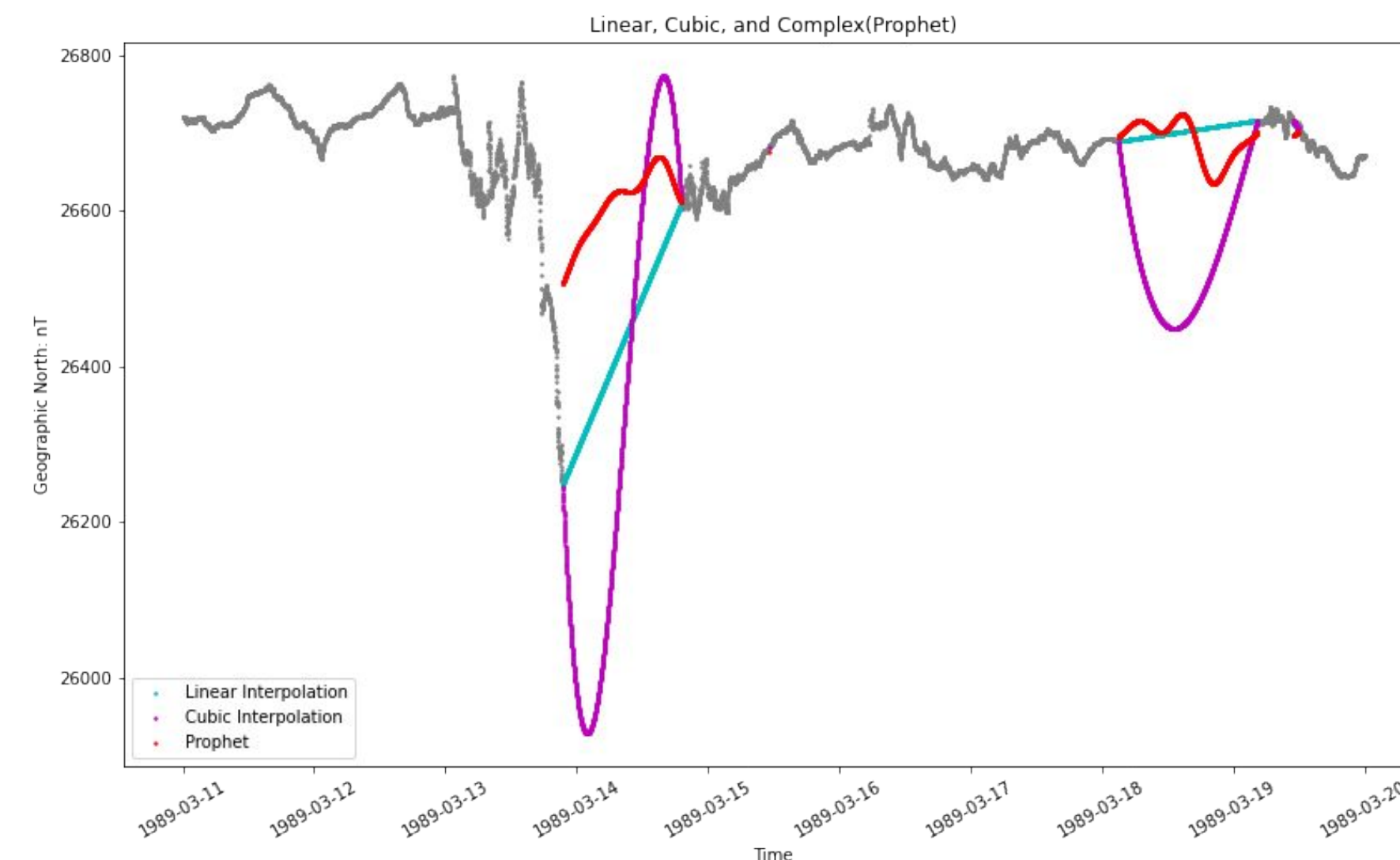


On the right, you will see the Magnetic-field components: (X,Y,Z) define the Cartesian components (north, east, down), (H,D,Z) are the components measured by instruments in every ground-based observatory.



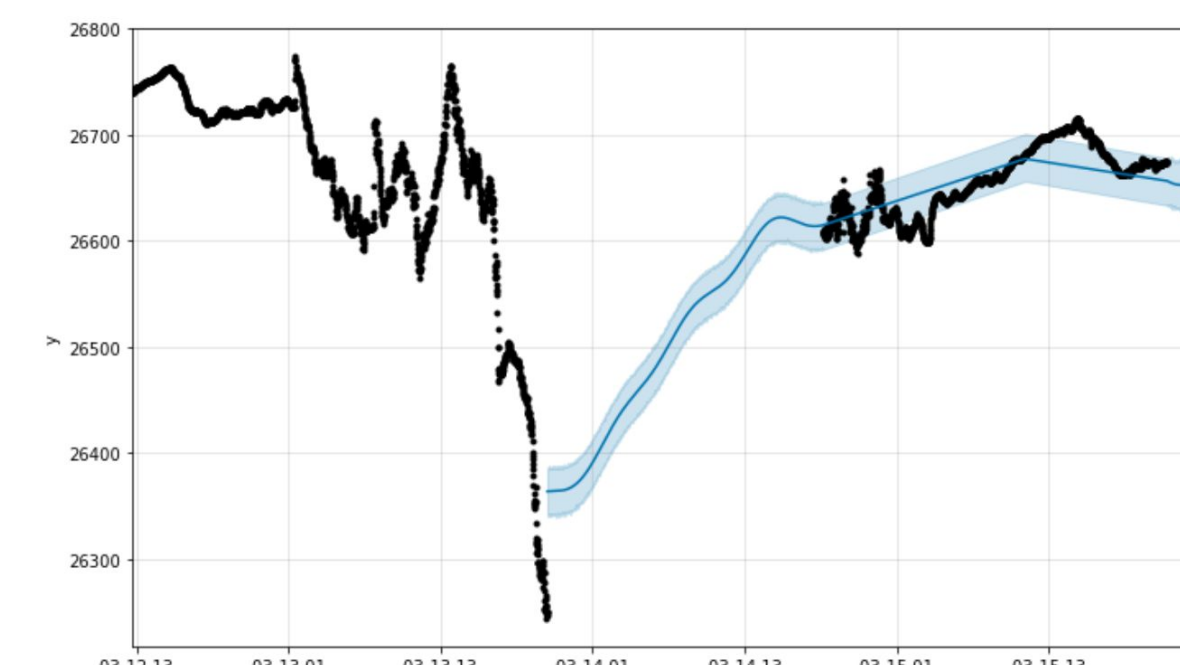
Statistical Analysis and Interpolation

Gap identification via NaN flagging and Interpolation: Linear, Cubic, and Complex.



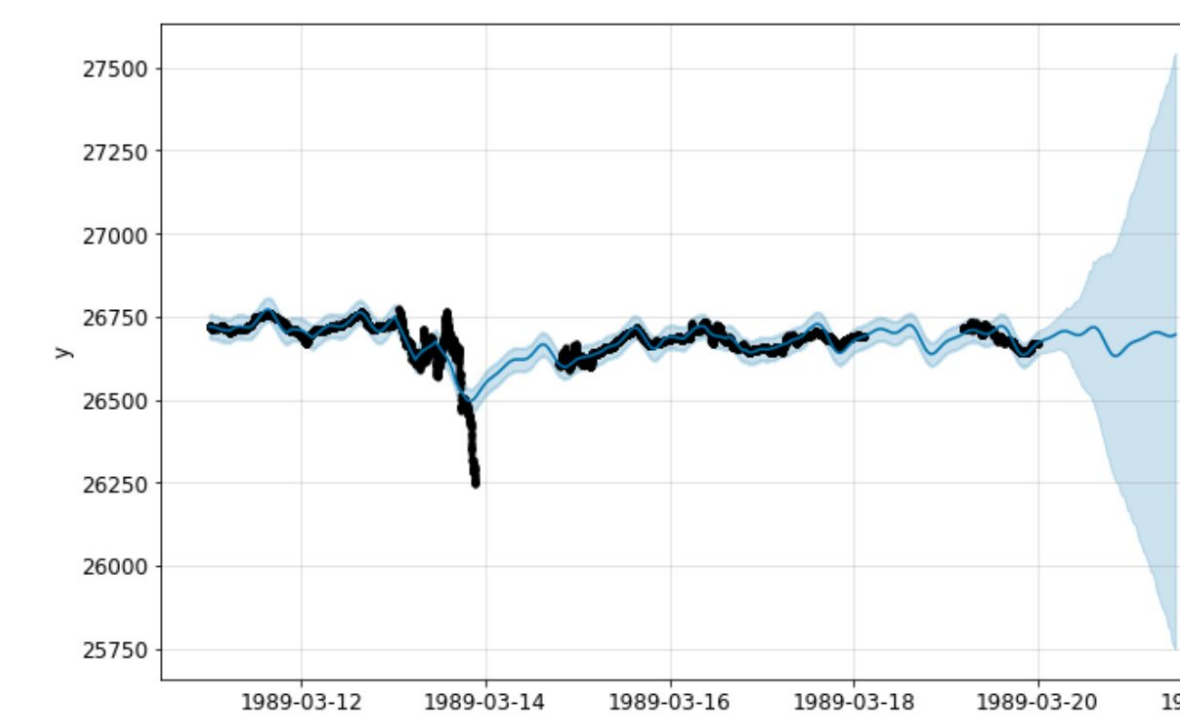
In this time-series data, there are gaps. This can be caused by local internet disturbances, local power outages, and even cars driving by. Finding the endpoints on either side of the data is important for interpolation. Numpy was used to find the NaN values. Linear interpolation is important because it can confirm the mean value of the data. Cubic interpolation is a more realistic method because it has continuous derivatives between points. However, finding an interpolation method that was more complex was essential.

Interpolation: Updated Python Package (Prophet)



Cubic interpolation produced unstable oscillation and instability when using big data. Using the python package Prophet to interpolate the NaN values proved to be effective. This interpolation method took historical values to build a non-linear additive fit between the NaN values. Prophet uses weekly, yearly and seasonality to fit the data. Prophet is similar to TensorFlow, in that it uses Fourier transform.

Model Testing: Interpolation and Prediction

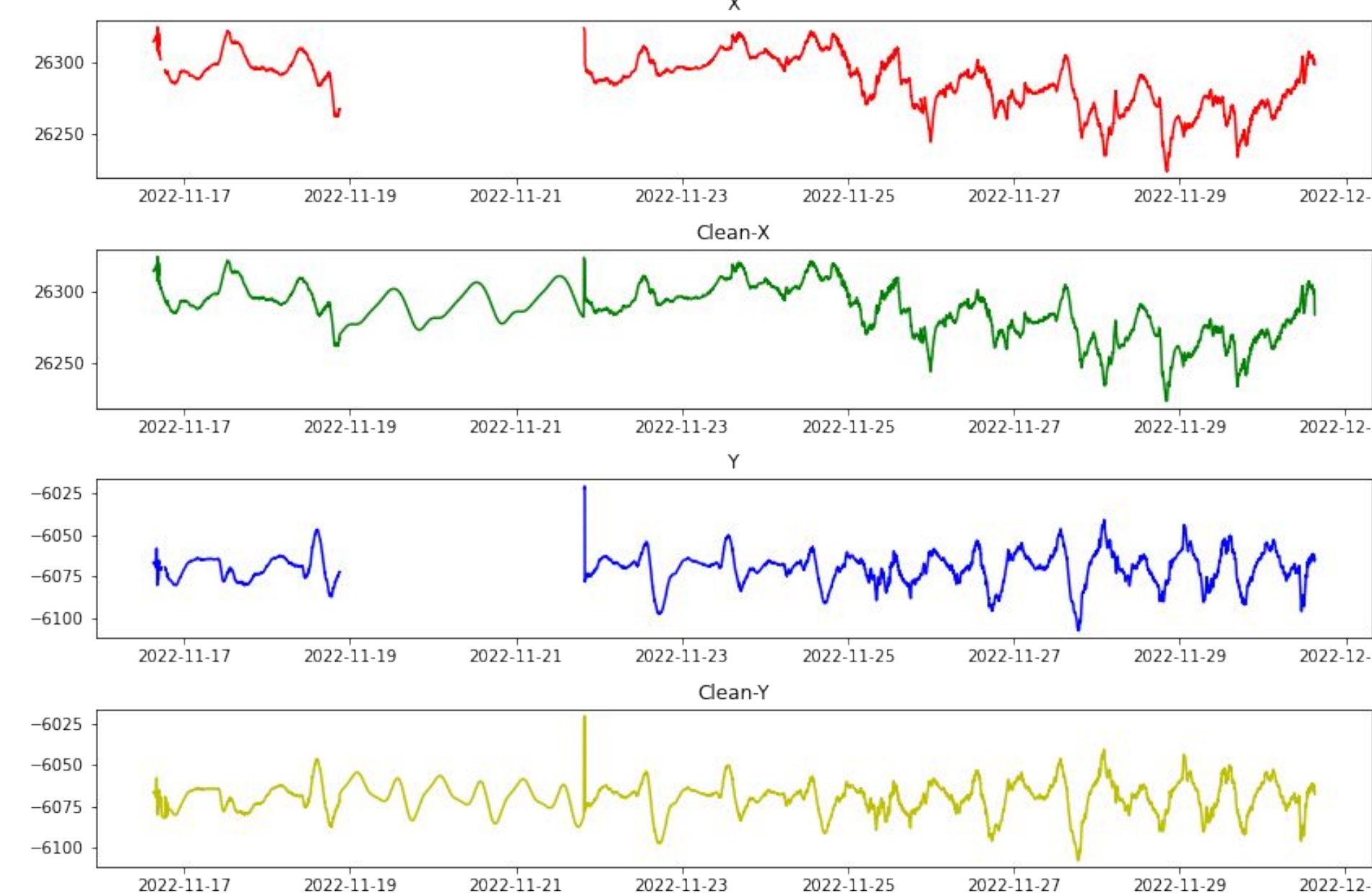


After the interpolation proved to be proficient, testing the model with different sample sizes was important to build a successful prediction model. The graph to the left shows a fit model that is interpolating values and predicts a forecast of 72 hours. This sample was one of many, that helped determine the historical to prediction ratio.

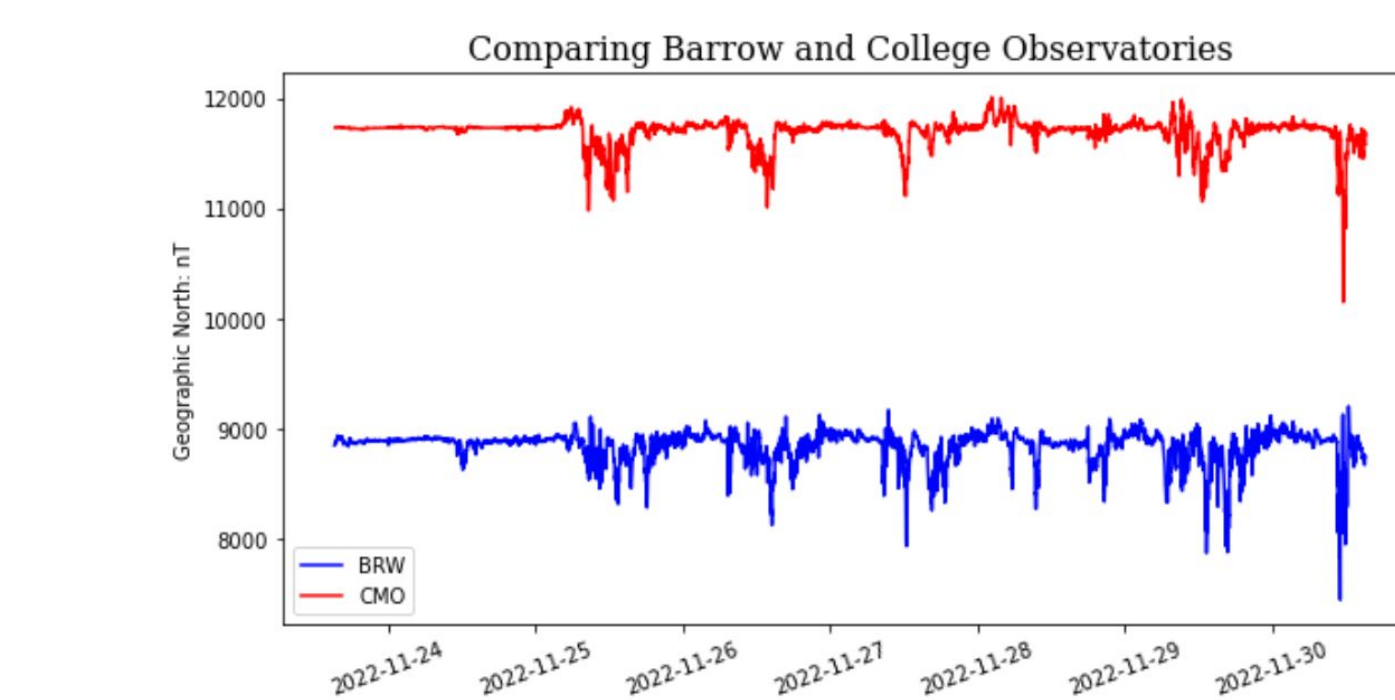
The graph on the left is an example of the model before any tuning of parameters. Testing the model with different prediction periods was crucial to learn about how the model was running.

Clean Data via Machine Learning model

Produce clean-data from real-time database: Multi-variant

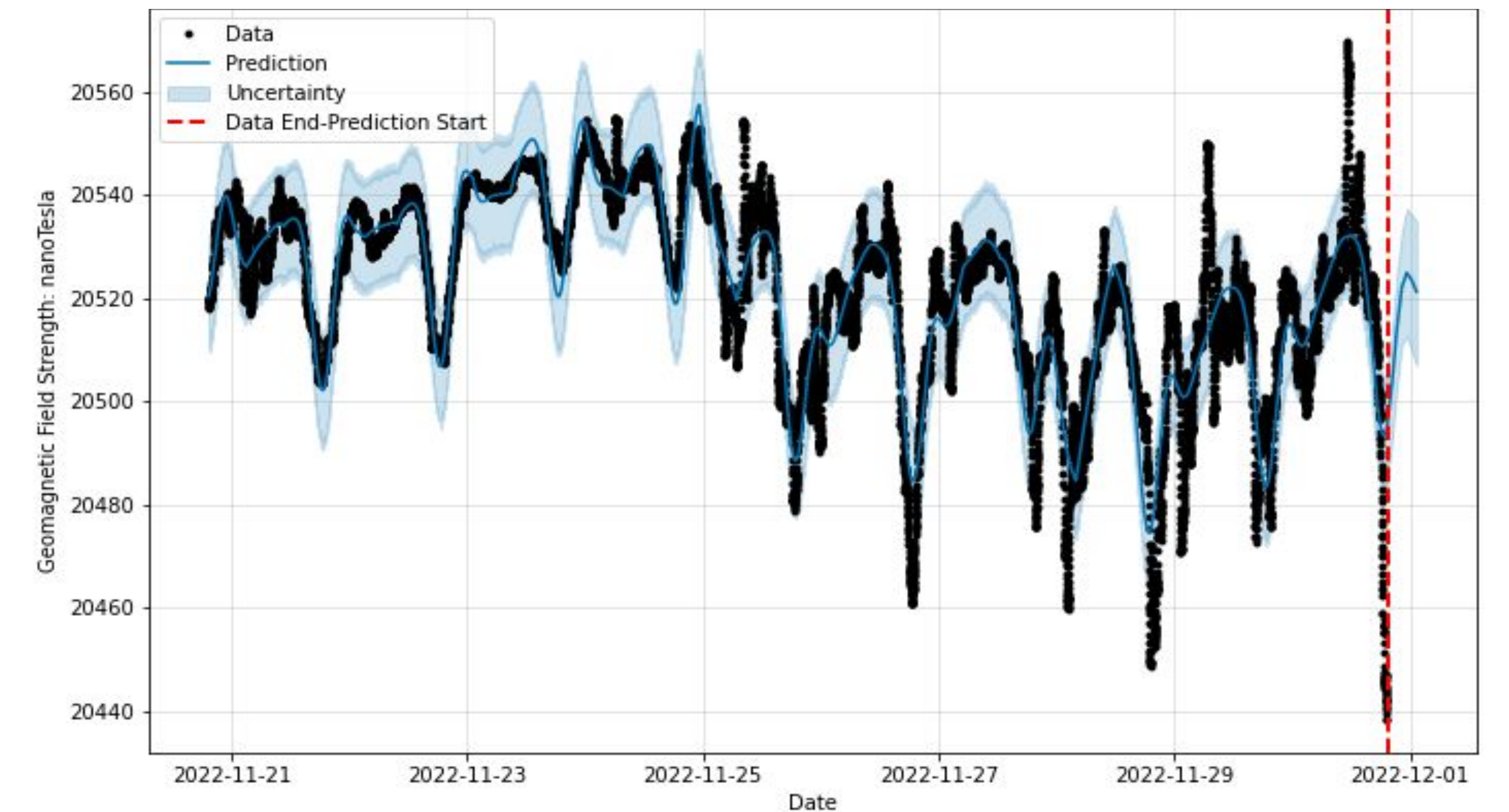


The model is currently producing clean data with no NaN values (graph on upper left). The model can be run on multiple stations, which is effective when looking at space weather events. The big graph on the left shows the model running on multiple spatial variables. It graph shows, in parallel, the model compared to data without the model.



The model can also be run on multiple stations. This can be seen in the graph on the left.

Result: Machine Learning model predicting a 6hr forecast

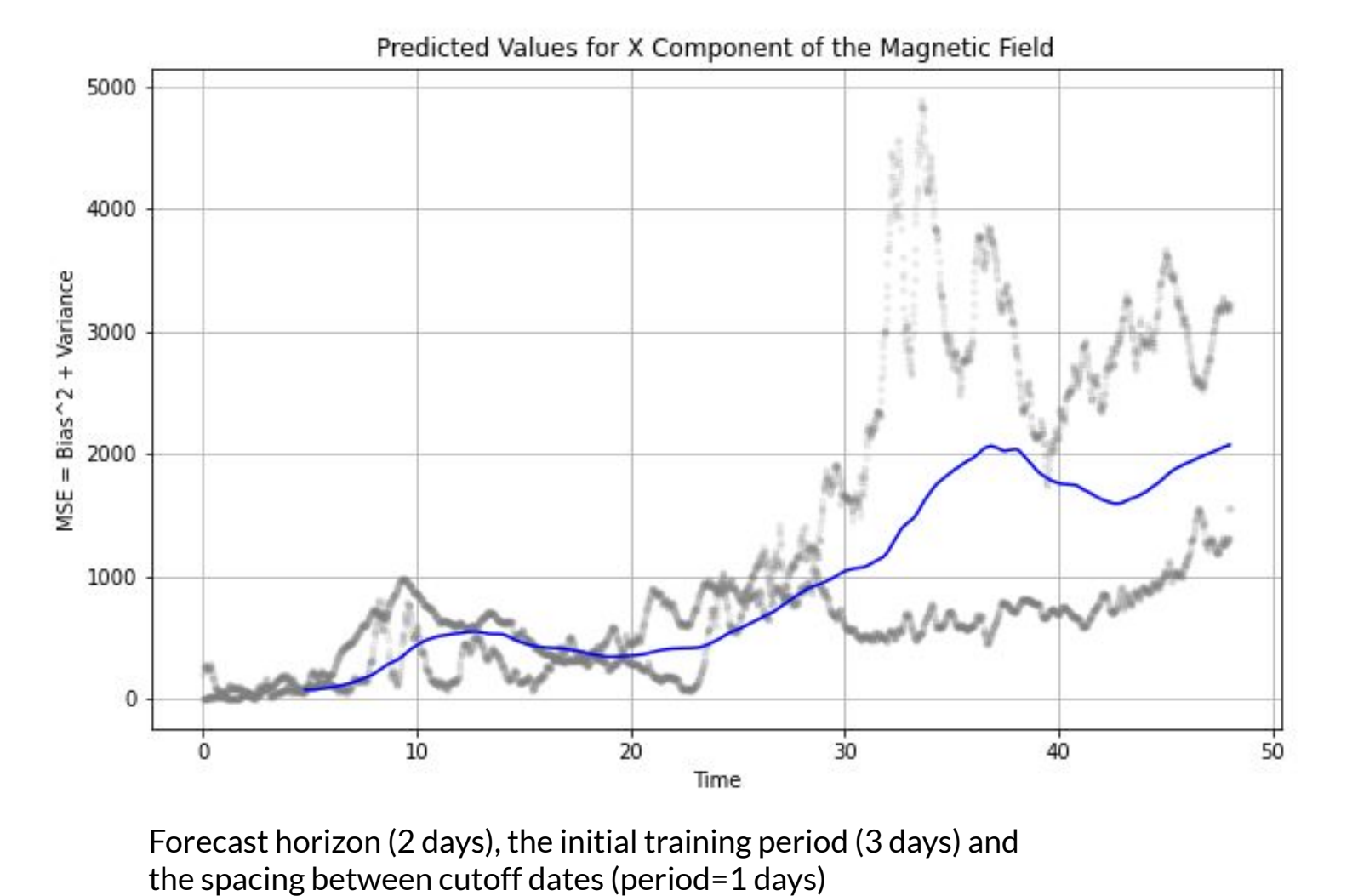


The result is a model that uses the historical data from the past 7 days and predicts the future forecast for up to 6 hours. The model also identifies and fills the gaps that occur in data. The Machine Learning model can be trained off of any amount of historical data. The model can be trained on all spatial variables (X,Y,Z) and can be used for any observatory. The current model has proved to have a greater confidence interval compared to the SARIMA/ARIMA models mentioned in the abstract.

On-going work: Hyperparameter tuning

Current work on the project will be testing and resampling the model in order to tweak the hyperparameters. Understanding and tuning the hyperparameters will allow for a greater accuracy in the data validation. Tightening the forecast error with Cross Validation will be essential for this ML model to run with more accuracy as shown on the plot.

The grey lines are the upper and lower bounds. The blue line is the Mean Squared Error(MSE) and bias variance decomposition. This allows for a good grasp of overfitting, underfitting and model capacity.



Future Work:

- Continue to sample, train, and validate the model, with potential to generate a Deep Learning algorithm.
- Implement model on geoelectric field maps for updated accuracy.
- Run model at the University of Colorado's Space Weather Technology Research and Education Center's (TREC) cloud-based Testbed environment.

Acknowledgements

This project was funded by NASA to the University of Colorado at Boulder: 80NSSC20K1477

Josh Rigler, Claire Raftery, Allison Liu, Maxine Hartnett, Matthew Bourque, Willow Reed, James Lende, Carter Woodson, Lila Uzzell, REU cohort, and USGS.

Prophet by Meta:
https://facebook.github.io/prophet/docs/quick_start.html