

## On the assessment of daily Equatorial Plasma Bubble occurrence modeling and forecasting

\*Brett A Carter<sup>1</sup>, Julie Currie<sup>1</sup>, Tam Dao<sup>1</sup>, Endawoke Yizengaw<sup>2</sup>, John Retterer<sup>3</sup>, Michael Terkildsen<sup>4</sup>, Keith Groves<sup>3</sup>, Ronald Caton<sup>5</sup>

1. RMIT University, 2. Aerospace, 3. Boston College, 4. Bureau of Meteorology, 5. Air Force Research Laboratory

The prediction of Equatorial Plasma Bubbles (EPBs) on a daily basis is an ongoing scientific challenge, despite decades of observations and research. Various methods for predicting the onset of EPBs have now been developed, however the research community is yet to investigate the methods for which these prediction models and techniques can be meaningfully evaluated and compared. Various assessment metrics, including percentage correct (PC), the Heidke Skill Score (HSS) and the True Skill Statistic (TSS), have been reported for space weather predictions, but the validity of their use for the purpose of EPB prediction assessment has yet to be investigated. In this study, 12 months of co-located GPS and UHF scintillation observations from locations spanning South America, Atlantic/Western Africa, Southeast Asia and Pacific sectors is used to evaluate the Generalized Rayleigh-Taylor (R-T) plasma instability growth rates calculated from the Thermosphere Ionosphere Electrodynamics General Circulation Model (TIEGCM). As part of this analysis, the limitations and caveats in using various assessment metrics are explored. In particular, the impact of employing significance testing on skill scores as a means of selecting thresholds is demonstrated. The sensitivity of the HSS, TSS and the Odds Ratio Skill Score (ORSS) to dataset type (i.e., GPS versus UHF) and dataset size (30, 50, 60 and 90 days/events) is also investigated, and it is shown that when a minimum of 50 days is used then the resulting TSS and ORSS values are similar. Further, it is shown in this analysis that to achieve statistically significant TSS and ORSS values 50 days/events are required. To investigate methods for conducting model-model comparisons, the TIEGCM R-T growth rate is compared to the ‘persistence’ forecast using the ORSS, but it is shown that the levels of 95% significance do not clearly indicate which model/method is higher quality. To that end, the concept of forecast ‘sufficiency’ is used as a direct comparison between the TIEGCM R-T growth rate the persistence forecast, and the results were found to be mixed across the different stations and EPB seasons. Importantly, throughout most of the datasets, these EPB prediction techniques were found to be of comparable quality (i.e., ‘insufficient for each other’). In this analysis, a valuable lesson learned is that it is important that the observation dataset exhibits an appropriate level of daily EPB variability in order to be used as a valid model/technique evaluation. The UHF observations from South America tended to show little daily variability in this analysis, therefore preventing any conclusive model-model comparison. Recommendations for how to meaningfully use skill scores and ‘sufficiency’ to track EPB forecast model/technique progress are given.

Keywords: Equatorial Plasma Bubbles, Ionospheric scintillations, Forecast assessment