

## Tweedie 回帰を組み込んだアンサンブル学習による赤道プラズマバブル最大高度推定

#牛王 悠輝<sup>1)</sup>, 成合 秀飛<sup>1)</sup>, 吉野 郁海<sup>1)</sup>, 藤本 晶子<sup>1)</sup>

<sup>(1)</sup>九州工業大学

## Estimation of Equatorial Plasma Bubble Altitudes with Ensemble Learning Models Incorporating Tweedie Regression

#Yuki Goo<sup>1)</sup>, Hideto Nariai<sup>1)</sup>, Ikumi Yoshino<sup>1)</sup>, Akiko Fujimoto<sup>1)</sup>

<sup>(1)</sup>Kyushu Institute of Technology

This study focuses on equatorial plasma bubbles (EPBs), ionospheric disturbances that may cause communication failures in satellite communication and navigation systems. These bubbles are depleted regions rising from the bottom side ionosphere to higher altitudes and extending along geomagnetic field lines, with their extent proportional to the equatorial development altitude. While typical EPBs reach ~500 km, strong solar activity can drive them above 1000 km [Huang et al., 2018]. Estimating their altitude distribution is therefore vital for assessing impacts across low-to-mid latitudes.

This study aims to estimate the maximum development altitude of equatorial plasma bubbles (EPBs) based on space weather parameters. The proposed approach is formulated as a supervised regression task, where space weather parameters such as solar activity, geomagnetic activity, and interplanetary magnetic field serve as explanatory variables, and the maximum EPB development altitude is the target variable. To achieve robust prediction and capture nonlinear relationships, we employ tree-based ensemble learning models, including Random Forest, Gradient Boosting, and XGBoost. In addition, we exploit the fact that high-altitude EPBs extend along geomagnetic field lines from the equator to middle and high latitudes, introducing altitude retrieval as an inverse problem.

In the first experiment, we develop a method to estimate EPB altitudes from ROTI (Rate of TEC Index) derived from GNSS data. ROTI signatures at middle and high latitudes are projected along geomagnetic field lines to reconstruct equatorial altitude distributions. A candidate altitude  $h(t)$  is identified as the maximum altitude  $H_{upper}(t)$  when it is at most 3000 km, with at least two detections within  $\pm 10$  minutes and variance exceeding 50 km.

In the second experiment, these derived altitudes were used as training labels for machine learning models. Ten geomagnetic storm events between 2016 and 2022 were analyzed. XGBoost with a Tweedie regression loss achieved the best performance, yielding a mean absolute error of 149.7 km and an adjusted  $R^2$  of 0.58, reflecting the suitability of the Tweedie distribution for modeling the target variable.

This study demonstrates a new approach for estimating EPB altitudes by integrating GNSS-based inversion with machine learning. XGBoost with Tweedie regression shows strong potential. Future work will focus on improving prediction accuracy through model design that incorporates EPB dynamics, by refining the loss function and introducing explanatory variables that account for seasonal, diurnal, and temporal variability.

本研究は、衛星通信や測位システムに障害を引き起こす可能性のある電離圏擾乱、プラズマバブルに着目する。プラズマバブルは電離圏下部に形成された低電子密度領域が高高度に上昇し、磁力線に沿って赤道から中・高緯度へ広がる現象であり、その緯度方向の広がりには赤道域での発達高度に比例する。典型的には高度約 500 km に達するが、太陽活動が活発な時期には 1000 km を超える場合もある [Huang et al., 2018]。したがって、その高度分布を推定することは、中・低緯度での影響評価に重要である。

本研究は、宇宙天気パラメータに基づいて赤道プラズマバブル (EPB) の最大発達高度を推定することを目的とする。提案手法として、太陽活動・地磁気活動・惑星間磁場などの宇宙天気パラメータを説明変数、最大発達高度を目的変数とする教師あり回帰タスクとして設計する。さらに、非線形関係を捉え、ロバストな予測が可能な決定木ベースのアンサンブル学習手法 (Random Forest, Gradient Boosting, XGBoost) を採用する。また、本研究では「高高度まで発達した EPB が磁力線に沿って赤道域から中・高緯度へ拡大する」という事実を利用し、高度を逆問題として推定する手法を導入する。

第 1 の実験では、GNSS 全電子量 (TEC) データから算出される ROTI (Rate of TEC Index) を用いて EPB 高度を推定する手法を開発する。本手法では、中・高緯度で観測された ROTI を磁力線に沿って射影し、赤道域の高度分布を再構成する。候補高度  $h(t)$  は、 $h(t)$  が最大で 3000 km 以下であり、 $\pm 10$  分以内に 2 点以上の検出があり、かつ同時刻幅における分散が 50 km を超える場合に最大到達高度  $H_{upper}(t)$  として特定する。

第 2 の実験では、実験 1 で推定する最大到達高度を教師データとして機械学習モデルを構築する。2016 年から 2022 年に発生した 10 件の磁気嵐イベントを解析した結果、Tweedie 回帰の損失関数を組み込んだ XGBoost モデルが最も高い性能を示し、平均絶対誤差 149.7 km、自由度調整済み決定係数  $R^2$  が 0.58 を得た。この結果は、目的変数の統計的特性を表現する上で Tweedie 分布が適していたためと考えられる。

本研究は、GNSS 由来の ROTI 逆解析と機械学習を統合し、EPB の最大発達高度を推定する新しいアプローチを提示する。特に、Tweedie 回帰を導入した XGBoost モデルの有効性を示し、磁気嵐時の高度推定において有望な結果を得た。今後は、損失関数の改良や季節・日変化、時間依存性を考慮した説明変数の導入により、EPB ダイナミクス考慮したモデル設計によって推定精度のさらなる向上を目指す。