Isolating Solar Eclipse Features using Supervised Machine Learning

Mateo Cardona Serrano[1] (they/them)

Sevag Derghazarian[2], Larisa Goncharenko[2], Shunrong Zhang[2], Dupinder Singh[2]

Abstract

The Millstone Hill Ionospheric Steerable Antenna (MISA), seen in Fig. 1, is a radar antenna that emits high powered radio waves to provide measurements of near space environments. The radar has an altitude range of 90 to 1000 km, and a latitudinal and longitudinal sweeping range that spans from the arctic circle to the state of Florida. In April 2024, MISA recorded a trove of measurements on and surrounding the day of the April 8th solar eclipse.

Fig 1: MISA pointed south-west in Westford, MA

Inconveniently, the eclipse event was preceded by a period of elevated geomagnetic activity earlier in the day, making isolating the eclipse's effects on the ionosphere challenging. By leveraging data from adjacent days that were part of a monthslong MISA measurement campaign [1], we trained a high-resolution high-precision supervised machine learning model to generate accurate approximations of data products as a function of geophysical indices and universal time (UTC). Our results illustrate the spatial and altitudinal variation of electron density and temperature in the ionosphere due to the 2024 solar eclipse.

Theory

On April 8th, 2024, MISA recorded the electron density (Ne), electron temperature (Te), and ion temperature (Ti) data from the solar eclipse event at an elevation angle of 6°. Unfortunately, the presence of a geomagnetic storm earlier that day introduced background variations to the recorded data products (Fig. 2). Fortunately, the radar also recorded a trove of data products for the entire month of April. By leveraging this data, we can train a linear regression model that can emulate these background variations.

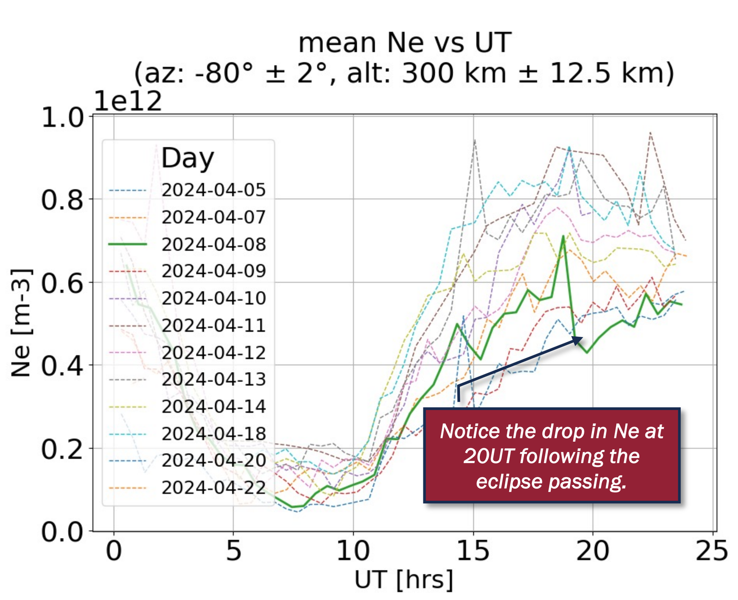

Fig. 2: Mean electron density in April. Notice the late evening variation. Eclipse day highlighted green.

This model can then be used to isolate the eclipse features from the geomagnetic storm by finding the difference between the model and recorded data, returning an approximate magnitude of the eclipse effects. To get the best model we want to select input features with a high degree of correlation to our output variables; the Flare Irradiance Spectral Model 2 (FISM2) [2] and geomagnetic Hpo [3] indices are high-resolution indicators of geophysical activity which demonstrate this high correlation. However, complicating our efforts to create a model is the presence of artifacts which may skew our results (Fig. 3). By first implementing an aggressive number of generous filters we can target multiple sources of error while maintaining prominent features of the data creating high-quality data to train our model.

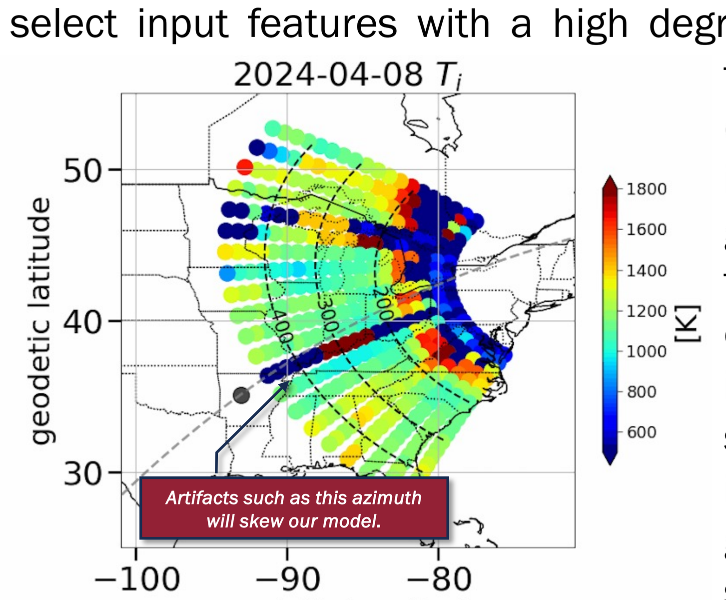

Fig 3: Recorded Ti at 19UT. Umbra & its path indicated in grey. Contour lines illustrate altitude.

Methods

The first step to creating our model was to increase the quality of our training data by filtering artifacts. We use a chain of layered filters with generous parameters to target multiple sources of error without degrading prominent features. The following figure illustrates the order and responses of the filters:

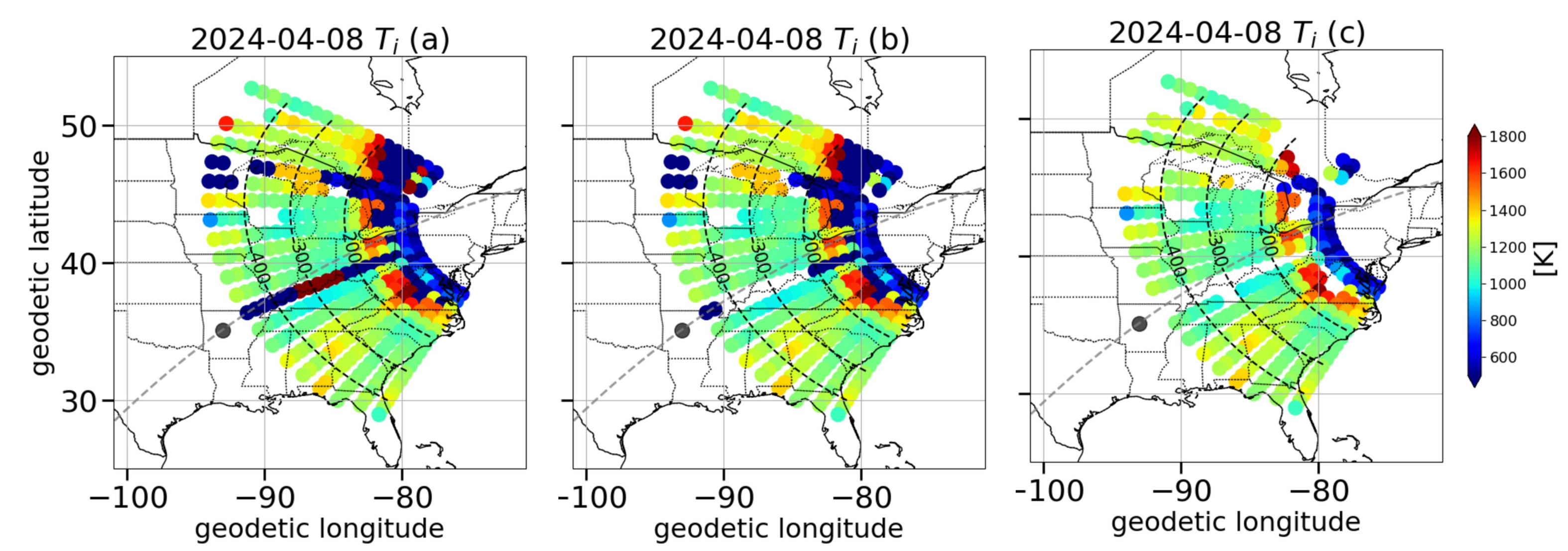

Fig 4: Recorded ion temperature at 19UT passing through each filter.
(a) Elements exceeding a 25% instrument rel. error are removed.
(b) Elements in each azimuth exceeding 2.5 standard deviations are removed.
(c) Data is binned by altitude to target differences in regional behavior. The bottom and upper ~5th percentile of data within each bin is removed.

This filtering is applied to all data products to generate a month of high-quality training data.

The space weather indices FISM2 and Hp30 are high-resolution measures of EUV irradiance [2] and geomagnetic activity [3]; they are appended to our data products along the time dimension. Fig. 5 illustrates their high degree of correlation with electron density (Ne), indicating that they are good input features to train our model. Days with space weather indices exceeding the eclipse day and the day of the eclipse itself were removed from the training data to prevent skewing. Next, we bin the data by azimuth and altitude to the following benefits:

- Bins increase flexibility in selecting the model's resolution & optimizing our system's compute.
- Bins capture changes in ionospheric phenomena at different regions.
- Bins maintain the original data architecture, granting more flexibility in contrasting the model with the recorded data.

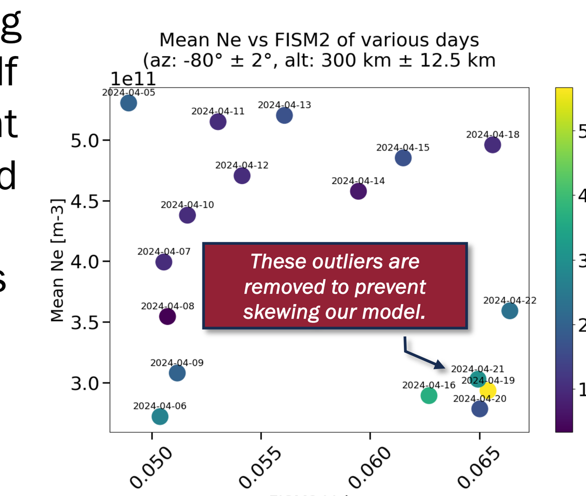

Fig 5: Daily mean electron density vs FISM2 index. Notice how days with high Hp30 index cluster together and correlate.

Figure 6(a) illustrates binned data from all the Ne data products in April in a particular azimuth and altitude region as a function of UT. A fourth-degree linear regression was trained for the electron density, electron temperature, and ion temperature as a function of UT, FISM2 and Hp30 from the binned data. Every permutation of altitude and azimuth bin was modeled and appended with the same dimensions as the original data. Figure 6(b) illustrates the recorded electron density of the eclipse day. Figure 6(c) illustrates the model output when it is given identical UT, FISM2, and Hp30 parameters as the eclipse day.

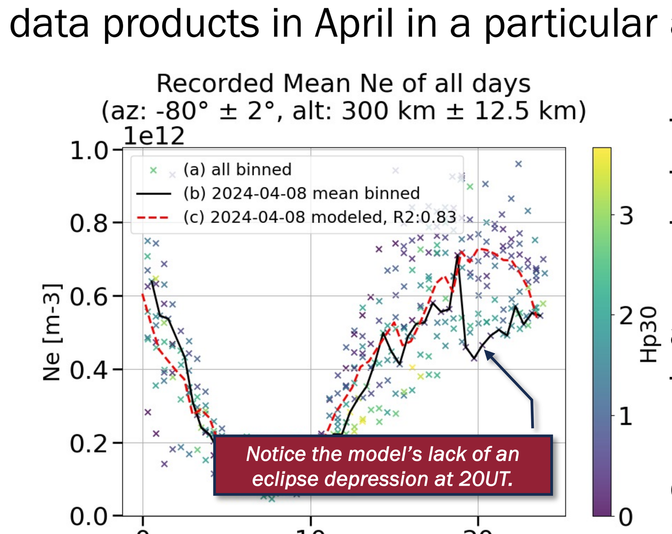

Fig 6: as a function of UT & Ne:
(a) mean training data for all remaining days
(b) mean of bins falling on eclipse day
(c) model with identical space weather parameters as eclipse day

Results & Conclusion

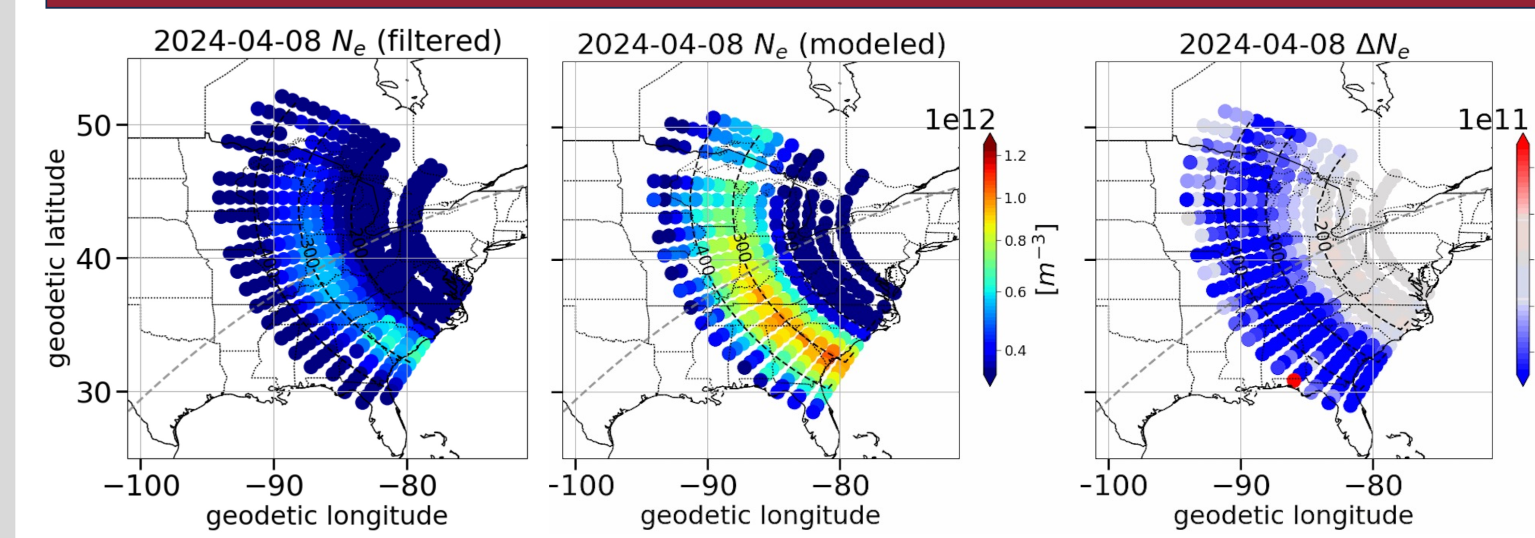

Fig. 7: Recorded, modeled background, and difference map of electron density (respectively) at 20.5 UT.

Figure 7 (left) illustrates the recorded Ne 1.5 hours after the umbra passes over at 19UT. The middle plot illustrates our modeled background results integrated over azimuth and altitude at the same time. The right plot illustrates the difference of the two former plots.

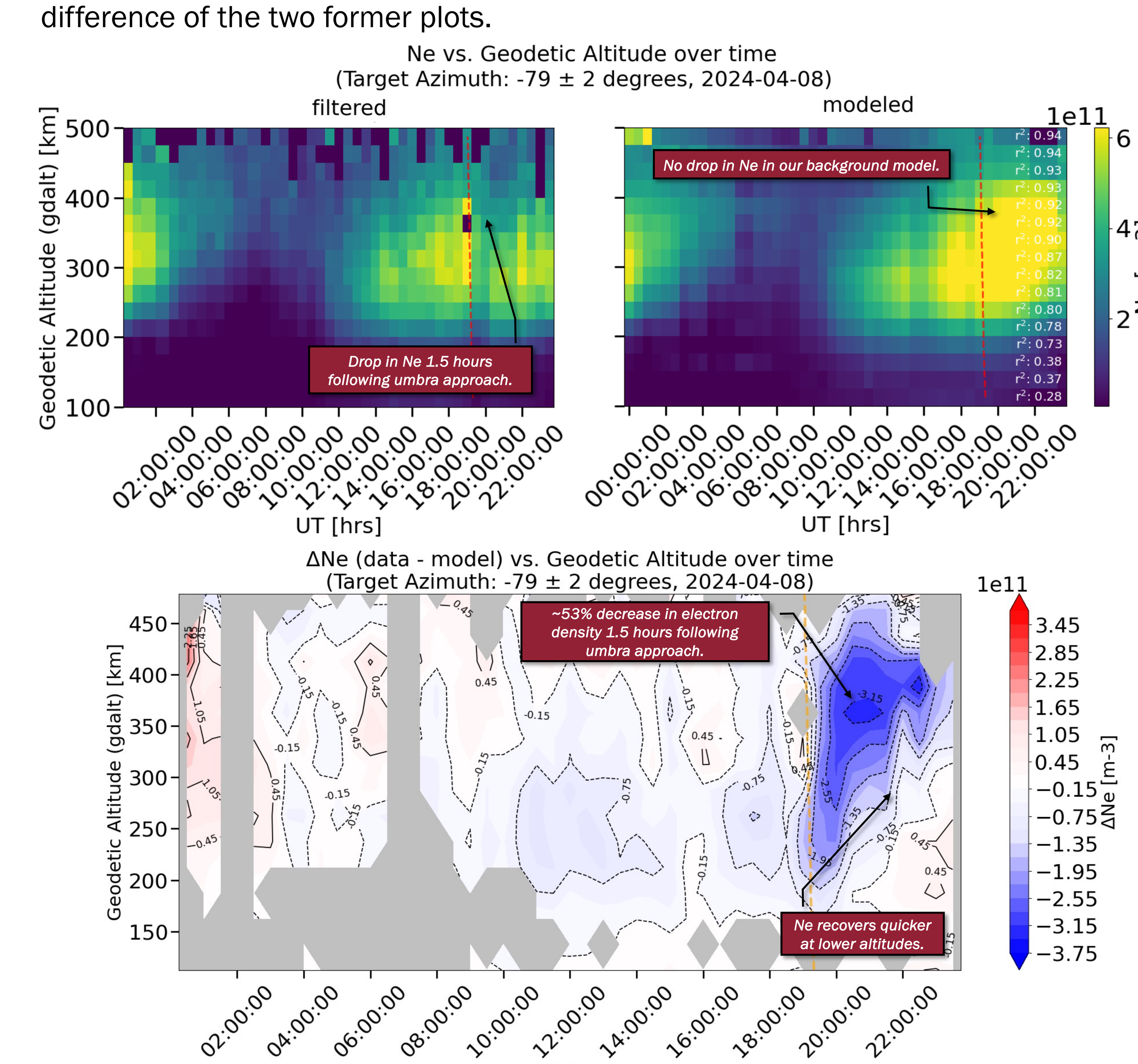

Fig. 8: As a function of altitude and UT, on eclipse day, in specified bin:
(left) recorded Ne (right) modeled Ne background
(bottom) recorded Ne minus background
Dashed line indicates time of closest umbra approach.

Figure 8 illustrates the altitudinal variation in electron density of a particular bin. The panels in the top row illustrate the model's remarkable ability to emulate Ne on April 8th without the presence of eclipse effects. The panel in the bottom row illustrates the result of removing the background from the recorded data. Our results aligned with physically expected results, demonstrating the effectiveness of our machine learning implementation in modeling background variations and highlighting the use of high-precision space weather indices in geophysical research.

[1]University of California, Berkeley, CA
[2]MIT Haystack, Westford, MA

[1] millstonehill.haystack.mit.edu
[2] Chamberlin, P. C., et al. (2020). The flare irradiance spectral model-version 2 (FISM2). *Space Weather*, 18, e2020SW002588.
[3] Yamazaki, Y., et al. (2022). Geomagnetic activity index Hpo. *Geophysical Research Letters*, 49, e2022GL098860.