XGBoost Algorithm for Estimating Equatorial Plasmaspheric Mass Density Using ULF Wave Measurements

R. Foldes¹†, A. Del Corpo¹, E. Pietropaolo¹ and M. Vellante¹

¹University of L’Aquila, Department of Physical and Chemical Sciences
† raffaello.foldes@graduate.univaq.it
The inner region of the Earth’s magnetosphere, is composed by by dense and cool (~ 1keV) plasma of ionospheric origin: it extends approx from topside of the ionosphere to 6 $R_E$, and it co-rotates with the Earth.

**Standing waves** can be excited by MHD compressive waves. This coupling produce Field Line Resonance (FLR) frequencies along a specific geomagnetic field line, which can be used to estimate plasmaspheric mass density.

Assuming: 1. a geomagnetic model 2. a functional form for the density

**FLR frequency** \( \rightarrow \) **equatorial plasma density**

Solving Singer wave equation (1981) we get the following relation:

\[
\rho_{eq} = \frac{\lambda B_{eq}^2}{4\pi^2 \mu_0 l^2 f_r^2}
\]
State of the art: **Gradient method** (Waters et al., 1991) from ground-based magnetometers (ULF measurements)

Assuming that:
1. Eigenfrequency linearly decreases poleward for stations slightly separated in latitude (this is not true passing through the plasmapause)
2. The stations are aligned along a geomagnetic meridian.
3. There are couples of stations sufficiently close to each other (1°-3°).

The FLR frequency can be inferred by computing the discrete Fourier cross-spectrum of two signals where:
1) **Cross-Phase** has its maximum (minimum) value
2) **Cross-Amplitude** crosses unity with positive (negative) slope
EMMA provides magnetic measurements with a resolution of 1s.

Real-time monitoring of the plasmasphere dynamics

Many authors created (semi-)automated tools for observing the plasmasphere via FLRs (Del Corpo et al., 2018; Wharton et al., 2018; Lichtenberger et al., 2013; Berube et al., 2003; Chi et al., 2013).

All the current methods rely on the cross-phase technique.
The data set created by Del Corpo et al. (2019) contains cross-phase spectra and validated FLR frequencies (first harmonic) with a time resolution of 30 mins. and an average relative error $\Delta f/f$ at any latitude. The fundamental frequencies range from few $mHz$ (MUO-PEL) to about 60 $mHz$ (SUW-BEL).

- 4 station pairs (SUW-BEL, TAR-BRZ, OUJ-HAN and MUO-PEL)
- 165 non-consecutive days (between 2012 and 2017)
- 13 geomagnetic storms (e.g. St. Patrick’s day storm, 2013)
- several different geomagnetic conditions
- about 4000 samples per stations pair

$\phi_{HRZ} - \phi_{AR}$

Input Matrix

$\begin{bmatrix}
    x_1^1 & x_1^2 & \cdots & x_1^M \\
    x_2^1 & x_2^2 & \cdots & x_2^M \\
    \vdots & \vdots & \ddots & \vdots \\
    x_N^1 & x_N^2 & \cdots & x_N^M
\end{bmatrix}$
We compute the **mutual information** between each input feature (cross-phase values) and the FLR frequency (output). We want to reduce the initial data set dimensionality to reduce the overfitting error maintaining all the information necessary to reach the best algorithm performance.

For each stations pair we obtain 95% of information with 80% of the spectral length.

<table>
<thead>
<tr>
<th># of features</th>
<th>Before</th>
<th>After</th>
<th>$f_{cut}$ (mHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUW-BEL</td>
<td>212</td>
<td>170</td>
<td>94</td>
</tr>
<tr>
<td>TAR-BRZ</td>
<td>212</td>
<td>170</td>
<td>94</td>
</tr>
<tr>
<td>OUJ-HAN</td>
<td>285</td>
<td>230</td>
<td>64</td>
</tr>
<tr>
<td>MUO-PEL</td>
<td>285</td>
<td>230</td>
<td>32</td>
</tr>
</tbody>
</table>

Foldes et al., DeepFLRID – EGUv2021
Data Selection & Data Preprocessing

- Data and feature selection
- Input/Output Transformation
- Train/Test splitting

Model Training

- XGBoost hyper-parameter tuning
- Cross-validation training with a different model for each pair of stations

Model Evaluation

- Evaluation of model performances on the test set (unseen data)
- The test set is overlaid for the 4 station pairs and it contains different geomagnetic conditions

Results Assessment and Comparison

- Analysis of FLRs estimation per latitude and geomagnetic conditions
Results 1: Global Results

**Day**: both field line footpoints sunlit  
**Penumbra**: only one footpoint sunlit  
**Night**: no footpoints sunlit  
→ Condition evaluated at an height of 120km

Scatter plot of the entire test set for the four station pairs. Points are the median value of the algorithm estimations and the error bar is the interquartile range:

- Algorithm performance shows high results for all the station pairs, especially for daytime frequencies
- Most distant points from the bisector usually have a greater error bar
- At every $L$ we can observe that the error does not increase with increasing frequencies
The case study refers to the geomagnetic storm of the 1\textsuperscript{st} June 2013.

DoY 151 is the day before the main phase of the storm (DoY 152). In the plot to the right is shown a detail of the FLRs estimation (white circles) with respect to the actual frequencies (magenta circles).

Except for MUO-PEL which shows some points far away from the validated frequencies, the estimation error does not show any dependence from the geomagnetic activity level.
By using the relation obtained from the Singer equation we compute the plasmaspheric mass density from both the validated and the estimated frequencies.

A detail of plasma mass density estimations for DoY 151 and DoY 152.

• They significantly decrease passing from a quiescent day to a geomagnetic disturbed one.

• Passing from daytime to nighttime (light and dark gray areas) values (and vice versa) we can often observe a sharp transition which is not physically meaningful.

Nighttime frequencies (and thus densities) are often questionable since they generally do not satisfy the approximation of infinite conductivity necessary to have the fixed field line footprints needed to sustain Alfvén standing waves.

Foldes et al., DeepFLRID – EGUv2021
Results 4: Latitudinal Dependence

Relative estimation error of the four pairs of stations for six consecutive days. In the top panels $Kp$ and $Dst$ indices show the evolution of the geomagnetic storm. Light and dark gray areas correspond to time when one or both footpoints are nightside.

• Error slightly increases with increasing $L$ probably because of fuzzier cross-phase spectra

• Average relative error is +1-2% from $L=2.4$ to $L=4.1$, for MUO-PEL is 4.5% meaning that overestimation errors have a heavier weight.

• We can observe more in detail that estimations do not depend on the level of geomagnetic activity, at least at low (and mid) latitudes.

Foldes et al., DeepFLRID – EGUv2021
Discussion and Future Works

Conclusion and next steps…

• Machine Learning algorithms (especially supervised ensemble methods with a feature-based approach) resulted a powerful tool for estimating FLRs from cross-phase spectra.

• The algorithm performances showed a little dependence on the station latitude, but it is worth noting that the estimation error remains small even during highly disturbed geomagnetic conditions (Space Weather tool for monitoring the plasmasphere dynamics).

• In order to obtain more robust models/predictors it is necessary to train the algorithms on a larger data set and using more stations along the EMMA network.

• This is only a preliminary result for evaluation purposes. To create a completely automated tool we need for an additional step which recognizes the presence of a FLR in the signals.
Input and Output Data Sets for this work:

- Input data: https://doi.org/10.5281/zenodo.4304662
- Output data: https://doi.org/10.5281/zenodo.4304911

[1] D. Berube et al., An automated method for the detection of field line resonance frequencies using ground magnetometer techniques, 2003
[3] L. Baransky et al., High resolution method of direct measurement of the magnetic field lines eigenfrequencies, 1985
[4] A. Del Corpo et al., Observing the cold plasma in the Earth's magnetosphere with the EMMA network, 2019
[6] H. Singer et al., Alfvén wave resonances in a realistic magnetospheric magnetic field geometry, 1981
[8] N.A Tsyganenko and M.I. Sitnov, Modelling the dynamics of the inner magnetosphere during strong geomagnetic storms, 2005
[9] C.L. Waters et al., The resonance structure of low latitude Pc3 geomagnetic pulsations, 1991
[10] S.J. Wharton et al., Cross-phase determination of ultralow frequency wave harmonic frequencies and thei associated plasma mass density distributions, 2018