Classifying and Predicting Equatorial Plasma Bubbles with Machine Learning

Sachin A. Reddy[1]*, Dhiren Kataria[2], Anasuya Aruliah[3], Gethyn Lewis[1], Colin Forsyth[1], and Andy Smith[1]

[1]UCL Mullard Space Science Laboratory [2]Southwest Research Institute [3]UCL Dept. of Physics & Astronomy *sachin.reddy.18@ucl.ac.uk

Introduction

Equatorial Plasma Bubbles (EPBs) are plumes of low density plasma that form in bottom side of the nightside F region. EPBs are a known cause of disruptive radio wave scintillations which can cause outages for GNSS, ground-to-satellite, and satellite-to-satellite communications.

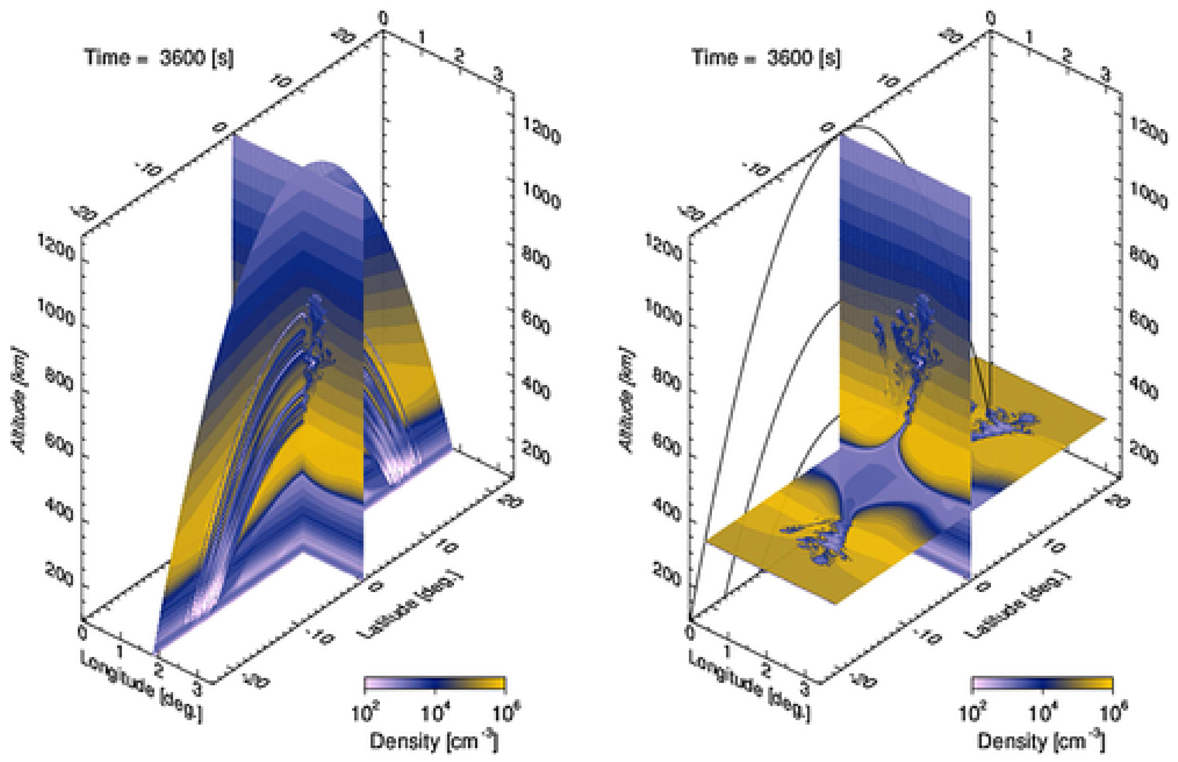

Figure 1: 3D simulation of an EPB. Reproduced from *Yokoyama, 2014* (https://doi.org/10.1002/2014JA020708)

As our dependence on space communications continues to grow, there is an ever-pressing need to better understand and predict EPBs.

The Mission

SWARM is an Earth observing constellation mission which launched into a near-polar orbit in 2013. We use the Langmuir Probe and Thermal Ion Imager on-board *Alpha* (~450km alt) to extract the ion moments and spacecraft potential to classify and predict EPBs. We use data from the 2015 equinox months.

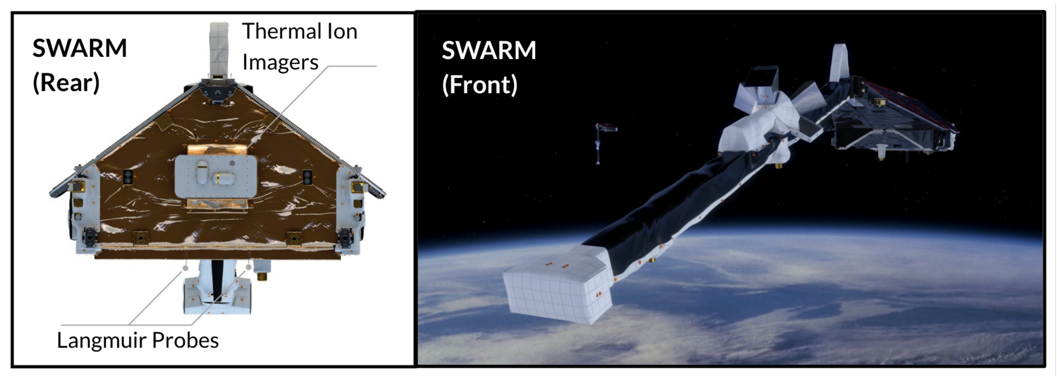

Figure 2: SWARM and its payloads. Image Credit: ESA

Classification

We develop a statistical EPB classifier to create a labelled dataset using a Savitzky-Golay smoothing filter

$$\mathcal{E}_N = \sum_{n=-M}^{M}\left(\sum_{k=0}^{N}\alpha_k n^k - x(n)\right)^2 \qquad (1)$$

where N is the polynomial order which is equal to 2, M is the window size which is set to 11, $x(n)$ is the number of samples and α are the coefficients to be optimised.

The S-G filter identifies EPBs as noise, which we then apply a threshold to: if the residual amount is $>$ 10,000 it is flagged as an EPB. We also add additional filters to reduce the false positive and negative rate. The full pipeline from *SWARM data → machine learning ready* is show below.

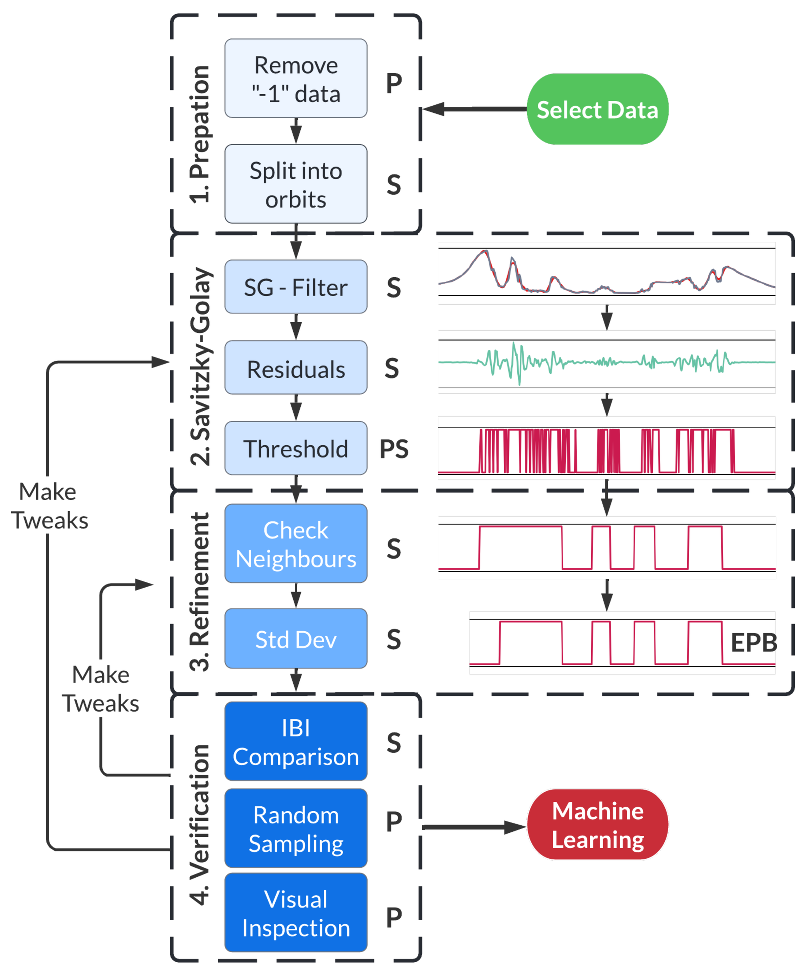

Figure 3: High-level overview of our EPB classification process. P = Physics based, S = Statistically based.

We assess the accuracy of the classifier by randomly sampling ~5% (n) of the total population. Assuming an error of $\sqrt{n}$, the accuracy of the classifier is 72% - 90%. The existing IBI processor on-board SWARM scores 66% - 83% under the same conditions.

Prediction

There are 858k instances in the training set with 43k of these tagged as an EPB. This represents a class imbalance of 20:1 (No EPB: EPB) which we redress with the SMOTE algorithm. We select four features from the SWARM dataset: longitude (∘), spacecraft potential (V), number density (cm^{-3}), ion temperature (K).

Random Forest classifiers have been shown to perform well on image classification tasks when the noise levels are below 40%. The worst case noise for our method is 28% (100% - 72%). Through trial-and-error we find that *train:test = 9:1*, and *number of trees = 175*, produces the best model performance.

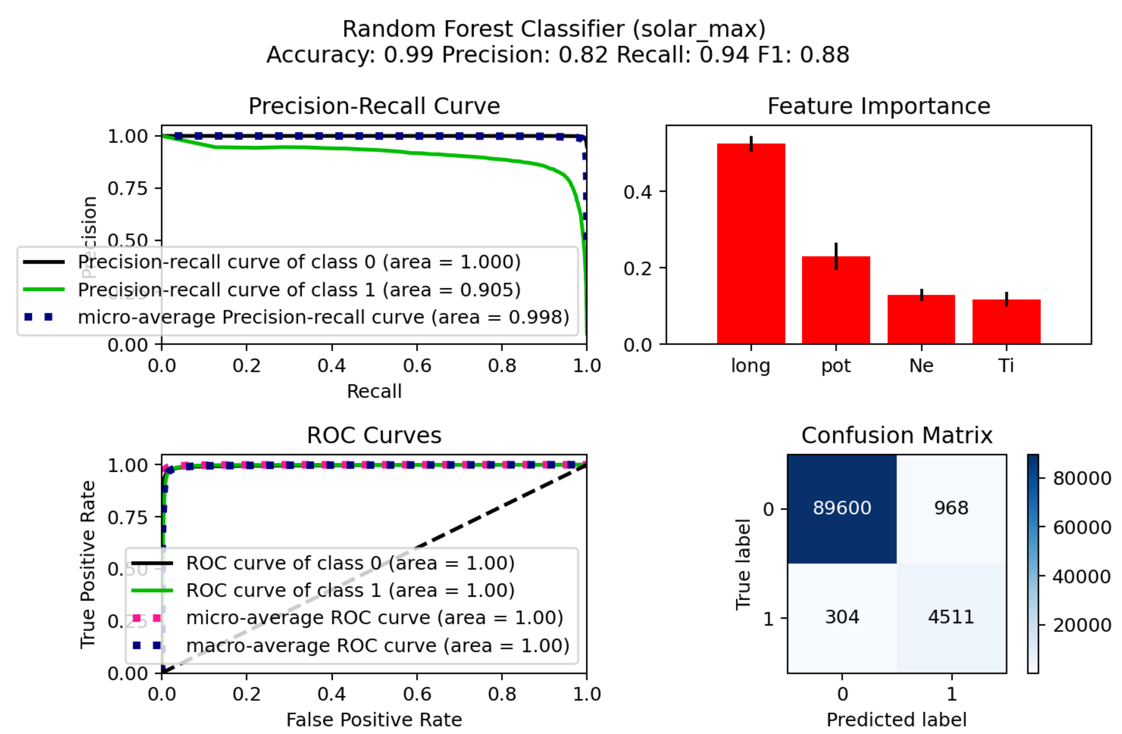

Figure 4: Results from the Random Forest model. No EPBs = 0 and EPBs = 1.

As our objective is to detect EPBs (minimize false negatives), *recall* is an important metric and we deem 94% to be acceptable. Although an *accuracy* of 99% is well received, it is not always the best reflection of actual performance. Instead, we chose the F1 score as our overall performance metric which at 88% is considered satisfactory.

The results also show that long(itude) is the most important feature. This aligns with previous non-ML work which has long shown the longitudinal dependence of EPBs.

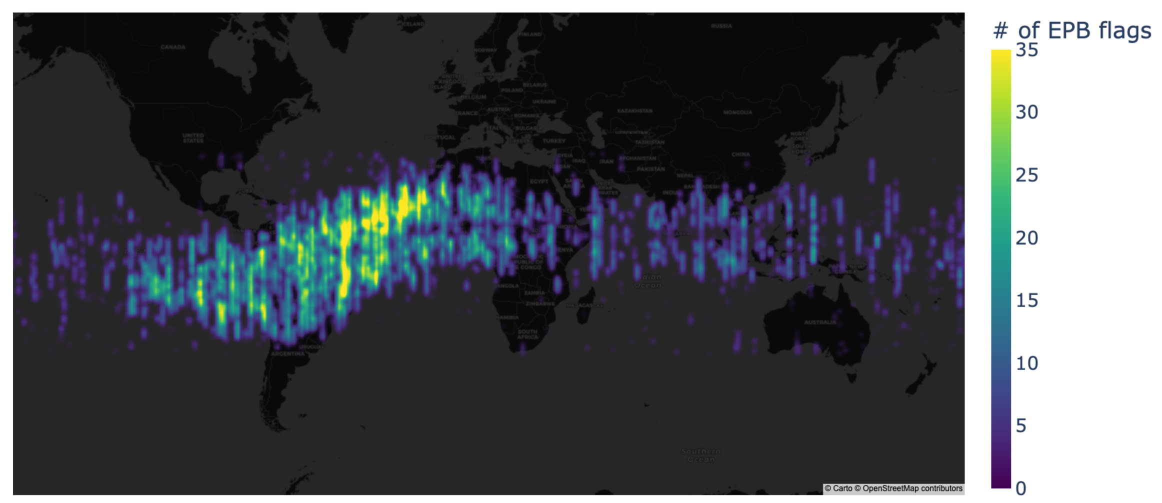

Figure 5: EPBs events identified by our classifier. The distribution of EPBs is in-line with previous work.

Figure 4 also highlights the importance of spacecraft potential (pot). As pot is a response to plasma density, electron temperature, and the spacecraft itself, it can be thought of as naturally occurring 'created feature'. We speculate that this is why the feature importance is higher than the plasma number density (Ne). The use of spacecraft potential to predict EPBs does not appear to have been reported previously.

Future Work

We are planning to integrate the Kp-index and solar flare data with the SWARM dataset as EPBs are linked to solar and geomagnetic activity.

Secondly, we are expanding the pipeline to include *Charlie* which is the other SWARM spacecraft at ~ 450*km* alt. This will double the temporal resolution of EPBs and should improve the classification accuracy.

Summary

We build a classifier to detect EPBs using SWARM data. We find it could be 6% more accurate than the existing processor. We then train a Random Forest classifier to predict EPBs and find that spacecraft potential is an important feature. Future work will add more features to improve prediction performance.