ML Model Forecasting of High and Mid Latitude Ionospheric Scintillation

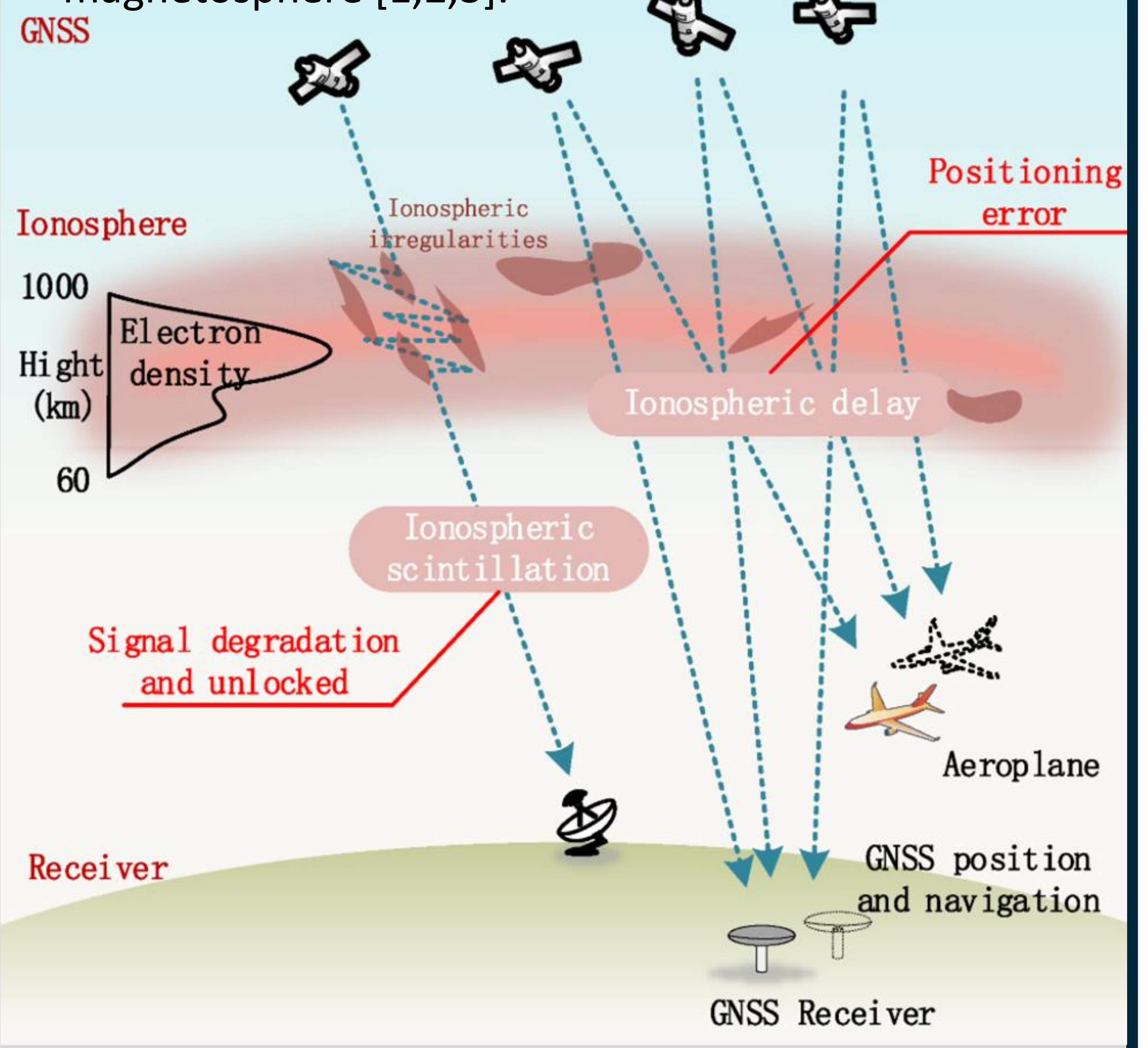
Jeffrey Marino[1], Luisa Capannolo[2], Jackson C. McCormick[1], Elyse Schetty[1], Benjamin E. McCrossan[1]
Ensemble Space Labs[1], Boston University[2]

Abstract

We have developed a proof-of-concept machine learning model that forecasts high and mid latitude ionospheric scintillation from solar and geomagnetic drivers. The training datasets include UNAVCO and CHAIN receivers, with temporal resolutions up to 5 minutes and spatial resolution of 1° by 1°, from approximately 25° to 80° latitude over 2015–2018. We calculated proxy scintillation indices from use geodetic receiver observations to mitigate the limited number of scintillation observations at mid-latitudes. The model's convolutional architecture captured spatiotemporal dependencies of TEC, space weather drivers, and phase and amplitude scintillation. The model generates one hour probabilistic forecasts for phase and amplitude scintillations, strongly outperforming a persistence model. The model is deployed on an AWS-hosted cloud environment that visualizes the model outputs on a map.

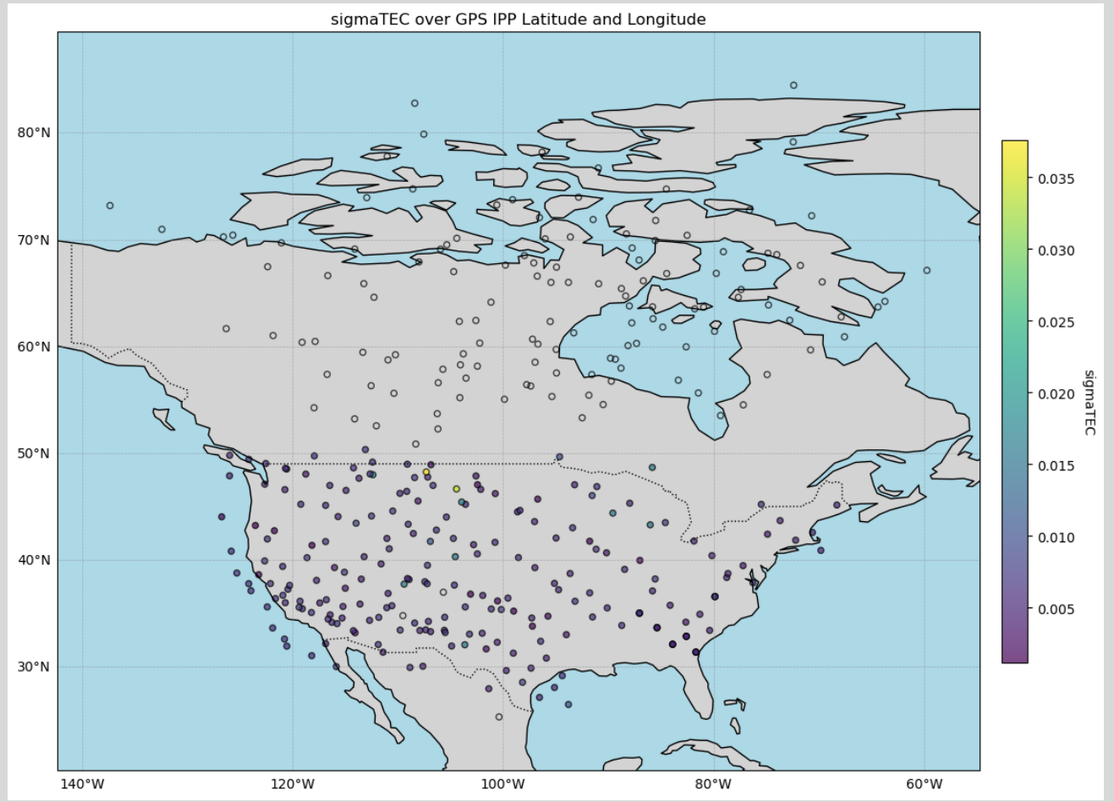
Ionospheric Scintillation

- Spurious noise observed in the trans-ionospheric satellite communication signals can cause errors in Position, Navigation and Timing (PNT).
- Trans-ionospheric signal noise is caused by rapidly varying electron density irregularities caused by particle precipitation from energetic particles in the magnetosphere [1,2,3].



- This precipitation is enhanced during increased geomagnetic activity (storms and substorms), triggered by solar wind fluctuations.
- Ionospheric scintillation is quantified by amplitude (S_4) and phase (σ_ϕ) indices.

Data & Model Description



- Canadian CHAIN receivers (25) provide high-resolution (50–100 Hz) amplitude (S4) and phase (σ_ϕ) scintillation indices.
- UNAVCO geodetic receivers provide 1-Hz signal-to-noise and phase measurements (1000s of stations in US).
- We follow the methods by Mrak et al. (2020) [4]. The amplitude and phase metrics SNR_4 and σ_{TEC} are:

$$SNR_4 = SNR_4' \cdot F(\theta)^{0.9};\ \ SNR_4' = \sqrt{<\delta SNR^2> - <\delta SNR>^2}$$

$$F(\theta) = \left(1 - cos^2(\theta)\left(\frac{R_e}{R_e + hipp}\right)^2\right)^{1/2}$$

$$\sigma_{TEC} = \sqrt{<\delta TEC^2> - <\delta TEC>^2}$$

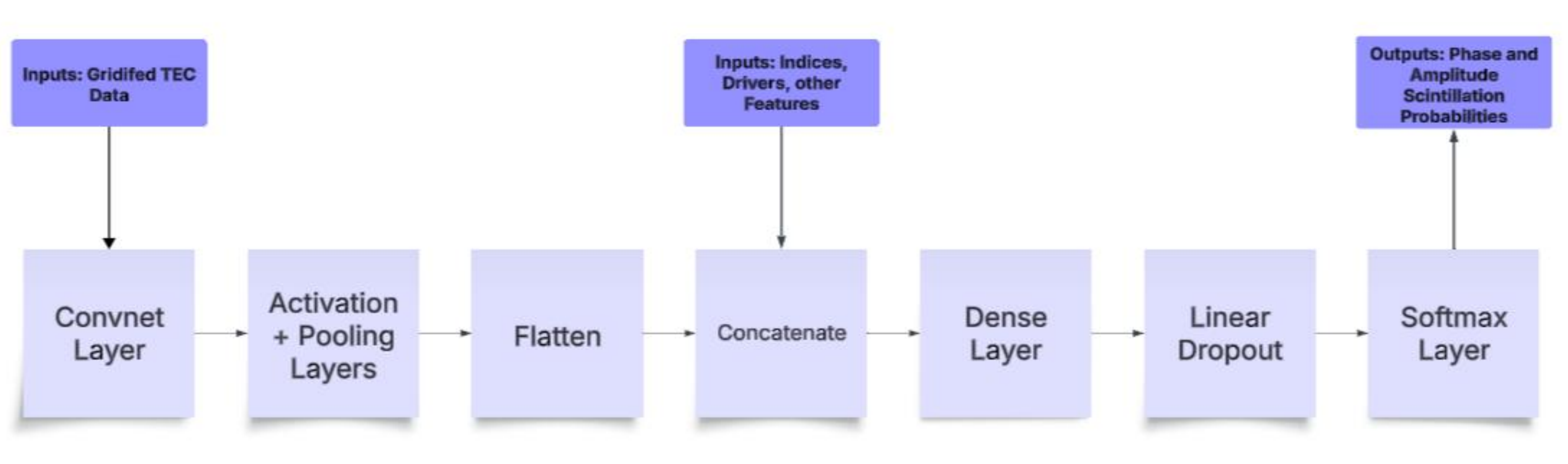
$$sTEC = \frac{1}{40.3}\frac{{f_1}^2{f_2}^2}{{f_1}^2 - {f_2}^2}(L_1 - L_2)10^{-16}TEC_u$$

- Where the δ indicates low-pass filtered data with a 0.1 Hz cutoff frequency and TEC is the total electron content. TEC is the vertical TEC found with the mapping F(θ)
- Scintillation at each receiver is defined individually based on the following heuristic by measurements 2 standard deviations above the quiet-time (defined by days with Kp<4, minutes with SMR> -30 nT and SML > -200nT)

Model inputs:

Data Type	Features
Spatial information	GPS IPP latitude longitude, MLT
Geomagnetic Indices [SuperMAG]	SME, SML, SMU, SMR (global & regional)
Temporal Data	Day of the year, time of the day
Solar Wind [OMNI]	Magnetic field (B_x, B_y, B_z), Speed (V_x, V_y, V_z, V), Density, Pressure, Temperature
TEC [Madrigal]	TEC, gradient of TEC
Scintillation indices [CHAIN, UNAVCO]	binary amplitude index binary phase index

- Feature selection:
  - Temporal features (e.g. time-of-day, day-of-year) and corresponding trigonometric transformations
  - Cutoff thresholds
- Training/Testing dataset selection:
  - Based on regional index activity (e.g. SMR_r)
  - Training: June-July 2015
  - Evaluation: July-September 2017
- Model Architecture selection:
  - Convolution layer synthesizes gridified TEC data;
  - Dense layers introduce other input features;
  - Outputs are probabilistic scalars



Inputs: Gridifed TEC Data

Inputs: Indices, Drivers, other Features

Outputs: Phase and Amplitude Scintillation Probabilities

Convnet Layer → Activation + Pooling Layers → Flatten → Concatenate → Dense Layer → Linear Dropout → Softmax Layer

References

[1] Kintner, P. M., B. M. Ledvina, and E. R. de Paula (2007), GPS and ionospheric scintillations, *Space Weather*, 5, S09003, doi:10.1029/2006SW000260.

[2] Ledvina, B. M., J. J. Makela, and P. M. Kintner, First observations of intense GPS L1 amplitude scintillations at midlatitude, *Geophys. Res. Lett.*, 29(14), doi:10.1029/2002GL014770, 2002.

[3] Nishimura, Y., Kelly, T., Jayachandran, P. T., Mrak, S., Semeter, J. L., Donovan, E. F., et al. (2023). Nightside high-latitude phase and amplitude scintillation during a substorm using 1-second scintillation indices. *Journal of Geophysical Research: Space Physics*, 128, e2023JA031402. https://doi.org/10.1029/2023JA031402

[4] Mrak, S., Semeter, J., Nishimura, Y., Rodrigues, F. S., Coster, A. J., & Groves, K. (2020). Leveraging geodetic GPS receivers for ionospheric scintillation science. Radio Science, 55, e2020RS007131. https://doi.org/10.1029/2020RS007131

Conclusions/Acknowledgements

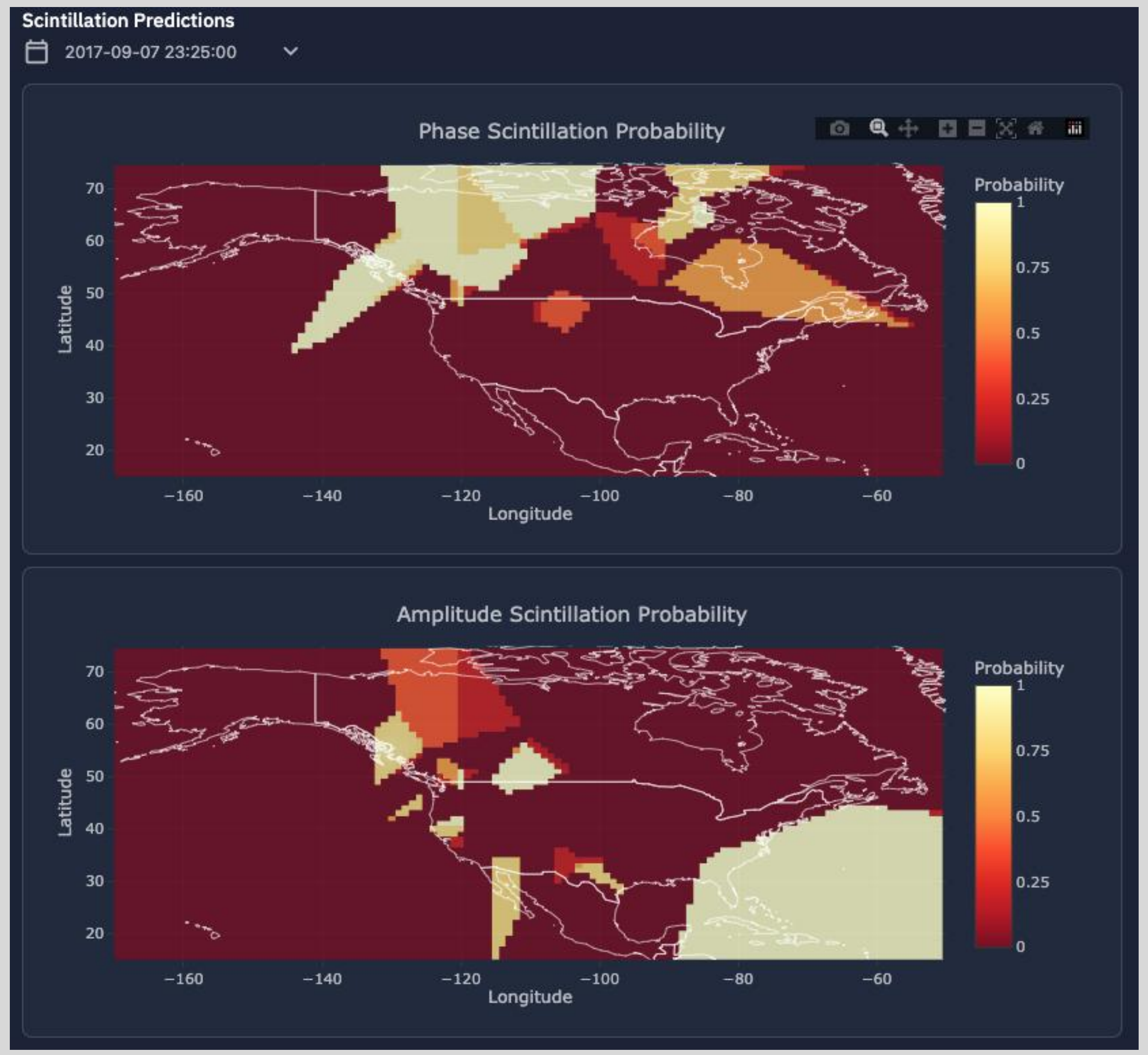
- We develop a ML model for mid and high latitude scintillation using space weather drivers and GNSS scintillation measurements.
- The model harnesses the spatially diverse GNSS measurements from ground-based receivers leveraging a CNN architecture.
- The model's TSS score is 0.53 and 0.44 for amplitude and phase scintillation, respectively.
- The phase scintillation forecast outperforms the baseline persistence model TSS of by nearly 100%.

Prototype Model Results

- To measure the model performance, we use the True Skill Score (TSS), which measures the model's ability to distinguish the two classes while accounting for the relative imbalance between the classes:

$$TSS = \frac{TP}{TP + FN} - \frac{FP}{FP + TN}$$

- The model generates probabilistic forecasts for phase and amplitude scintillations one hour into the future, achieving TSS of 0.53 and 0.44, respectively. The phase scintillation forecast outperforms the baseline persistence model TSS of .27 by nearly 100%.



Scintillation Predictions

2017-09-07 23:25:00

Phase Scintillation Probability

Amplitude Scintillation Probability

- The confusion matrix compares true observations (binary variable of yes (1) and no (0) phase scintillation) on the y-axis with the predictions from the model on the x-axis, on a per-sample basis.
- The model enjoys:
  - a relatively high specificity of 0.96;
  - a relatively strong recall of 0.62;
  - and a modest precision of 0.42.