Where Does Ionospheric Predictability Reside?

Univariate, Multivariate, and Cross-Station Forecasting of Ionosonde Time Series

Alvaro Garcia[1], Enrique Rojas[2]

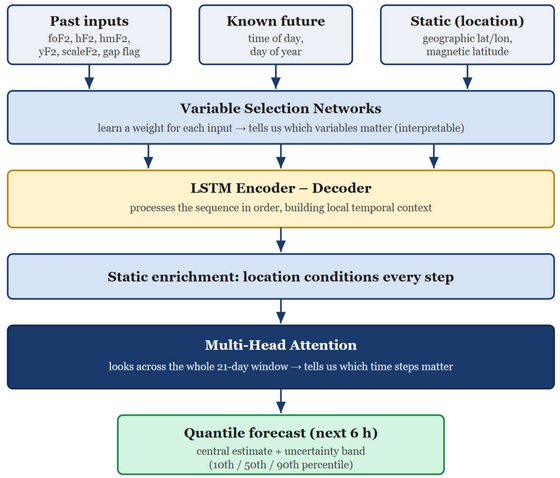
1. Universidad de Ingenieria y Tecnologia
2. MIT Haystack Observatory

Introduction

The **F2-layer critical frequency (foF2)** sets the maximum frequency reflected by the ionosphere and is key for HF communication, navigation, and space-weather operations. Its evolution is driven by solar radiation, geomagnetic activity, and neutral-atmosphere coupling, making it highly variable and hard to forecast, especially near the magnetic equator.

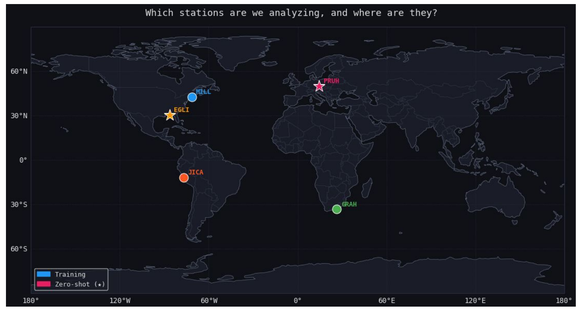
It remains open **where the predictive information resides:** in a parameter's own history, in coupled observables, in station location, or in exogenous solar and geomagnetic drivers.

We use the **Temporal Fusion Transformer (TFT)** because its variable-selection and attention mechanisms make it **interpretable**: it exposes which inputs and time steps drive each forecast, exactly what is needed to locate that information. It also handles multivariate inputs, static covariates, and probabilistic outputs.



Motivation

We use **forecast skill itself as a probe** of ionospheric coupling and regional coherence, addressing three questions: how skill depends on the **forecast horizon**, how it depends on station **location and transfers across sites**,.

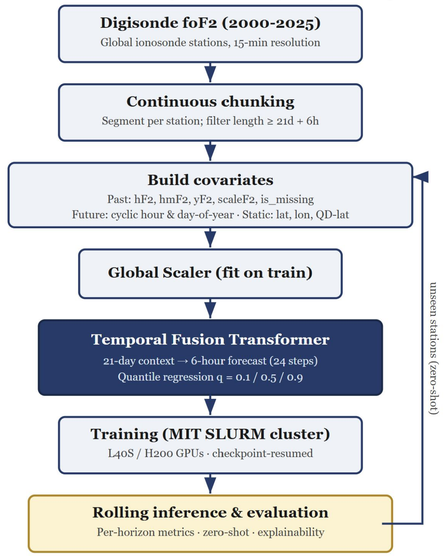


Data & Model

We use Digisonde **foF2** time series from global ionosonde stations spanning **2000-2025** at **15-minute resolution**, segmented into continuous segments. Each segment carries the forecast target (**foF2**) and three groups of inputs: **past observables** (hF2, hmF2, yF2, scaleF2, and a flag marking gap-filled values); **temporal inputs** (time of day and day of year); and **station location** (geographic latitude, longitude, and magnetic latitude).

The model is a **Temporal Fusion Transformer (TFT)**. It reads a **21-day window** of past data to forecast the **next 6 hours**. Rather than a single value, it predicts a **central estimate with an uncertainty range** (10th, 50th, and 90th percentiles).

Models are trained on three stations: **Millstone Hill (MILL)**, **Jicamarca (JICA)**, and **Grahamstown (GRAH)**; and tested on two stations never seen in training, **Pruhonice (PRUH)** and **Eglin AFB (EGLI)**.



Forecast Skill vs Horizon

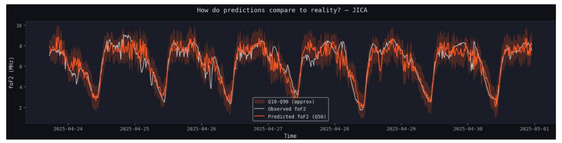
We evaluate rolling 6-hour forecasts over a held-out test set (2024-2025), reporting error at each 15-minute step from h+15 min to h+6 h.

Global test metrics (median quantile):

Station	MAE	RMSE	MAPE %	Med AE
MILL	0.532	0.726	11.5	0.405
JICA	0.925	1.260	15.3	0.679
GRAH	0.801	1.213	14.5	0.526

All errors in **MHz**. **MAE** is the average absolute error; **RMSE** is similar but penalizes large misses more heavily; **MAPE** is the error as a percentage of the observed value; **Median AE** is the typical (50th-percentile) error, less sensitive to outliers. **MILL** (mid-latitude) is the most predictable site; **JICA** (magnetic equator) carries the largest error, consistent with strong equatorial F-region variability.

Skill **degrades monotonically with horizon,** but the rate is station-dependent. From h+15 min to h+6 h, MAE grows by **+95% at MILL, +157% at GRAH, and +68% at JICA**, the equatorial site starts worse but degrades more slowly, while GRAH degrades fastest.



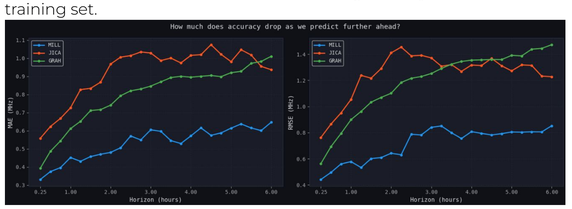
A small **signed bias** remains: the model slightly **underestimates** foF2 at MILL (-0.017 MHz) and GRAH (-0.022 MHz) and slightly **overestimates** at JICA (+0.049 MHz). Quantile bands widen with horizon, reflecting calibrated uncertainty growth.

Spatial Dependence & Cross-Station Transfer

A central goal is to test whether ionospheric dynamics learned at a few stations **transfer to sites never seen in training**; the basis for regional, longitudinal forecasting chains. We apply the trained model with no retraining to PRUH and EGLI.

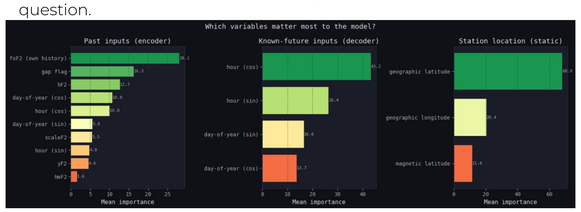
Transfer is **strongly station-dependent**. **EGLI generalizes remarkably well** (MAE 0.810 MHz, on par with the trained station GRAH at 0.801), **while PRUH carries the largest error of all** (1.083 MHz, exceeding even the equatorial JICA). This shows that a substantial part of foF2 predictability is **shared across mid-latitude sites,** but generalization is not uniform.

Horizon degradation amplifies the gap: PRUH grows **+228%** from h+15 min to h+6 h, the steepest of all stations, whereas EGLI (+130%) stays close to the trained sites. Cross-station skill therefore **holds best at short horizons** and for stations magnetically similar to the training set.



Where the Information Resides

We use the TFT variable-selection weights to quantify which inputs drive the forecast, directly addressing the univariate-vs-multivariate question.

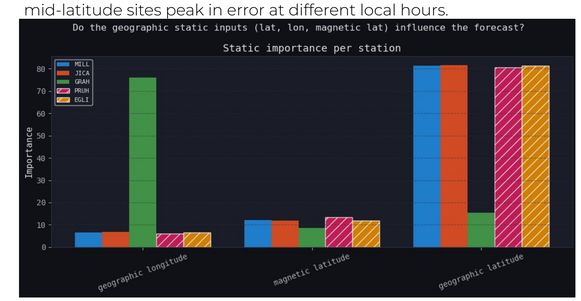


Among **past observables**, the target's own history dominates (foF2, 28.1), **but coupled parameters add measurable skill:** hF2 (12.7) and the missing-data flag (16.3) rank above the cyclic encoders, while hmF2 contributes least (1.6). This confirms that **multivariate context improves on a purely univariate forecast.**

Among **known-future inputs**, the **diurnal cycle dominates** (hour cosine 43.2, hour sine 26.4), reflecting the strong day-night control of the F2 layer.

Among **static covariates, geographic latitude is by far the strongest** (68.0), above longitude (20.4) and magnetic QD-latitude (11.6); the model organizes site identity primarily by geographic position.

Error is also **diurnally structured**: JICA is worst near **02 h UTC** (post-sunset equatorial irregularities) and best at midday, while mid-latitude sites peak in error at different local hours.



Conclusions & Ongoing Work

- foF2 is forecastable 6 h ahead with MAE ≈ 0.5-0.9 MHz; skill decays with horizon at station-dependent rates (+68% to +157%).
- Multivariate context (hF2, missing-flag) measurably improves on univariate forecasts; the diurnal cycle is the dominant known driver.
- Models transfer zero-shot to unseen sites, but unevenly; EGLI matches trained stations while PRUH does not.

Acknowledgements

Computational resources were provided by the MIT SLURM cluster. We thank MIT Haystack Observatory and UTEC for their support.