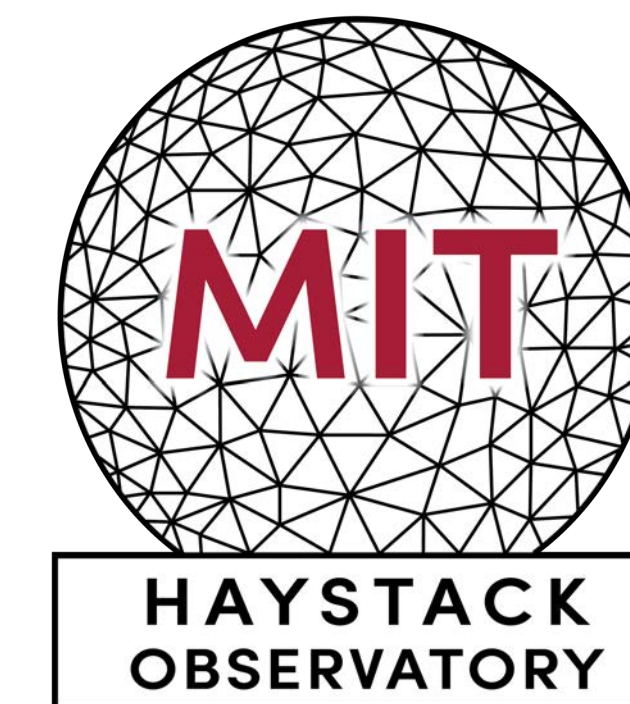


# A New Empirical Model for Ionospheric Total Electron Content

Cole Tamburri<sup>1,2</sup>, Larisa Goncharenko<sup>2</sup>, William Rideout<sup>2</sup>, Anthea Coster<sup>2</sup>

<sup>1</sup>Institute for Scientific Research at Boston College; <sup>2</sup>MIT Haystack Observatory



## Introduction

Ionospheric Total Electron Content (TEC) fluctuates based on a number of factors, including solar flux, geomagnetic activity, and seasonal variability. Models of these changes are constructed with the intention of determining the nature of TEC fluctuations as a result of external parameters. This project presents preliminary results for a new model for ionospheric TEC with unprecedented spatial and temporal resolution and data reliability, applied to all longitudes at 45 degrees North latitude.

## Data Sources

GNSS-based TEC, 10.7cm radio flux (f10.7), and geomagnetic activity data are provided by the CEDAR Madrigal Database. Extreme Ultraviolet (EUV) data are provided by the TIMED Solar Extreme Ultraviolet Experiment (SEE).

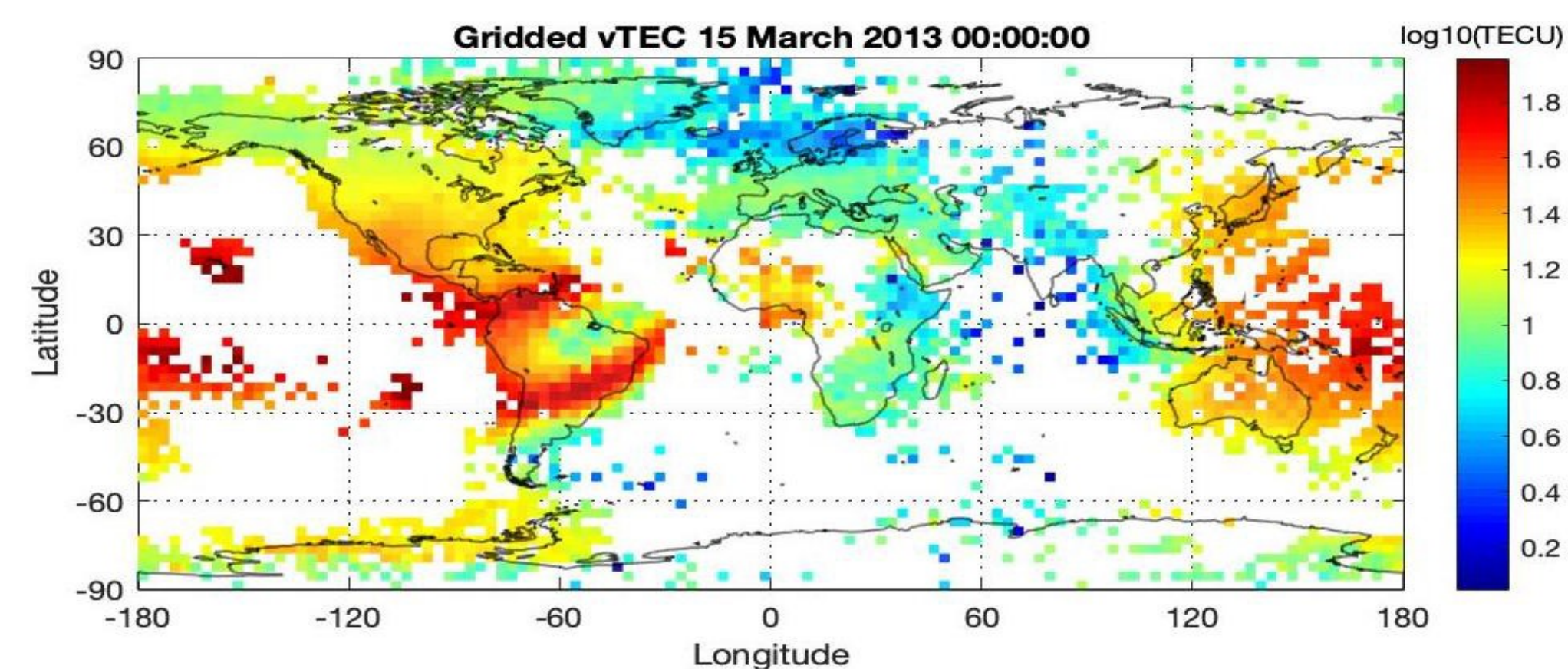


Figure 1. Sample gridded TEC data provided by the CEDAR Madrigal Database

## Data Preparation

Periodic recurrence of artificial data variability in TEC that is related to the sidereal motion of satellites is more severe at solar minimum, while regions of low GPS coverage offer less data than other locations. Several steps are taken to account for this, including (1) the application of a Hampel filter for outlier removal, (2) a cubic spline interpolant with low tolerance, and (3) "iterative model-building," a method which removes points with excessively high differences between the model and data under the condition that geomagnetic activity and solar flux are relatively low.

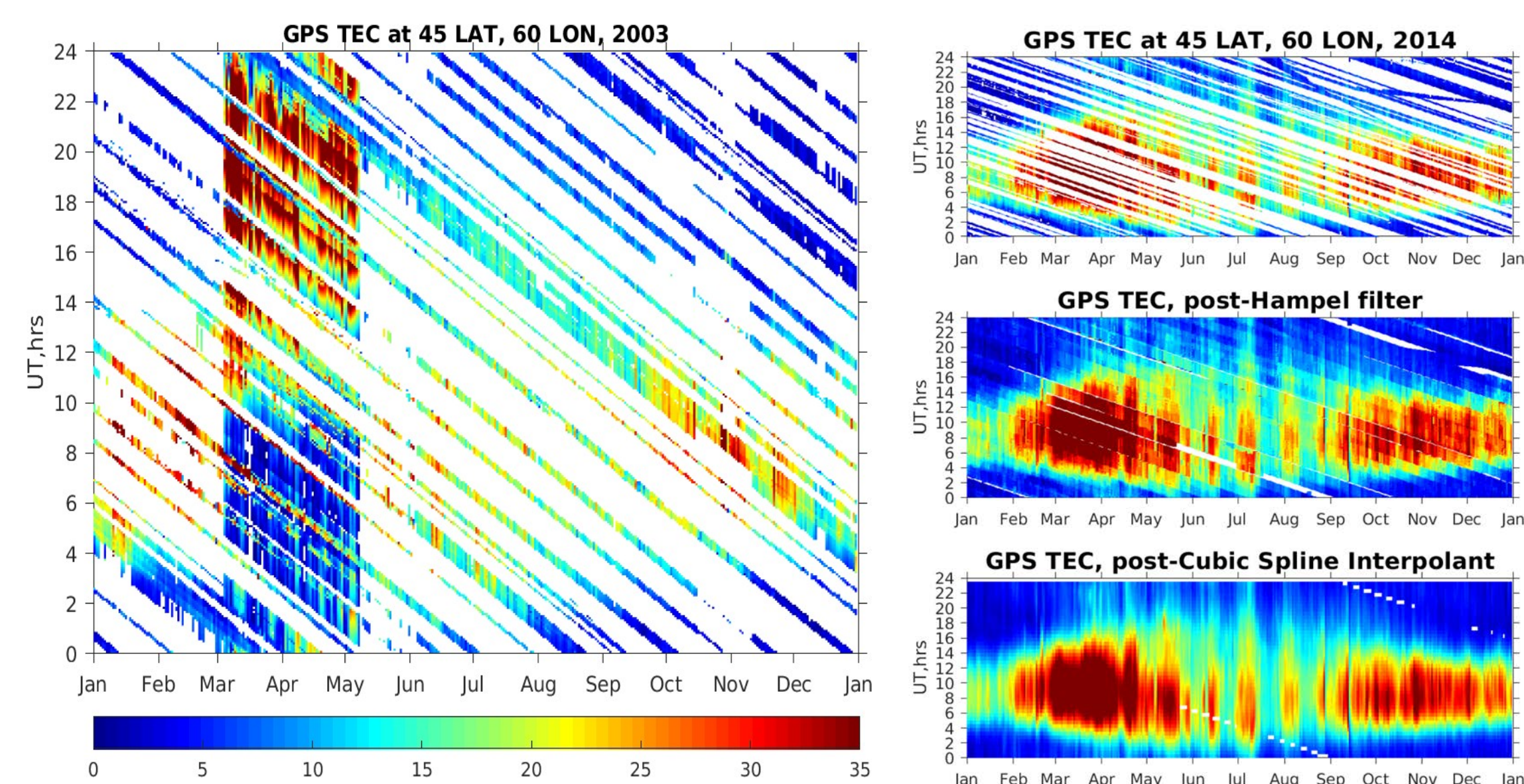


Figure 2. Sample raw TEC data with 5 minute temporal resolution. The effect of sidereal motion of satellites is evident.

Figure 3. The result of cleaning the data by application of a Hampel filter alongside a cubic spline interpolant with low tolerance.

## Model Description

The new empirical linear regression model is constructed with TEC as the response variable and several parameters representing solar EUV flux, geomagnetic activity, and season as independent variables. The model is based on GNSS TEC data from the years 2002-2019. The model is constructed at 45°N latitude in 3° longitude bins with a 30 minute temporal resolution and shows an average RMS error of 1.78 TECu for this location.

## Input Parameters

**Solar Flux:** Relative importance of EUV and the F10.7 radio flux was examined. While the latter captures the last solar cycle of data well, the EUV data are more comprehensive in capturing the solar spectral output. Modulation of the real-time EUV by seasonal terms is included, as are temporal delay terms. Unlike prior models, the parameter is not a moving average of solar flux.

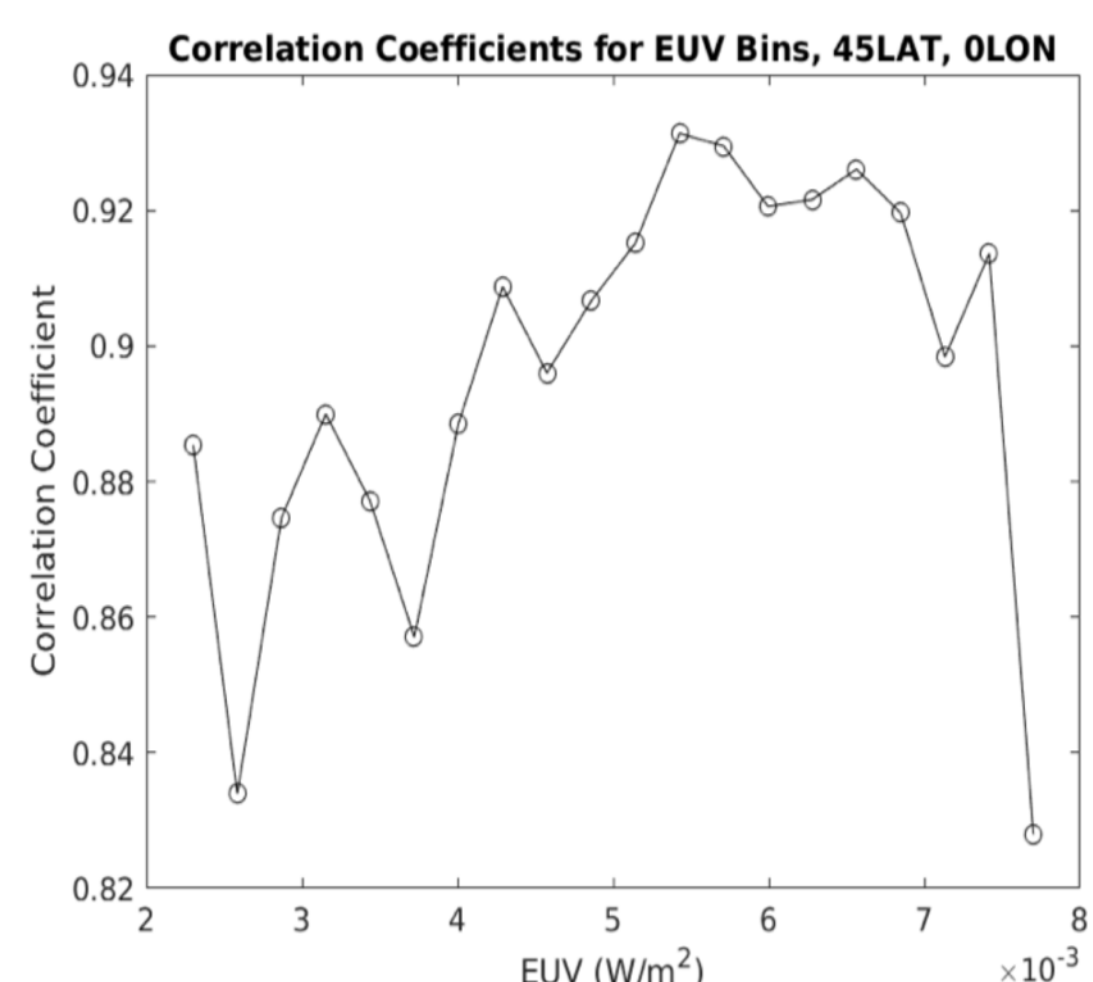


Figure 4. Correlation coefficients associated with data organized into bins according to solar flux value for all local times at 45°N latitude, 0°E longitude.

**Geomagnetic Index:** Relative significance of geomagnetic activity indices Ap3, Kp, and various delays was investigated. Ultimately it was found that integration of the Ap3 Index over the past 33 hours was unnecessary, but that several delays showed significance.

**Seasonal Variation:** Besides the modulation of EUV by seasonal terms, terms corresponding to annual, semi-annual, four-monthly, and three-monthly terms are themselves included in the model. The annual term was deemed superior to a solar zenith angle input for reasons of practicality.

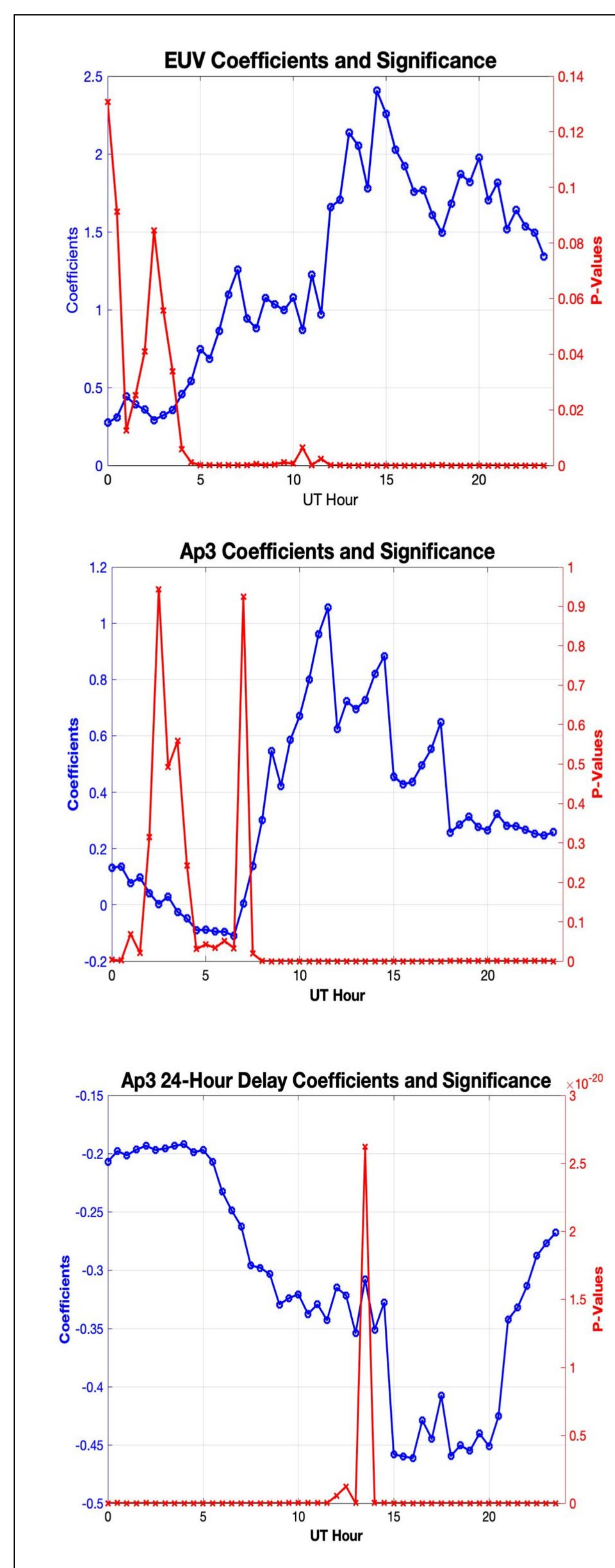


Figure 5. Sample coefficients and associated p-values at 45°N latitude, 0°E longitude corresponding to EUV (top), Ap3 (middle), and Ap3 with a 24 hour delay (bottom). A p-value of less than 0.05 indicates statistical significance. A positive storm effect is observed for the Ap3 data, while a negative storm effect is observed for the Ap3 input with a 24 hour delay.

## Input Parameters: Final Selection

The final selection of 36 input parameters includes a quadratic EUV term, EUV delays of 1 day, 8 days, 24 days, and 36 days, a quadratic Ap3 term, Ap3 delays of 3 hours, 24 hours, 48 hours (quadratic), and 72 hours, seasonal terms (annual, semiannual, four-monthly, three-monthly), modulation of the EUV linear term by annual, semiannual, and three-monthly terms, and modulation of the EUV quadratic term by all periodic terms.

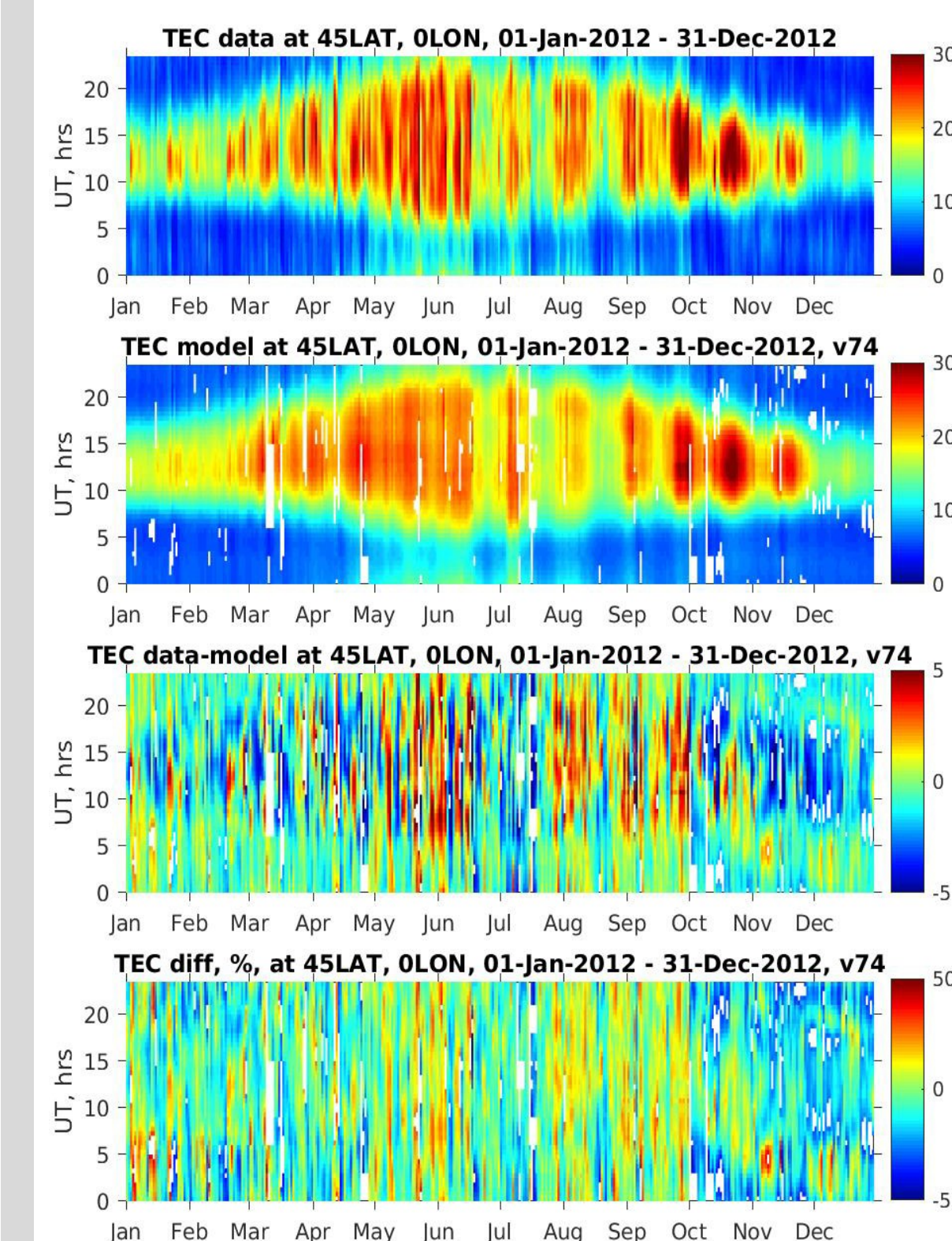


Figure 7. Sample comparison of data to model, 2012, 45°N latitude, 0°E longitude. (a) TEC data. (b) Constructed model. (c) Absolute difference between data and model. (d) Percentage difference between data and model.

## Future Work

Further study will include (1) expansion of the model to other latitudes beyond 45 degrees North, (2) expansion of the model to a longer time period, as it may be possible to expand the limiting dataset (TIMED SEE EUV) to the full extent of TEC data available (2000-2019), (3) investigation of additional drivers/independent variables, and ultimately, (4) identification of ionospheric anomalies using the final model.

## Acknowledgements

This work was funded by the National Science Foundation through its Research Experience for Undergraduates (REU) Program in cooperation with the MIT Haystack Observatory. Thanks are given to the development team of the CEDAR Madrigal Database and the TIMED SEE Mission.

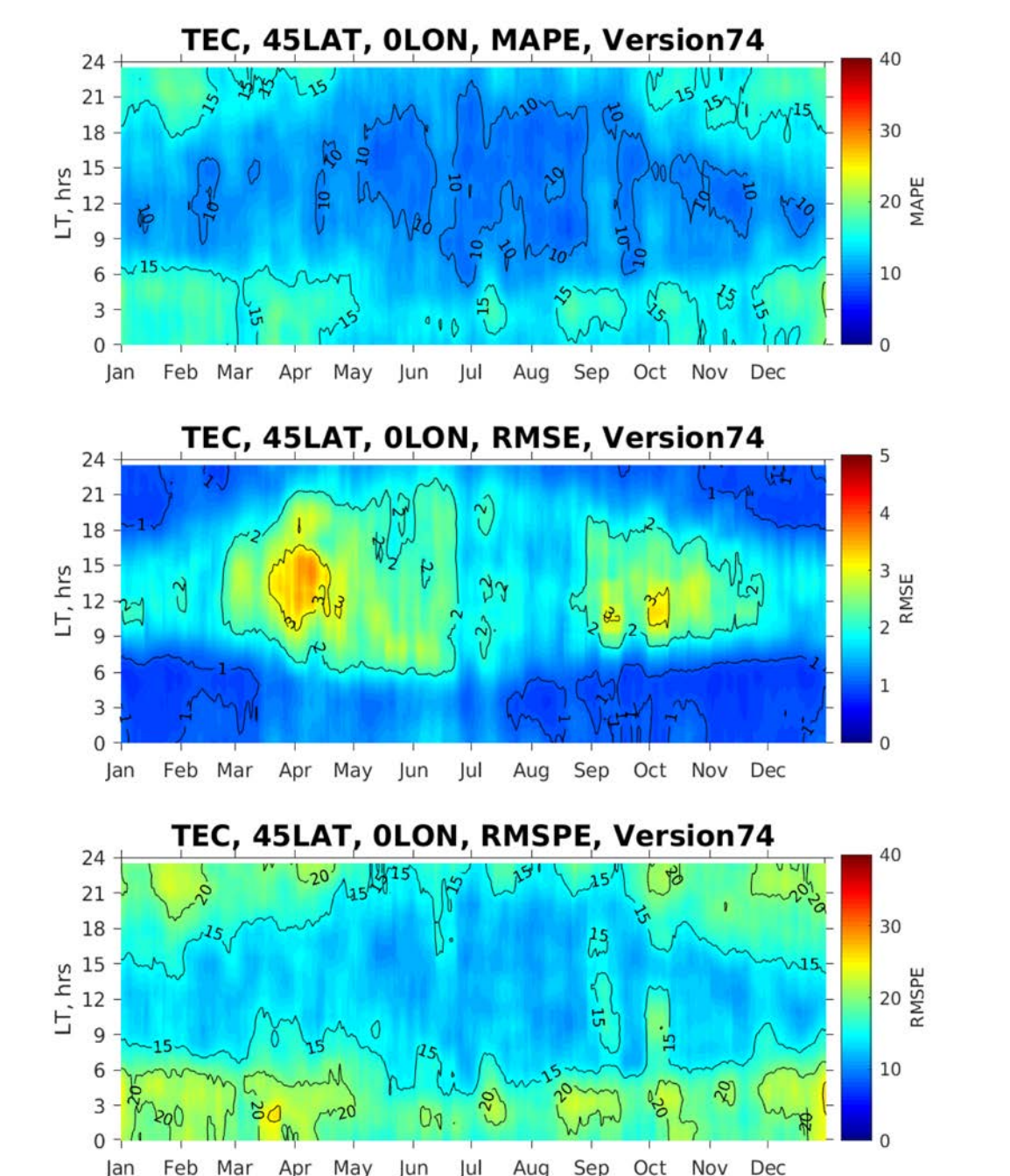


Figure 6. Sample statistical comparison of data to model, 45°N latitude, 0°E longitude. (a) Mean absolute percentage error. (b) Root mean squared error. (c) Root mean squared percentage error.

## Conclusions

- Preliminary results are presented for an empirical model for ionospheric TEC that has been produced with unprecedented spatial (3°) and temporal (30 min) resolution.
- The model is based on comprehensive global GPS data, allowing for less interpolation of original TEC data over areas of low coverage than prior models.
- Emphasis is placed on including only parameters with physical meaning in model construction.
- The model-to-data comparison shows small error values and good statistical agreement.