Machine Learning for Local VTEC Forecasting from GNSS Observations

A first single-station assessment toward regional ionospheric modelling

Year:2026

Outline

- Introduction
- Objective
- Methodology
- Outcomes
- Ongoing and Future Works
- Conclusion

Introduction

Motivation: why ionosphere matters for GNSS

- Precise GNSS positioning is affected by ionosphere delay, which depends on the electron content along the signal path.
- N-RTK networks usually mitigate this effect using reference stations and spatial interpolation, but rapid ionospheric variability, seasonal changes and geomagnetic disturbances can limit these approaches.
- Global VTEC products such as IONEX are valuable reference, but they are not specifically optimized for local/regional correction networks. This motivates local GNSS-derived VTEC forecasting using Machine Learning.

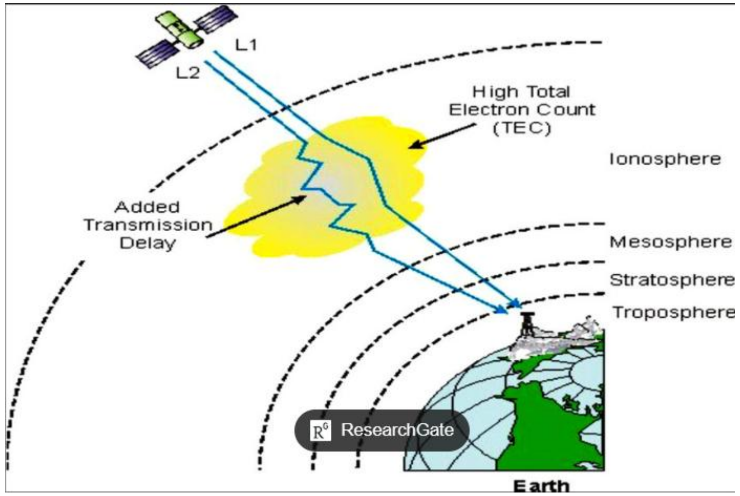


fig1. Satellite-receiver path

What affects the ionosphere?

- Solar Radiation
- Solar Cycle
- Geomagnetic activity
- Solar wind
- Earth magnetic field
- Local time
- Season

The ionosphere is driven by solar radiation and geomagnetic activity; therefore, VTEC changes with time, season and geospace conditions

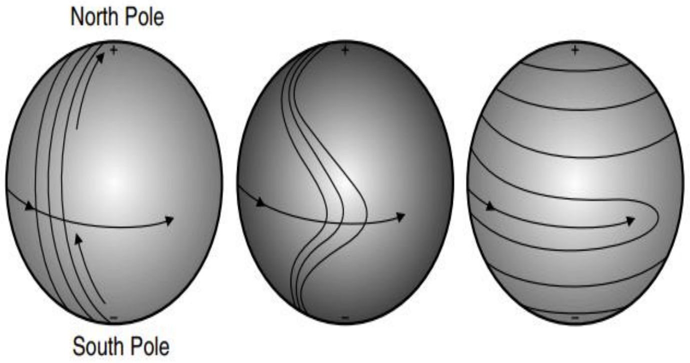


fig1. Sun rotation

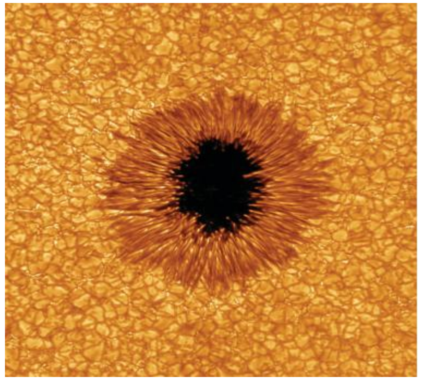


fig1. Sun spot

Introduction

What is TEC/VTEC?

- TEC = Total Electron content along path.
- STEC = Slant TEC along the satellite-receiver path.
- VTEC = Vertical TEC mapped to a vertical path.

STEC → Mapping function → VTEC

$$VTEC = STEC * F(elevation)$$

GNSS observation principle

The ionosphere delay depends on frequency, so dual-frequency GNSS observations allow ionospheric information to be extracted.

$$I \; \alpha \; \frac{1}{f^2}$$

I: *ionospheric delay*
f: *frequency*

From GNSS observations to local VTEC

$$RINEX\ observations \rightarrow DCB\ corrections \rightarrow Stec\ estimation \rightarrow arc\ filtering \rightarrow phase\ levelling$$

$$\downarrow$$

$$Robust\ local\ GNSS - derived\ VTEC\ target \leftarrow 15 - min\ time\ bins \leftarrow mapping\ function$$

Objective

Objective 1 – VTEC reference and validation

- Use IONEX global VTEC as an external reference to assess the consistency of the robust local GNSS-derived VTEC estimated from BRTS muti-GNSS observations and IGS DCB products.

Objective 2 – Machine Learning modelling and performance assessment

- Develop and evaluate XGBoost-based VTEC forecasting models using local GNSS-derived VTEC as the prediction target.
- Compare two XGBoost configurations: one without temporal memory using physical/environmental descriptors, and one with temporal memory using past VTEC values for short-term forecasting.

Methodology

Why DCB and phase levelling?

- Code measurements are noisy but absolute.
- Carrier phase measurements are precise but ambiguous.
- Phase levelling combines both advantages.

The objective is to obtain a cleaner and more stable local VTEC target for Machine Learning.

Arc correction / cycle slips

- GNSS observations are not continuous forever.
- A satellite track is divided into arcs.
- Cycle slips or data gaps break the continuity.

$$\boldsymbol{Arc = continuous\ satellite\ observation\ segment}$$

15-minute time bins

Instead of using each raw 30-second observation directly, observations are grouped into 15-minute time bins.

Methodology

Feature generation

Feature group	Examples	Why useful
Time-related	hour_sin, hour_cos, doy_sin , doy_cos	Daily/seasonal cycles
Solar	Solar_zenith, solar local time	Ionization driver
GNSS geometry	Elevation, IPP lat/lon	Signal path/spatial sampling
Geospace	Kp, Dst, F10.7	Solar/geomagnetic activity
Observation quality	n_satellites, n_observations	Reliability of the bin
Temporal memory	VECT_lag_1, VTEC_lag_4	Recent VTEC state

Table1. Feature group

Why sine/cosine features?

- ✓ Hour is cyclic: 23:45 is close to 00:00.
- ✓ DOY is cyclic: day 365 is close to day 1.

$$\boldsymbol{hour\_sin = sin\left(2\pi * \frac{hour}{24}\right)}$$

$$\boldsymbol{hour\_cos = cosine\left(2\pi * \frac{hour}{24}\right)}$$

Methodology

What is Machine Learning doing here?

$$Input\ features\ \rightarrow ML\ model \rightarrow predicted\ VTEC$$

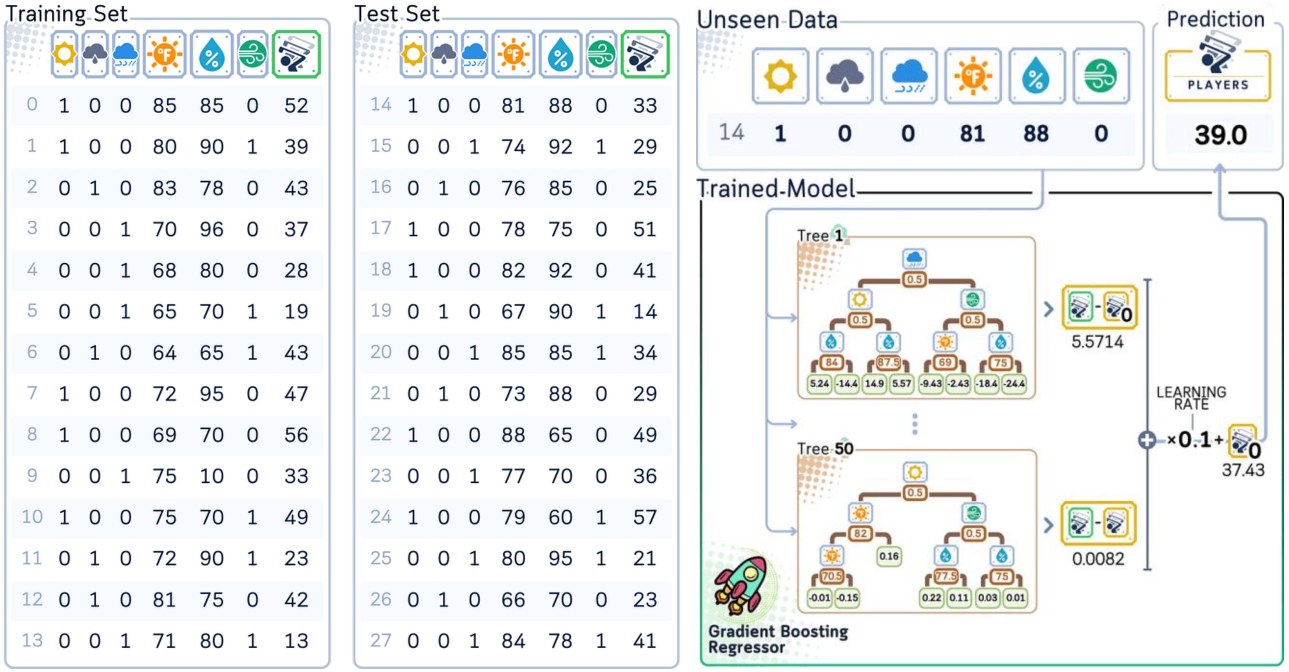
The model learns relationships between the input conditions and the local VTEC target from historical data.

What is XGBoost?

- XGBoost is an ensemble of decision trees.
- Each tree makes simple decisions.
- The model combines many trees to reduce errors.

Decision tree intuition

A decision trees splits the data into conditions, XGBoost builds many trees sequentially, each one correcting previous errors.



Training Set

0	1	0	0	85	85	0	52
1	1	0	0	80	90	1	39
2	0	1	0	83	78	0	43
3	0	0	1	70	96	0	37
4	0	0	1	68	80	0	28
5	0	0	1	65	70	1	19
6	0	1	0	64	65	1	43
7	1	0	0	72	95	0	47
8	1	0	0	69	70	0	56
9	0	0	1	75	10	0	33
10	1	0	0	75	70	1	49
11	0	1	0	72	90	1	23
12	0	1	0	81	75	0	42
13	0	0	1	71	80	1	13

Test Set

14	1	0	0	81	88	0	33
15	0	0	1	74	92	1	29
16	0	1	0	76	85	0	25
17	1	0	0	78	75	0	51
18	1	0	0	82	92	0	41
19	0	1	0	67	90	1	14
20	0	0	1	85	85	1	34
21	0	1	0	73	88	0	29
22	1	0	0	88	65	0	49
23	0	0	1	77	70	0	36
24	1	0	0	79	60	1	57
25	0	0	1	80	95	1	21
26	0	1	0	66	70	0	23
27	0	0	1	84	78	1	41

fig4. XGBoost model Example

Methodology

Hyperparameters

Feature group	Why useful
N_estimators	Number of trees
Max_depth	Complexity of each tree
Learning_rate	How strongly each tree corrects errors
Subsample	Fraction of samples used
Reg_alpha/reg_lambda	Regularization to reduce overfitting

Table2. Hyperparameters

Why lags?

- VTEC_lag_1 = VTEC 15 minutes before
- VTEC_lag_4 = VTEC 1 hour before
- VTEC_lag_96 = VTEC 24 hours before

Past VTEC values provide temporal memory. They are useful because VTEC usually changes gradually over short periods.

Cross-validation/TimeSeriesSplit

- Time-based split or TimeSeriesSplit
- The model is trained on past data and tested on future data, closer to real forecasting conditions.

Two XGBoost configurations

Configuration	Examples	Why useful
Without temporal memory	Time-related, solar, GNSS geometry, geospace descriptors	Evaluate physical/geospace descriptors
With temporal memory	Same + past VTEC values	Improve 15-min short-term forecasting

Table3. XGBoost configurations

Both configurations use XGBoost. The difference is the input information.

Outcomes

Period	Model Configuration	MAE(TECU)	RMSE(TECU)	R^2
Winter	Without Temporal Memory	0,810	1,009	0,832
Winter	With Temporal Memory	0,182	0,269	0,988
Summer	Without Temporal Memory	1,367	1,605	0,172
Summer	With Temporal Memory	0,196	0,266	0,977

Table4. XGBoost VTEC forecasting performance for winter and summer periods

Summer is harder without temporal memory. Adding past VTEC strongly improves 15-minute forecasting.

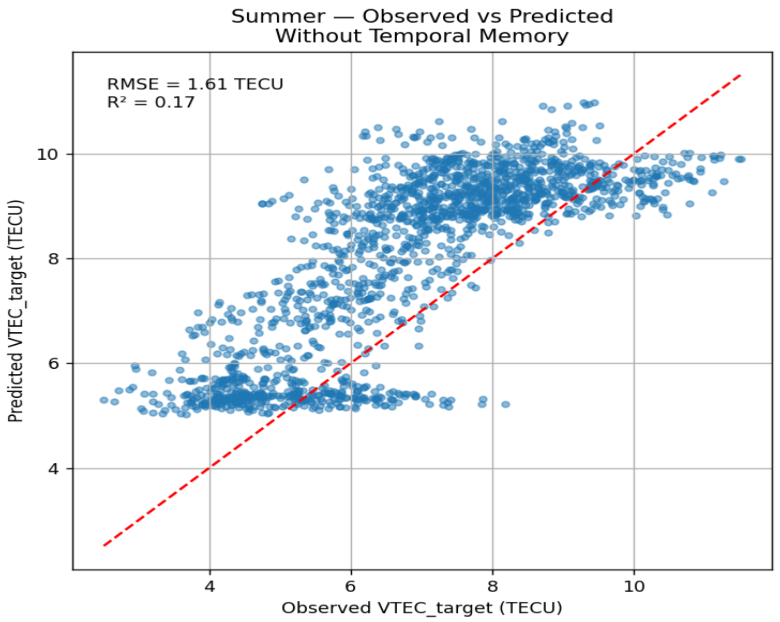


fig5. Summer prediction without temporal memory

The model captures part of the general behavior, but summer short-term variability is not fully represented by descriptors alone.

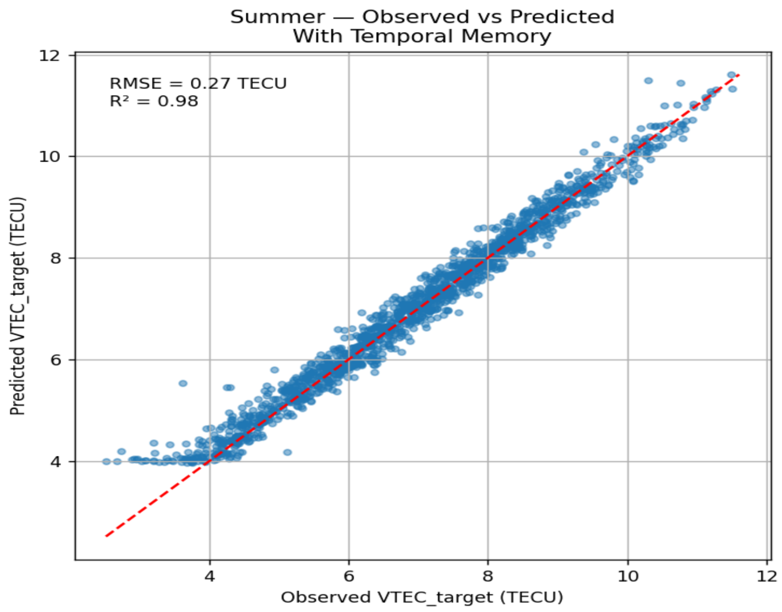


fig6. Summer prediction with temporal memory

Adding past VTEC values significantly improves the prediction, showing strong short-term persistence.

Outcomes

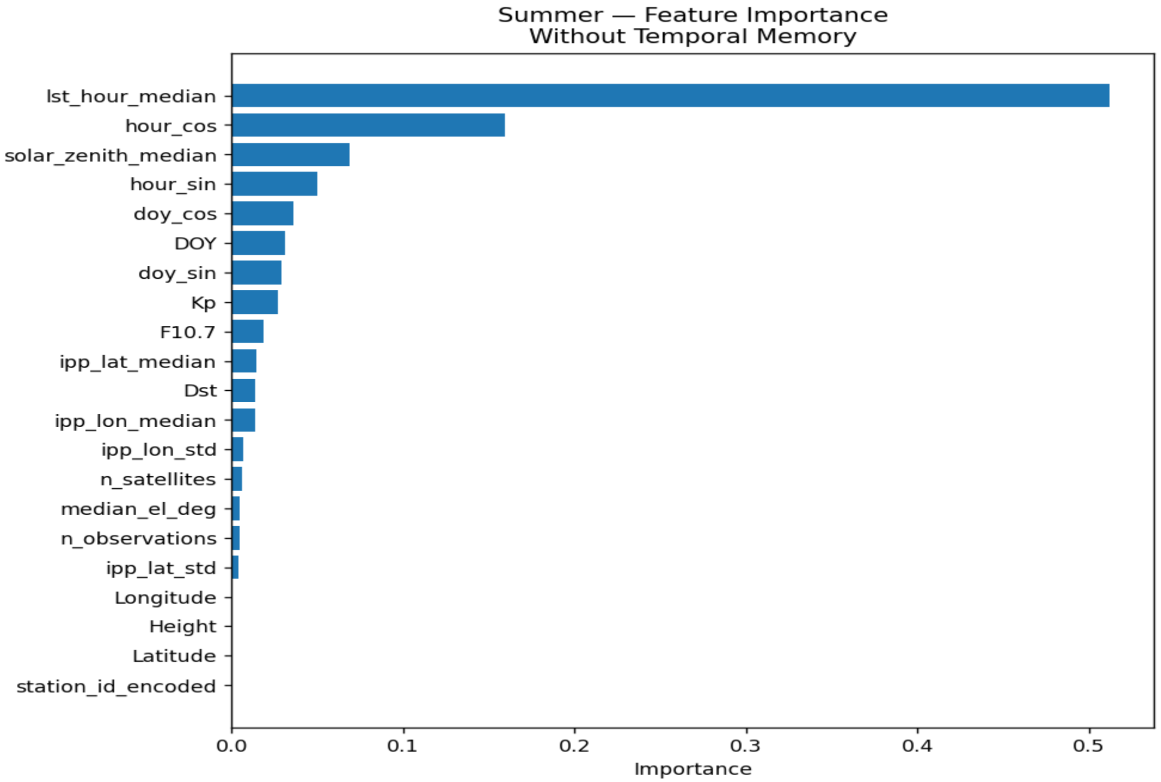


fig7. Feature importance without temporal memory

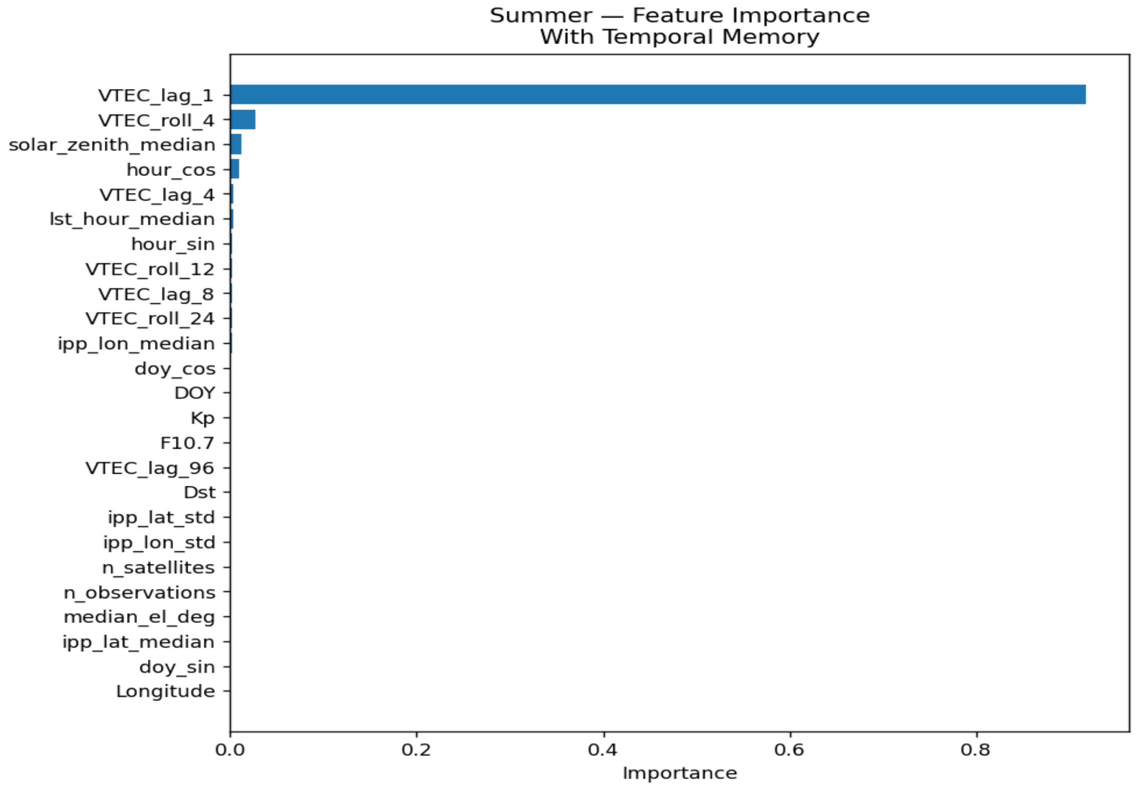


fig8. Feature importance with temporal memory

- At 15-minute horizon, recent VTEC dominates.
- However, geospace descriptors remain important for future regional, multi-station and disturbed-conditions modelling.

Current finding:

Temporal memory is very effective for 15-min VTEC nowcasting.

Scientific direction:

Future work will assess the role of geospace descriptors using more years; disturbed conditions and multiple stations.

Future work

- More years and disturbed geomagnetic conditions.
- Multiple Exagone-Teria stations.
- Spatial Validation.
- Different forecast horizons.
- Conversion from VTEC to ionospheric delay.
- Future integration into GNSS/N-RTK correction studies.

Real-time perspective

$$\boldsymbol{GNSS\ NTRIP\ stream \rightarrow real-time\ VTEC\ estimation \rightarrow lastes\ geospace\ indices \rightarrow trained\ ML\ model\ \rightarrow VTEC\ forecast \rightarrow future\ ionospheric\ correction}$$

Conclusion

- A local GNSS-derived VTEC processing and ML workflow was validated using BRST data .
- XGBoot without temporal memory captures general VTEC behavior, but summer is more challenging.
- Past VTEC values strongly improve 15-minute forecasting, while future work will focus on regional multi-station modelling and geospace-driven variability.