

JGR Space Physics

RESEARCH ARTICLE

10.1029/2020JA029008

Special Section:

Probing the Magnetosphere through Magnetoseismology and Ultra-Low-Frequency Waves

Key Points:

- We investigated the performances of several machine learning algorithms to identify geomagnetic Field Line Resonance (FLR) frequencies from ultralow frequency signals detected at European quasi-Meridional Magnetometer Array
- Tree-based methods have higher performance than kernel-based methods and they result a robust technique to estimate FLR frequencies
- The estimation error of our approach shows a dependence on the station latitude, but it is stable on highly variable geomagnetic conditions

Correspondence to:

R. Foldes,
raffaello.foldes@graduate.univaq.it

Citation:





Foldes, R., Del Corpo, A., Pietropaolo, E., & Vellante, M. (2021). Assessing machine learning techniques for identifying field line resonance frequencies from cross-phase spectra. *Journal of Geophysical Research: Space Physics*, 126, e2020JA029008. <https://doi.org/10.1029/2020JA029008>

Received 7 DEC 2020
Accepted 4 MAY 2021

© 2021. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Assessing Machine Learning Techniques for Identifying Field Line Resonance Frequencies From Cross-Phase Spectra

Raffaello Foldes¹ , Alfredo Del Corpo¹ , Ermanno Pietropaolo¹ , and Massimo Vellante¹ 

¹Department of Physical and Chemical Sciences, University of L'Aquila, L'Aquila, Italy

Abstract Monitoring the plasmasphere is an important task to achieve in the Space Weather context. A consolidated technique consists of remotely inferring the equatorial plasma mass density in the inner magnetosphere using Field Line Resonance (FLR) frequencies estimates. FLR frequencies can be obtained via cross-phase analysis of magnetic signals recorded from pairs of latitude separated stations. In the last years, machine learning (ML) has been successfully applied in Space Weather, but this is the first attempt to estimate FLR frequencies with these techniques. We survey several supervised ML algorithms for identifying FLR frequencies by using measurements of the European quasi-Meridional Magnetometer Array. Our algorithms take as input the 2-hour cross-phase spectra of magnetic signals and return the FLR frequency as output; we evaluate the algorithm performance on four different station pairs from $L = 2.4$ to $L = 5.5$. Results show that tree-based algorithms are robust and accurate models to achieve this goal. Their performance slightly decreases with increasing latitude and tend to deteriorate during nighttime. The estimation error does not seem to depend on the geomagnetic activity, although at high latitudes the error increases during highly disturbed geomagnetic conditions such as the main phase of a storm. Our approach may represent a prominent space weather tool included into an automatic monitoring system of the plasmasphere. This work represents only a preliminary step in this direction; the application of this technique on a more extensive data set and on more pairs of stations is straightforward and necessary to create more robust and accurate models.

1. Introduction

The near-Earth environment is composed of different plasma populations that cover an energy range of several orders of magnitude. The plasmasphere is the coldest one (~ 1 eV) and typically extends from the top of the ionosphere up to 4–6 Earth radii (R_E). Its shape, composition and extension change in response to geomagnetic activity variation and, due to its preponderant contribution to the mass density, plays a crucial role in exciting plasma-waves and driving their interaction with more energetic particles (Liemohn, 2006; Thorne, 2010). During major storms the outer boundary, the plasmopause, also defined as the plasmasphere boundary layer (PBL, Carpenter & Lemaire, 2004), can move earthward as near as 1.5–2 R_E , and the recovery to the pre-storm conditions can last for several days.

For its central role in the space weather context, a real-time monitoring of the cold plasma in the inner magnetosphere would be an essential task to achieve. A prototype of such system was developed inside the PLASMON FP7 European project (Lichtenberger et al., 2013) during which the European quasi-Meridional Magnetometer Array (EMMA) was established. EMMA provides a unique opportunity to monitor the plasmasphere dynamics in near real-time. It consists of 27 ground-based magnetic stations approximately located along the same magnetic meridian and mapping into magnetic L -shells in the range 1.6–6.1, where L is the McIlwain parameter evaluated using the IGRF. Using Ultra Low Frequency (ULF) signals simultaneously observed by pairs of stations it is possible to infer the radial distribution of the equatorial plasma mass density in the longitudinal sector identified by the network (Del Corpo et al., 2019; Lichtenberger et al., 2013).

The core of this monitoring system is the gradient method proposed by Baransky et al. (1985) and further developed by Waters et al. (1991): By performing Fourier cross-spectral analysis of the magnetic signal recorded by two stations slightly separated in latitude, it is possible to estimate the resonance frequency of the field line whose footprint is halfway between the stations. Around the field line resonance (FLR) frequency and its higher harmonics the cross-phase presents a maximum and the amplitude ratio crosses one with a positive slope. One or both these characteristics can be used to identify FLR frequencies which, in turn, depend on the magnetic field strength and plasma mass density distribution along the field line. By making assumptions on the magnetic field

topology and the field line plasma distribution, the governing wave equation (Singer et al., 1981) can be numerically solved, providing an estimate of the plasma mass density in the equatorial crossing point of the field line.

Using EMMA observations, Del Corpo et al. (2019) created a data set of validated frequencies considering only the fundamental harmonic and employing a semi-automated algorithm which performs the above procedure. They used the TS05 model (Tsyganenko & Sitnov, 2005) to represent the geomagnetic field lines and assumed a power-law dependence of the plasma mass density along them. The validation procedure takes care of possible failures of the automated part. One of the most critical situations concerns the detection of FLR frequencies at high latitudes. They can be very close to the spectral resolution, so the fundamental harmonic could be hard to detect, especially during prolonged quiet period, when the FLR frequency usually decreases. In this case higher harmonics could be accidentally picked up by a fully automated procedure (Del Corpo et al., 2019). Other issues can emerge when a pair is sounding the region near the PBL where the selected FLR frequency can be mismatched or can even disappear (Fraser et al., 2005; Menk et al., 2004; Milling et al., 2001).

Based on the gradient method, mentioned above, a few authors in the past developed algorithms to automatically detect FLRs (Berube et al., 2003; Chi et al., 2013; Lichtenberger et al., 2013; Wharton et al., 2018). In these works, either statistical or morphological properties of the cross-phase spectra were used to individuate the resonance peaks. In this study, we propose an alternative approach based on machine learning methods, using the cross-phase spectra by Del Corpo et al. (2019) as input data and their data set of validated frequencies to train the algorithms.

Machine learning techniques represent a subset of artificial intelligence and they have turned into one of the most powerful tools for addressing space weather problems. Since the early 90s machine learning has benefited from the high amount and quality of both space and ground-based data. Nowadays machine learning methods are applied to the most varied space weather topics, and they can be identified according to their final goal into four macro categories (Camporeale et al., 2018): (i) algorithms for event identification, such as solar image classification (Armstrong & Fletcher, 2019), near-Earth plasma regions classification (Breuillard et al., 2020) and the classification of periods with ULF waves activity (Balasis et al., 2019); (ii) methods for revealing causality between high dimensional data sets and specific events, see Wing et al. (2018) and Heidrich-Meisner and Wimmer-Schweingruber (2018); (iii) forecasting algorithms, widely used for predicting solar flares (Massone et al., 2018), the arrival time of coronal mass ejections (Liu et al., 2018) and the behavior of various geomagnetic indices (Chandorkar & Camporeale, 2018); and (iv) algorithms for modeling non-linear relationships which try to reveal the physical action of the system starting from first principles (Boynton et al., 2018).

Despite the wide application, machine learning techniques were applied to the characterization of plasmasphere regions only in the past few years. Main contribution is from Zhelavskaya et al. (2016, 2017) who applied a convolutional neural network for estimating electron mass density. To the best of our knowledge, the only attempt of estimating FLR frequencies using a ML approach has been developed by Fujimoto et al. (2019) for a pair of stations in New Zealand (mid-latitude).

In this work, we assess various machine learning methods for addressing the challenging task of identifying FLR frequencies by using ground-based measurements.

In Section 2 we briefly introduce the machine learning algorithms adopted for this analysis. Section 3 widely explains the set of data obtained from the EMMA network. Section 4 focuses on pre-processing, feature and model selection procedure which compose the whole machine learning process. Finally, in Section 5 we discuss the results and in Section 6 we outline the significant breakthrough and development produced by this analysis.

2. Machine Learning Algorithms

Machine Learning (ML) algorithms can be principally divided into three macro categories: supervised (labeled data set y_i), unsupervised (not labeled data set y_i) and reinforcement learning (goal-oriented algorithms). Since our analysis is the first attempt to identify ULF wave resonance frequencies from ground-based measurements using a ML approach, we assess several different algorithms, widely used for regression problems, that is, identification and/or prediction. A general regression problem within this framework can be expressed as the estimation of a real-valued function $F(X)$, where X is a d -dimensional input variable defined by an input/output pair (X_i, y_i) . It is possible to divide regression algorithms into three main groups: classification and regression trees

(CART, i.e., Decision Tree by Breiman et al., 1984), ensemble methods (i.e., Random Forest and Gradient Boosting algorithms) and kernel methods (linear methods using the “kernel trick,” that is, Support Vector Machines and Kernel Ridge).

2.1. Kernel Methods

Ridge regression employs ordinary least squares (OLS) for minimizing the error between predictions \hat{y}_i and targets y_i , by adding a penalty term ($\alpha \geq 0$) that controls the complexity of the learned model (Hoerl & Kennard, 1970).

Kernel Ridge Regression (KRR) combines such a described method with the kernel trick (Theodoridis & Koutroumbas, 2008). Kernel functions transform data into higher dimensional space, by computing the inner product in the input space. There exist several different kernel functions; the Gaussian radial basis function (RBF), polynomial and sigmoid kernels are the most used, just to cite a few.

As well as KRR, several ML algorithms can operate with the kernel method; among them, *support vector machines* (SVMs, Cortes & Vapnik, 1995) are the best known, though they were initially used for classification.

SVM models employ the so called *support vectors* (i.e., data points nearest the division boundary) for identifying the optimal hyperplane in the feature d -dimensional space, which divides the points by maximizing the margin around the hyperplane itself. Because of its universality and robustness many variants were developed (see Kumar et al., 2019, for a review), and they have been widely used also for regression problems (Bao et al., 2020; Yang et al., 2009). In particular, the so-called ϵ -SVM addresses regression problems by employing an ϵ -range which represents the approximation region of the real-valued target function.

After the training process SVM and KRR have learned the same model shape, though they use different cost functions, that is, quadratic for KRR and ϵ -insensitive for SVM. Another difference between these two algorithms is the execution time; since KRR can be computed in a closed form, it results faster with medium-sized (less than 10^4 samples) data sets, but SVM can learn sparse models resulting in faster performance with massive sets (Murphy, 2012).

2.2. Decision Trees

Differently from single-stage algorithms, like those described in the previous section, decision trees (DTs) are multi-step models which recursively divide data in two partitions (binary split). Starting from the root node, containing all data, the tree structure is composed by intermediate nodes (called splits) and terminal nodes (leaves).

The process involves two macro phases: the splitting procedure and the pruning phase. The main objective of pruning techniques (see Esposito et al., 1997, for a review) is to generalize the tree structure and hence avoid over-fitting, see Figure 1. Pruning techniques can be controlled by various parameters which drive the growth of DTs, such as the minimum number of samples in a split/leaf or the maximum number of features to consider for the best split.

The splitting procedure determines the growth of a DT; at each iteration the central operation is finding the best split for the data set according to some error measures (Buntine & Niblett, 1992). The tree growth continues until no more splitting produces a significant gain in the error measure or a stopping criterion is satisfied. DTs represent optimal algorithms in processing massive data sets, handling missing values and mixed data (ordinal, categorical and continuous), performing features selection, ignoring redundant features, but even for their user-friendly final structure.

One of the major shortcoming is a poorness of generalization which make this kind of algorithms prone to over-fitting and highly unstable with respect of slight variation in the input data (Li & Belford, 2002). This error, namely variance error, can be addressed using ensemble (or “forest”) models which combine results from numerous trees (usually averaging, i.e., Random Forest), or combine various DTs during the growth phase (i.e., Gradient Boosting methods).

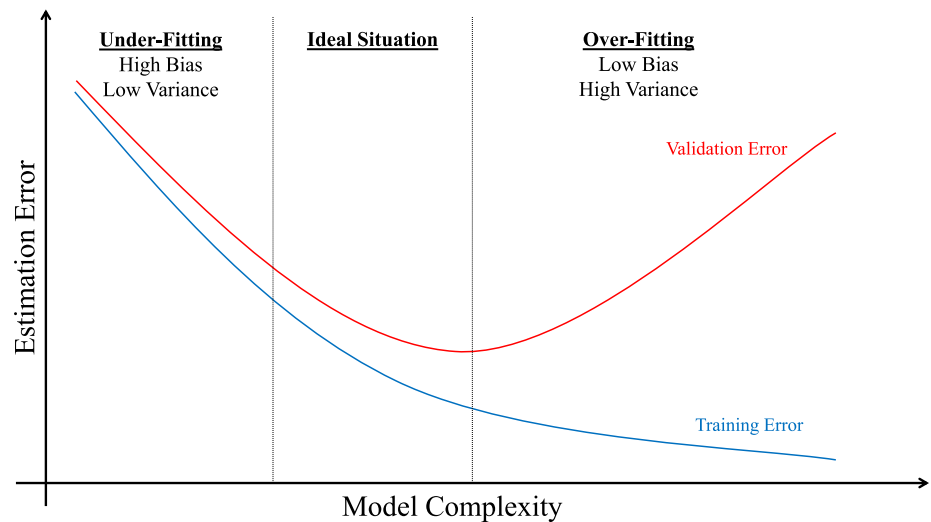


Figure 1. Illustration of a typical scenario occurring during the procedure of hyperparameter tuning for every machine learning method. As the model complexity increases the estimation error for the training set decreases; still, the model parameters are too specific and hence they are not able to reproduce the same high performance with unseen data (i.e., validation set), namely over-fitting.

2.3. Ensemble Methods

The central concept besides ensemble methods is that multiple weak estimators can achieve better results than a single complex model; they result more efficient than DTs in reducing the variance error (Breiman, 1996).

Let us consider many shallow DT regressors (usually hundreds) which give K solutions of a real-valued function, $\hat{F}_1(X), \hat{F}_2(X), \dots, \hat{F}_K(X)$. We can build the final prediction using a linear combination of the estimation:

$$\hat{F}_{Ens}(X) = \sum_{j=1}^K a_j \hat{F}_j(X)$$

where the coefficients, a_j , determine the type of procedure adopted in constructing the ensemble. Two main approaches are commonly used: bootstrap aggregating (or Bagging) methods use the average as an estimator, that is, $a_j \equiv 1/K$, and boosting techniques, which employs coefficients dependent on the number of estimators (Buhlmann, 2012).

Similarly to bagging methods, the Random Forest (RF) algorithm averages the solutions of many estimators; still, it considers for splitting data only a subset of features, $d < D$, producing highly varied estimators and resulting in faster performance with high dimensional data sets. In RF, each tree grows independently, while boosting algorithms create one tree at a time to improve the prediction of the previous model recursively. Usually, RF obtains better performance than bagging methods but worse than boosting ones.

A wide group of boosting methods are those related to the gradient descent technique, widely known as Gradient boosting models. Considering, for instance, a quadratic loss function $L = \sum (y_i - \hat{F}(x_i))^2 / 2$, these methods act to minimize L by adding further estimators according to the value of the negative gradient, $r(x_n) = y_n - F(x_n) = -\partial L / \partial F(x_n)$, which represents how far the model is from a reliable estimation. Gradient boosting algorithm can be directly generalized to different loss functions, without modifying any part of their procedure.

In this analysis, we focus our attention on two particular boosting algorithms: eXtreme Gradient Boost (XGB, Chen & Guestrin, 2016) and Light Gradient Boost Model (LGBM, Ke et al., 2017). Both methods have shown in the last years to address regression/classification problems with high and robust performance.

More in detail, XGB is a recent algorithm resulted in being highly efficient in regression/classification competition (see www.kaggle.com) and many real-world applications (Ivanov et al., 2020; Kunte & Panicker, 2020;

Luckner et al., 2017). XGB employs second-order derivative during gradient descent resulting in faster and more refined performance. It results particularly reliable in handling massive data sets because of its parallel implementation. Despite the numerous novelties and improvements of XGB, this algorithm can result barely efficient with high dimensional data set, because of its splitting method which turns out very time-consuming.

For speeding up the splitting procedure, LGBM combines two additional methods, namely Gradient-based One-Side Sampling (GOSS) and Exclusive Features Bundling (EFB). GOSS, in combination with EFB, picks (usually randomly) only a small portion of data having small gradients for estimating the information gain, since these points give minor contributions to the gradient descent. LGBM, in comparison to XGB, can simply handle with categorical data without requiring any pre-processing phase.

3. Data

The plasmaspheric mass density can change drastically during active geomagnetic periods and the associated field line resonance frequencies change accordingly. Therefore, a good data set, suitable to train ML algorithms, should cover the most comprehensive range of geomagnetic disturbance levels. The data set by Del Corpo et al. (2019) was created following this idea and covers 165 days along 5 years: 2012 (DoY 266–336), 2013 (DoY 72–86 and 145–162), 2014 (DoY 45–68), 2015 (DoY 72–90 and 169–178), and 2017 (DoY 146–153). It comprises prolonged quiet geomagnetic periods as well as disturbances of different intensity, including 13 geomagnetic storms, nine of them having a minimum *Dst* value less than -100 nT. For each period, the analysis has been carried out over a variable number of station pairs, depending on the data coverage. About 20 pairs were analyzed for each period, but only eight were common to all the data set.

This work aims to test the viability of a machine learning algorithm to pick up trusted FLR frequencies from magnetic field cross-phase spectra collected by a network like EMMA. Since the frequency generally decreases with latitude, such algorithm should be tested not only in different geomagnetic conditions, but also for data collected at different latitudes. For this reason, we selected four pairs among the eight common to the entire data set, that cover at the same time the largest range of latitudes. Figure 2 shows the position of the selected pairs inside the EMMA network: Suwalki-Belsk (SUW-BEL) in black, Tartu-Birzai (TAR-BRZ) in green, Oulujärvi-Hankasalmi (OUJ-HAN) in red and Muonio-Pello (MUO-PEL) in blue. More details about coordinates and the *L*-parameters are reported in Table 1.

The data set consists of daily Fourier cross-phase spectra computed with a 2-h moving window every half an hour (Del Corpo & Vellante, 2020). Each 2-h window is divided into a variable number of sub-intervals that depends on the latitude (increasing from 1 to 7 with decreasing latitude). The spectra are evaluated for each sub-interval and successively averaged. This implies a spectral resolution that increases with increasing latitude, a condition necessary to detect FLR frequencies that vary by more than a factor of 10, moving from the lowest to the highest latitude station pair. Each final spectrum is smoothed over 9–11 samples and the fundamental resonance frequency is eventually provided following a selection and validation process composed of both automated and manual steps.

The fundamental resonance frequencies are provided with a time resolution of 30 min, although the coverage is not continuous. In fact, for several reasons, it is not always possible to detect FLR frequencies from cross-phase spectra: (i) data gaps in one or both stations; (ii) noise in the magnetic signals that inhibits the appearance of the resonance signatures in the spectra; (iii) quiescent LF wave periods that prevent the excitation of standing Alfvén waves (Balasis et al., 2019). Statistical works have shown that the FLR detection rate changes also with latitude and local time with a broad peak around noon; in particular, during nighttime it can be as low as 10% (Chi et al., 2013; Del Corpo et al., 2019).

To efficiently sustain a standing Alfvén wave, the field line should be fixed at the footprints, which implies an infinite conductivity. This approximation is generally satisfied during daytime, when the ionospheric conductivity is sufficiently high, but after sunset the conductivity decreases due to the predominant role of recombination processes and it should be assumed finite (Allan & Knox, 1979). Whether the ionosphere conditions are able to sustain normal modes also during nighttime, is still under debate. If the conductivity is sufficiently small, the observed modes should be interpreted as free-end ones. Recently Takahashi et al. (2020) observed FLRs across the midnight sector compatible with a fixed-end mode configuration. In some cases, only one footprint is

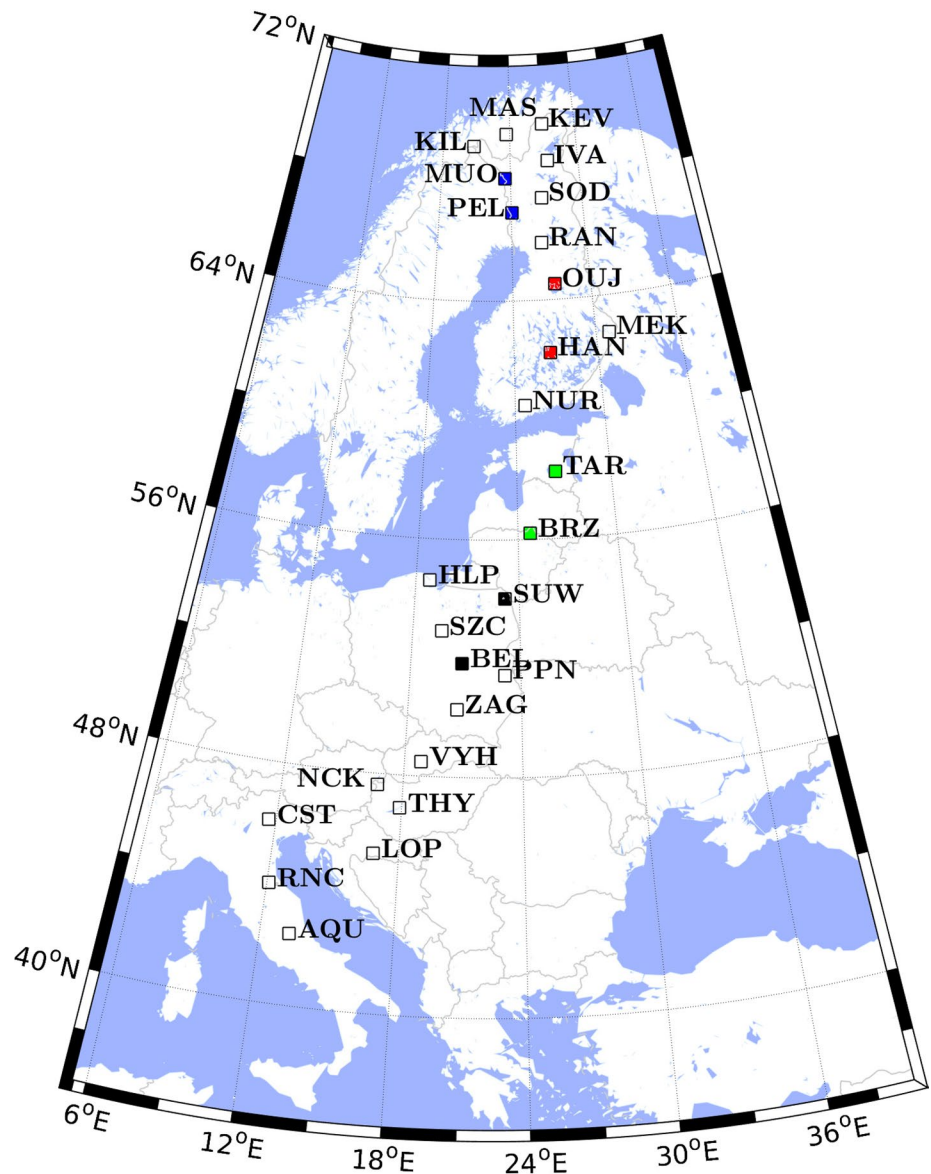


Figure 2. The European quasi-Meridional Magnetometer Array network map in geographic coordinates. The four pairs used in this analysis are highlighted in different colors.

Table 1

Station Pairs of the EMMA Network Considered in This Work With Their Mid-Point Geographic Coordinates and the Corresponding McIlwain Parameter

Pair code	Stations name	Latitude (deg.)	Longitude (deg.)	<i>L</i>
MUO-PEL	Muonio-Pello	67.5	23.8	5.5
OUI-HAN	Oulujärvi-Hankasalmi	63.4	26.9	4.1
TAR-BRZ	Tartu-Birzai	57.2	25.6	2.9
SUW-BEL	Suwalki-Belsk	52.9	22.0	2.4

Abbreviation: EMMA, European quasi-Meridional Magnetometer Array.

sunlit while the other is in darkness, generating quarter-wave modes (Obana et al., 2008, 2015). These situations make less reliable the nighttime validated frequencies.

Despite the data set contains some dubious frequencies, that is, especially during nighttime hours, we kept all the frequencies for the training and used sunrise and sunset information to interpret the results according to the reliability of validated frequencies; the main reason is that we want to maximize the number of samples for the training phase.

Excluding times without selected frequencies, we eventually obtain 3514 samples for SUW-BEL, 4721 for TAR-BRZ, 3270 for OUI-HAN and 3112 for MUO-PEL.

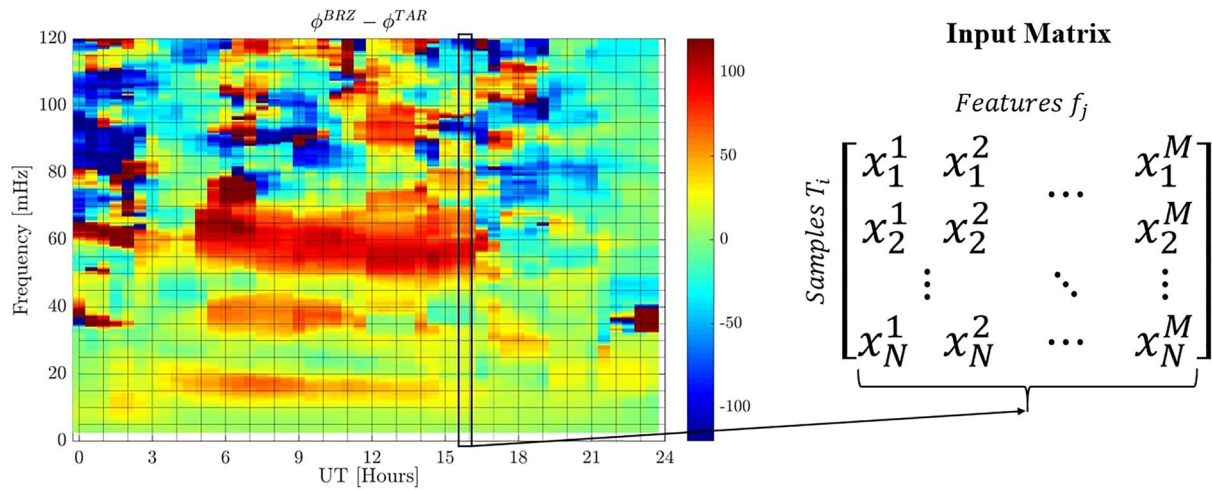


Figure 3. Graphical representation of the input matrix preparation. Every entry x_i^j represents the cross-phase value at a specific time and frequency value.

3.1. Input and Output Data

Every ML algorithm adopted in this analysis requires a so-called features matrix as input data, where rows represent different samples and columns are the input features. As previously stated, following the well-established cross-phase technique, our input for the ML algorithm is the cross-spectrum, as shown in Figure 3. The output will be the estimated value of the resonance frequency f_i for that sample.

Since we would like to test our ML algorithms on different pairs of stations, the input matrix is created separately for each couple. Typically, the resonance frequency decreases with increasing latitude except across the PBL where the behavior is the opposite. Thus, to obtain reasonable average spectral resolutions ($\Delta f / \bar{f} \sim 3\%$) at every latitude along the EMMA network, different pairs of stations have a different length of cross-phase spectra, resulting in features matrix with different dimensions: TAR-BRZ and SUW-BEL spectra are composed by 212 frequency bins (between 0 and 120 mHz), while OUJ-HAN and MUO-PEL have both 285 input features, but for OUJ-HAN they vary between 0 and 80 mHz while for MUO-PEL between 0 and 40 mHz.

Input data are transformed using two subsequent operations to improve the performance of ML algorithms: standardization, $(x - \mu)/\sigma$ and normalization, $(x - x_{\min})/(x_{\max} - x_{\min})$. The same procedure is applied to the output frequencies.

As can be observed in Figure 4, the first transformation gives a more Gaussian-like shape to the distribution of frequency, while the second one forces values between 0 and 1. Both operations are commonly adopted in various ML applications resulting in higher accuracy with every algorithm. After these transformation data are split in training and test set with roughly a 7:1 ratio (Figure 4); the former will be used to create the ML model, while the latter will evaluate the model performance.

4. Methods

In Figure 5 is depicted the entire ML pipeline built for FLR frequencies identification. After the stage of data pre-processing, described in the previous section, our ML procedure involves the application of typical ML techniques for creating the optimal model and subsequently improving its performance: feature selection and model selection.

Every ML algorithm is then trained via the cross-validation technique to further avoid over-fitting, producing a final model which is used to estimate frequencies from the test set.

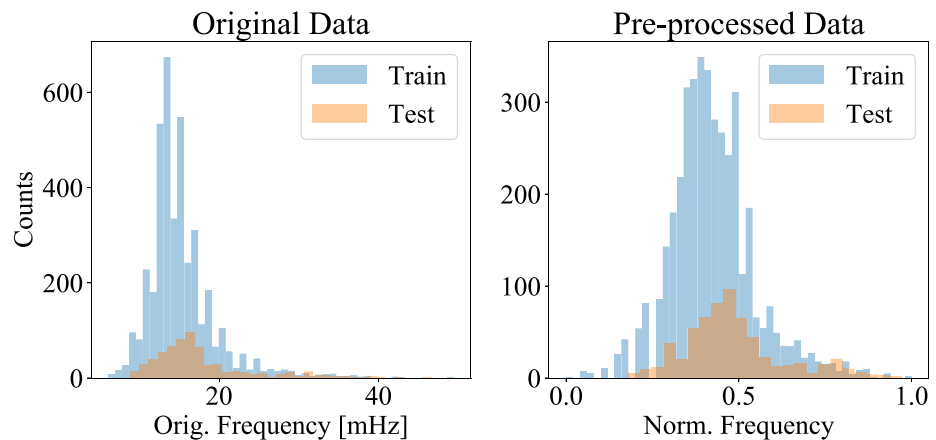


Figure 4. The left panel shows the original data set, while the right panel represents the same data after the pre-processing stage for Tartu-Birzai. In both panels distributions are divided in training (blue) and test (orange) set. This division is performed before the pre-processing procedure.

4.1. Feature Selection

Feature selection techniques are one of the central concept in ML methods, since the data set dimensionality, as well as the feature choice, severely affects the algorithm performance. Various works showed that numerous and redundant features may result in decreasing the algorithm accuracy and robustness (Guyon & Elisseeff, 2003; Hua et al., 2004; Kapetanos, 2007; Peng et al., 2005). The reasons for implementing and including feature selection routines in a ML pipeline can be summarized in four main points: lower dimensional data sets are simpler to interpret and visualize for end-users (the curse of dimensionality), increasing the model generalization (avoid over-fitting), improving the algorithm accuracy and shortening data processing times. For all these reasons we include two different feature selection techniques for reducing the data set dimensionality as well as evaluating additional features.

Feature selection methods can be categorized in embedded, filters, and wrapper methods (Guyon & Elisseeff, 2003). Embedded methods are ML models which implement feature selection using different regularization techniques, that is, l_1/l_2 -norm for linear models and a feature sub-sampling for DTs (Lal et al., 2006). Filters are model-independent methods which simply employ some correlation measure to find the optimal subset of features. Commonly known correlation measures are mutual information, point-wise mutual information and the Pearson's R correlation coefficient. Filters are computationally cheaper than other methods. Eventually, wrapper methods use model performance (accuracy and robustness) as a metric for evaluating feature subsets.

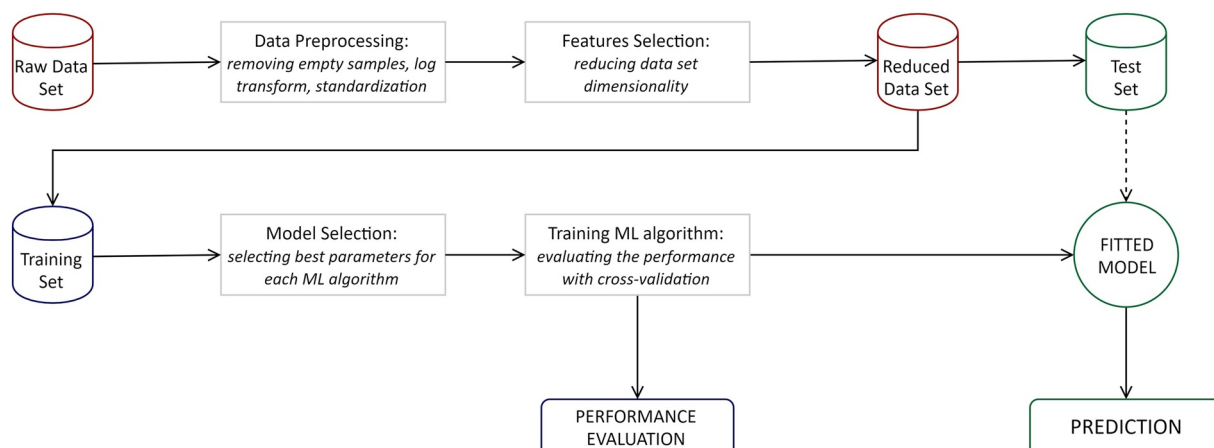


Figure 5. Graphical representation of the entire machine learning pipeline we built.

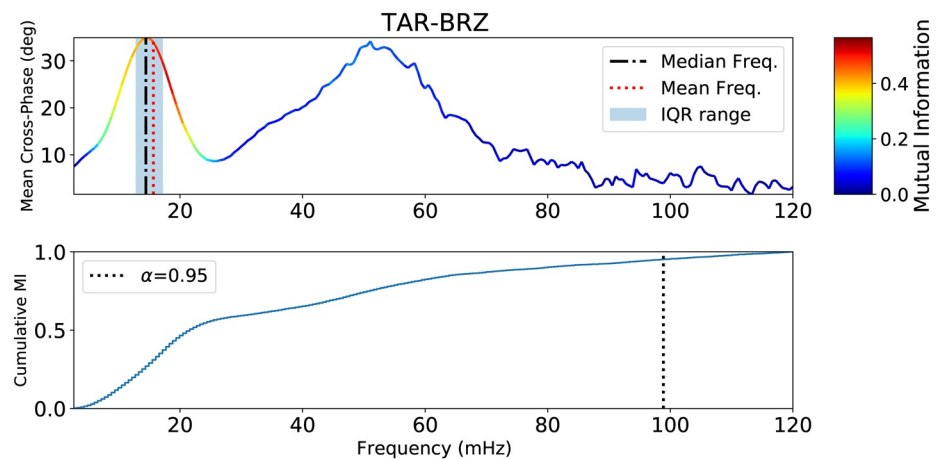


Figure 6. Results from the first feature selection technique. The upper panel represents the average cross-phase spectrum over the entire data set of Tartu-Birzai; black dash-dotted and red dotted vertical lines are the median and the average value of the validated frequency, respectively, while the blue interval is built with the first and third quartile of the distribution of frequencies. The average spectrum shows a peak close to the median value of the frequency, with a highly right-skewed distribution (high frequencies). The bottom panel is the cumulative mutual information (MI) indicating the 95% (vertical dotted line) of information subsisting between inputs and output. This confidence level is reached around 100 mHz.

These methods are computationally heavy, especially with high-dimensional data set, and often they allow only a random (or sparse) search between all the possible subsets.

For reducing the initial matrix dimensionality, we first compute the mutual information (MI) at every frequency bin and for the whole data set. Mutual information can be interpreted as a correlation measure between inputs and outputs, and it represents a powerful tool to identify even non-linear and non-monotonic correlations. Then we apply a time average over the entire period and compare mutual information values, represented in Figure 6 as a color scale, with the average cross-phase spectrum (Figure 6, top panel). By imposing a confidence level over the cumulative function of the mutual information (bottom panel of Figure 6) at $\alpha = 0.95$ (vertical dotted line), we can keep only frequencies below this threshold value. In this way, we ensure that, despite the lower number of features, the data set contains at least 95% of the information to estimate the output frequencies.

Figure 6 summarizes the results of this procedure for TAR-BRZ. As expected, highest values of mutual information correspond to the first peak of the average cross-phase spectrum, meaning that, on average, this is the most informative part of the spectrum for predicting the frequencies. The first peak occurs at $f_1 \approx 14$ mHz and is related to the first harmonic since it perfectly overlaps with the blue interval (top panel in Figure 6) which represents the 0.25 and 0.75 percentile of the validated FLR frequency distribution (i.e., interquartile range, IQR). The second peak occurs at $f_2 \sim 50$ mHz; the frequency ratio between the first and second peak is $f_2/f_1 \sim 3.6$ which is compatible with the expected ratio between the third and fundamental harmonics (Schulz, 1996).

Applying this procedure to the four pairs, we found that, independently from the couple of stations, the 95% threshold is reached using only the 80% of the initial number of features, resulting in 170 for TAR-BRZ and SUW-BEL, and 230 for OIJ-HAN and MUO-PEL.

The second part of the feature selection stage is the evaluation of additional features using wrapper methods. These novel features are spatio-temporal information, solar activity proxies, and geomagnetic activity indicators derived from geomagnetic indices. All the additional features analyzed in this section are summarized in Table 2. Temporal features are represented by the day of year (DoY) to highlight the seasonal variation of the resonance frequency (Vellante et al., 2007) and the decimal hour, because it is commonly known that resonance frequencies show a diurnal modulation (e.g., Chi et al., 2013; Del Corpo et al., 2019).

To consider the effects of solar activity, mostly for its impact on the production of ion-electron pairs at ionospheric height, we take into account the daily radio solar flux at 10.7 cm (F107, see Table 2) as provided by OMNIWeb. From the same database we also consider the y-component of the solar wind electric field (EField in Table 2) to investigate any impact of the convection electric field.

Table 2
Summary of all the Other Features Evaluated During the Features Assessment Procedure

Col. Name	Description
DoY	Day of Year
Hour	Decimal hour with a resolution of 0.5 h
Kp	Instantaneous <i>Kp</i> value
Kp_w_XX	<i>Kp</i> index weighted at different intervals 12, 24, ..., 72 h
Kp_m_XX	Maximum value of <i>Kp</i> index during last 12, 24, ..., 72 h
DST	Instantaneous <i>Dst</i> value (nT)
DST_m_XX	Minimum value of <i>Dst</i> index during last 12, 24, ..., 72 h (nT)
AE Index	Auroral Electrojet index
F107	Radio flux emitted from the entire solar disc at 10.7 cm (sfu or $W m^{-2} Hz^{-1}$)
EField	Convection electric field ($mV m^{-1}$)
r_{eq}	Field line equatorial crossing point (in Earth radii) with the Tsyganenko and Sitnov (2005, TS05) model

Notes. The geomagnetic indexes, the solar radio flux, and the convection electric field are derived from the OMNIWeb database.

Since the frequency also depends on the geomagnetic field line length we consider the equatorial crossing (r_{eq}) using the TS05 models to map the field line from the station pair position.

Geomagnetic activity indicators are evaluated to consider the FLR frequency variations due to the geomagnetic activity. The use of such indicators is widespread in the literature (e.g., Carpenter & Anderson, 1992; Del Corpo et al., 2019, 2020; Gallagher et al., 1988; Moldwin et al., 2002; Sheeley et al., 2001; Takahashi et al., 2006). The simplest indicators are the geomagnetic indices themselves. For this analysis, we consider the *Kp*, the *Dst* and the *AE* indices provided by the OMNIWeb database (<https://omniweb.gsfc.nasa.gov/>). More sophisticated indicators that take into account some effects of the magnetospheric dynamics can be derived from the *Kp* and *Dst* indices. In this work we consider two main different approaches: in the first, following the work by Moldwin et al. (2002), the maximum/minimum *Kp/Dst* are evaluated over a period τ preceding the time analyzed t_0 ; the second is made by following the work by Gallagher et al. (1988) and by evaluating the average *Kp* over a period τ preceding t_0 , weighted with an exponential function of the form $w(t) = \exp[-(t - t_0)/\tau]$. Both the procedures are applied to values of τ in the range 12–72 h with a step of 12 h, producing three groups of indices. The first is related to the *Kp* maximum in the preceding period and is reported in Table 2 and Figure 7 as *Kp_m_XX*, where XX are the τ values expressed in hours. The second group is related to the minimum *Dst* in the preceding period and it is indicated as *DST_m_XX*. The third one consists of *Kp* weighted averages and are represented as *Kp_w_XX*.

Figure 7 summarizes the evaluation of the additional features for all the six ML methods applied to the TAR-BRZ data set. Method performance, obtained using only the cross-phase as input, is reported as the initial value of the coefficient of determination, namely R_0^2 (reported in the lower right side of each panel of Figure 7). The additional features are evaluated by adding recursively the quantity which produces the best variation of the R^2 -score with respect to the previous value. This technique, commonly known as Sequential Forward Selection (SFS, Ferri et al., 1994), represents the most powerful tool to investigate medium and low-dimensional data sets. At each step, the selected feature is evaluated multiple times using the cross-validation technique; in this way we can produce a confidence interval defined by the standard deviation of the cross-validation score (orange area in Figure 7).

From the six panels in Figure 7 we can observe that the additional features have lower influence as the method complexity (and model performance) increases. Except for KRR and SVR, which result in poor performance, the other ML methods considered for this analysis (DTR, RF, LGBM, and XGB) basically express the same behavior. Algorithms that reach higher performance tend to return similar results also adding novel features, meaning that for these algorithms information from the cross-phase is sufficient to achieve the best accuracy. Another aspect is that it is impossible to select optimal quantities resulting in best performance for all the considered methods. Similar results are found for the other pairs of stations.

Hence, since no additional feature results in a significant improvement of the algorithms performance, we consider as input the cross-phase spectra only.

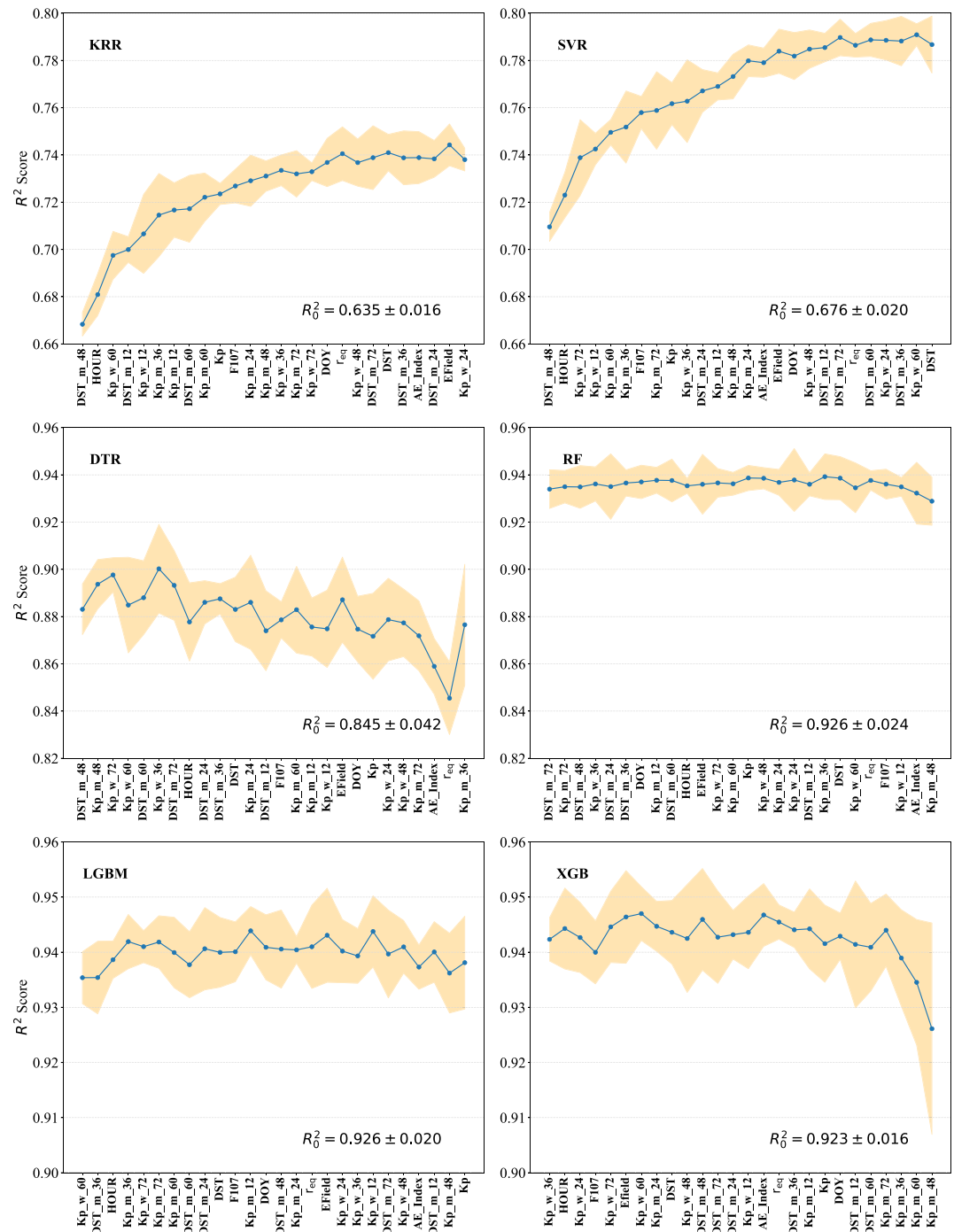


Figure 7. Results of the performance evaluation of all the additional features for Tartu-Birzai. The behavior for the other station pairs is the same. R_0^2 is the score obtained using only the cross-phase as input features. Methods with lower performance (Kernel Ridge Regression and Support Vector Regression) have an improvement with additional features while, as the model accuracy increases, they become less relevant. A comparison of these panels makes difficult to conclude which features could be more relevant for improving predictions.

4.2. Model Selection

The choice of the optimal ML method for a specific task and the procedure for tuning the model hyperparameters are both objectives of model selection. In Section 5.1 we will address the choice of the most suitable ML method for predicting resonance frequencies, by observing the model complexity and the accuracy for all the four

Table 3
Table of the Hyperparameters Optimized During Model Selection for Every Pair of Stations

Method	Hyperparameters	SUW-BEL	TAR-BRZ	OUI-HAN	MUO-PEL
KRR (w. RBF)	α	$7 \cdot 10^{-3}$	0.06	0.23	0.06
	γ	0.01	0.1	0.1	0.1
SVR (w. RBF)	C	1.5	1.5	1.5	1.5
	γ	0.1	0.1	0.5	0.5
DTR	Max. depth	19	17	17	5
	Min. samp. leaf	19	4	12	6
	Min. samp. split	14	8	19	8
RF	Max. depth	15	19	19	19
	# estimators	400	500	200	500
	Max. features	0.2	0.2	0.2	0.2
LGBM	# estimators	1,500	1,500	1,000	1,000
	# leaves	19	18	19	19
XGB	# estimators	1,000	200	300	200
	Max. depth	5	6	5	6
	Learn. rate	0.1	0.1	0.1	0.1

Abbreviations: KRR, Kernel Ridge Regression; LGBM, Light Gradient Boost Model; SVR, Support Vector Regression; XGB, eXtreme Gradient Boost.

station pairs. In this section we focus the attention on the selection of the best hyperparameters which control the model complexity of the six ML methods described in Section 2.

The number, typology and value of these parameters determine the model complexity and its performance. It is worth noting that hyperparameters, optimized in this step, are significantly different for the various methods; their optimal values are summarized in Table 3.

The first two rows in Table 3 are kernel methods which control the model complexity via the regularization terms, α for KRR and C for SVM. Higher values of α reduce collinearity between coefficients leading in higher generalization, while as α approaches zero KRR reverts to a simple OLS. In SVMs applied to regression problems (Support Vector Regression, SVR), the regularization parameter, C , controls the tolerance region width around the division boundary; increasing C results in more points close to the boundaries and hence more support vectors are required. Conversely, higher values of C mean a narrower tolerance and thus fewer points near boundaries.

In both methods input data are transformed using an RBF kernel for creating non-linear regression models (see Section 2.1). The hyperparameter γ for kernel methods refers to the standard deviation adopted for the Gaussian kernel (RBF kernel).

For decision trees and ensemble methods the main hyperparameters controlling the tree structure and the ensemble population are the maximum depth (i.e., Max. depth) and the number of estimators (i.e., # estimators), as briefly explained in Section 2.3.

Since the hyperparameter space is not too wide, for selecting the optimal values we perform a random search ($N = 3,000$ points) over the hyperparameter grid of each ML method. Real-valued hyperparameters are sampled using a logarithmic scale with $M = 1,000$ points, except for the maximum number of features which is picked from the values [0.0, 0.2, 0.5, 0.7, 1.0]. Kernel methods hyperparameters, α , C , and γ , vary in the range $[10^{-5}, 10]$, and the learning rate in the $[10^{-5}, 0.5]$ interval. Discrete-valued hyperparameters are all selected from specific lists of values: the maximum depth, the number of leaves and minimum samples in leaf/split are selected in the interval $[0, 1, \dots, 20]$, and the number of estimators is picked between [100, 200, 500, 700, 1000, 1500, 2000].

The selection of optimal model hyperparameters is essential into the ML pipeline for avoiding model over-fitting, reducing the variance error and increasing model accuracy. We select the best method and model for identifying resonance frequencies by considering two main aspects: the accuracy of the method and the ability to create a common model for all the four station pairs, reflected by more stable hyperparameters across the pairs (see Table 3).

After this stage, the next step is training/validation, during which all the ML models are trained using the cross-validation (CV) technique, with a number of CV folds, $k_{CV} = 5$. CV divides the training set into k_{CV} partitions; at each step one of the partition is excluded from the training phase, and it will be only used for validation. By optimizing results obtained with the validation set model, over-fitting is avoided and the model generalization increases.

The training set for each pair of stations is selected using two main criteria: all the station pairs must have the same proportion between training and test set, and the four pairs must share periods in the test set for a comparison. In particular, we selected DoY 148–162 (2013) as common test for the extremely varied geomagnetic conditions, consisting of a geomagnetic storm (June 2013, Stepanova et al., 2019; Thaller et al., 2015; Zou et al., 2017) preceded by prolonged quiescent days. It follows that the test set (~13% of the whole data set) consists of DoY 78–82 and 145–162 (2013) for TAR-BRZ, DoY 148–162 (2013) for OUI-HAN, and DoY 145–162 (2013) for SUW-BEL and MUO-PEL. The remaining portion, 87%, composes the training set.

5. Results and Discussions

5.1. Models Evaluation

The first part of the results focuses on the performance comparison of the various ML methods. The accuracy of regression problems can be evaluated by using various estimators, the coefficient of determination (R^2), the mean absolute error (MAE), the mean absolute percentage error (MAPE) and the root mean squared error (RMSE):

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|,$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2},$$

where y_i are the actual frequencies, \hat{y}_i the predicted frequencies and \bar{y} the average value of y_i .

For the CV procedure we use R^2 , since it weakly depends on the distribution function of data and therefore it is more reliable for regression problems. The results from the CV procedure are shown in the boxplot of Figure 8 for TAR-BRZ, the other station pairs show a similar behavior; hence we show only one pair for the sake of simplicity. The first aspect of this graph is a significant difference in accuracy between kernel-based (KRR and SVR) and tree-based models. The main reason for such a difference could rely on data: by using discrete cross-phase spectra the validated frequencies can assume only a discrete set of values and, for their implementation, tree-based methods are more suitable, and therefore accurate, to estimate discrete-like data.

From Figure 8 we can also argue that DTR has a high variance error, meaning that the model is too simple for the specific task. This statement is also supported by the high variability of hyperparameters across the station pairs (see Table 3). On the other hand, RF results in high accuracy, but it does not appear sufficiently robust with high frequency outliers, as highlighted by a large skewness and the presence of a flier point out of the ends of the whiskers. LGBM and XGB show a similar behavior even in their performance, resulting in robust algorithms which avoid over-fitting and reduce the variance error. These two methods also result in similar hyperparameter variability between the four station pairs (see Table 3), the main difference relies on the execution time (see Table 4); XGB processes the same number of sample in half the time with respect of LGBM.

After the CV stage each ML method is applied to the prediction of resonance frequencies of the test set for each station pair. The frequencies obtained during both the training and the test phase are available from <https://doi.org/10.5281/zenodo.4304911> (Foldes et al., 2020). The results for TAR-BRZ (DoY 78–82 and 145–162, 2013,

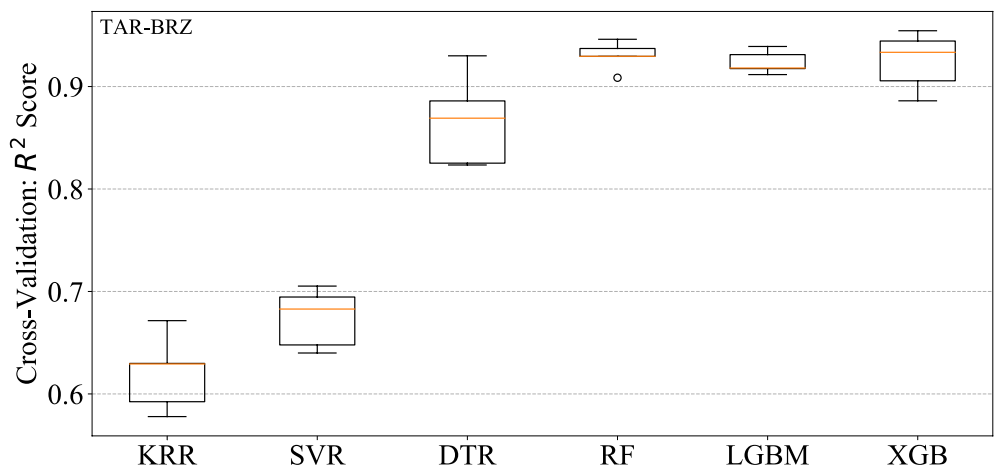


Figure 8. Comparison of the 6 ML methods trained in this analysis for Tartu-Birzai. Results for other station pairs have the same tendency. At first glance, we observe a substantial difference between kernel-based (Kernel Ridge Regression and Support Vector Regression) and tree-based models. Light Gradient Boost Model (LGBM) and eXtreme Gradient Boost (XGB) are quite similar methods as we see even from results, but XGB tends to be more robust for all the four pairs of stations since it reaches performances comparable with LGBM and RF using a lower number of estimators (i.e., simpler model, see Table 3); hence we consider XGB as the best-performed algorithm for this task.

Table 4
The General Performance of the Six ML Algorithms for TAR-BRZ

	R^2	MAE (mHz)	MAPE	RMSE (mHz)	CV time (s)
KRR	0.613	2.4	0.128	3.9	2.83
SVR	0.688	2.0	0.106	3.5	17.8
DTR	0.828	1.0	0.057	2.6	1.82
RF	0.840	0.93	0.042	2.5	32.7
LGBM	0.878	0.98	0.052	2.3	46.6
XGB	0.875	0.95	0.046	2.2	24.3

Abbreviations: KRR, Kernel Ridge Regression; LGBM, Light Gradient Boost Model; ML, machine learning; SVR, Support Vector Regression; TAR-BRZ, Tartu-Birzai; XGB, eXtreme Gradient Boost.

Notes. These results are obtained on the test set, $N_{\text{test}} = 619$. The best results for each metric are in bold. XGB does not show the best performances even though they are always close to the best results.

$N_{\text{test}} = 619$) are summarized in Table 4. By observing parameter stability in Table 3 and the algorithm performance in Table 4, we can observe that XGB and LGBM show almost the same robustness and accuracy. For these reasons, we choose XGB as most performing algorithm between them by considering two main aspects: it always employs fewer estimators than LGBM to reach the same performance, resulting as a simpler model, and because of its faster execution time it is more suitable for real-time applications.

To further explore the performance of the six ML methods we show, as an example, the relative error as a function of time (Figure 9) for a particular case study: 2013, DoY 151 (31 May). The relative error is the difference between the estimated frequency (f_{meth}) and the observed one (f_{obs}), normalized to f_{obs} . The intervals in which one or both the field line footprints are in darkness are represented as light or dark shadowed areas, respectively. The sunrise and sunset times (edges of the shadowed area) are evaluated at the height of 120 km as suggested by Del Corpo et al. (2020). The error increases during nighttime for all the methods, probably because, as stated above, the night frequencies are harder to detect and the relative cross-phase spectra can result poorly informative. During daytime the prediction error is remarkably

low for ensemble methods (especially LGBM and XGB), as previously pointed out in Table 4, and their fluctuations show a Gaussian-like behavior.

As stated above, XGB is the most suitable method to estimate FLR frequencies, hence, starting from this point, we perform further analysis using only the XGB method.

In Figure 10, we show an overview of the results obtained on the test set of TAR-BRZ. The first scatter plot (left panel) contains the median value of the predicted frequencies with the interquartile range (IQR) represented with bars. To compute the uncertainty of the model estimation we use two different approaches; for single estimator algorithms (i.e., KRR, SVR, and DTR) the uncertainty is computed using $N = 1,000$ different runs, while for multi-estimator algorithms (i.e., RF, LGBM, and XGB) the IQR is derived from the distribution of the estimators. The error increases when we observe a broader spread of the estimators; the uncertainty seems to increase slightly with frequency, but surely the error is greater when points are farther from the actual frequency value.

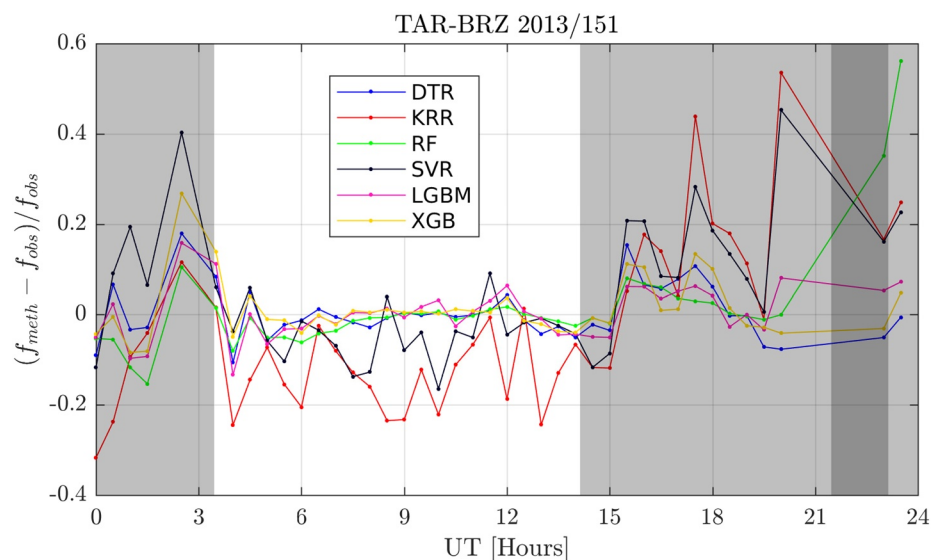


Figure 9. Comparison of the six machine learning methods for a quiescent day, 31 May. The light-gray area indicates when only one of the field line footprints is in the night-side. The dark-gray area occurs when both footprints are in darkness. In the daily part of the graph the relative error is significantly low ($\leq \pm 10\%$) for ensemble methods.

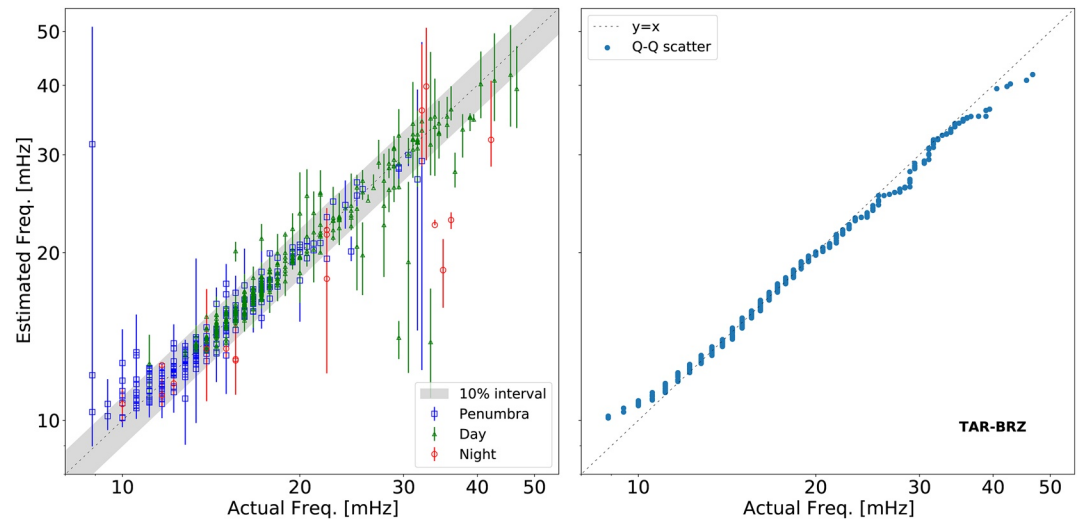


Figure 10. Global results from the test set of Tartu-Birzai. The left panel shows the median value of all the 619 test frequencies with their interquartile range as error bars; colors distinguish between diurnal (green), nocturnal (red) and penumbral (blue) frequencies. Right panel is the quantile-quantile (Q-Q) plot; it summarizes simultaneously several aspects of the target (actual frequencies) and obtained (estimated frequencies) distribution. The dashed line is the bisector and it indicates that the two sets come from a population with the same distribution.

Colors of the scatter plot represent the day, night and penumbral frequencies; the latter are those selected when only one footprint of the field line is sunlit. The division is made using the same criteria of Figure 9. Night and penumbral frequencies are harder to be observed (Balasis et al., 2019; Del Corpo et al., 2019) and indeed they are less numerous. The main reason is that cross-phase spectra are barely informative and hence FLR identification results more difficult. During penumbra, frequencies could correspond to quarter-wave mode events (Del Corpo et al., 2019; Obana et al., 2015). Right panel in Figure 10 represents a quantile-quantile plot (Q-Q plot). The Q-Q plot describes various distributional aspects of the two set of frequencies, actual and estimated; since points in Figure 10 (right panel) accurately follow the identity line (dashed line), the mean value of the two distributions does not show a significant shift. At frequency values higher than 38 mHz, and between 25 and 32 mHz, points move below the bisector, which means that estimated frequencies are slightly lower than actual frequencies in these ranges. As can be noticed from the left panel in Figure 10, the lack of points at higher frequencies (≥ 40 mHz) produces higher errors which nevertheless can be considered as statistical fluctuations.

5.2. The June 2013 Geomagnetic Storm: A Case Study

As already stated in Section 4.2, the geomagnetic storm occurred in June 2013 offers an excellent opportunity to test the algorithm performance since it comprises highly variable geomagnetic conditions with comprehensive data coverage for all the considered station pairs. The two top panels of Figure 11 show the time variations of the Dst and Kp indices, respectively. The main phase starts at the beginning of DoY 152 and proceeds until noon when Dst reaches a minimum value of about -120 nT and Kp rises up to 7. The recovery phase lasts for a few days during which both geomagnetic indices slowly recover to quiet time values. A peculiarity of this event is the prolonged very low geomagnetic activity conditions that characterize the days preceding the storm, allowing us to test the efficiency of the algorithm also in such an unusual situation.

The last four panels in Figure 11 show, from top to bottom, the relative error of the estimated frequency for the four station pairs arranged in decreasing latitude order. As for Figure 9, light and dark shadowed areas correspond to periods in which the field lines have one or both footprint in dark, respectively. The average relative error at $L = 2.4, 2.9, 4.1$ is about $+1-2\%$ suggesting that the distribution of the deviations has a Gaussian-like shape. At $L = 5.5$ the average relative error is about $+4.5\%$ suggesting a systematic overestimation of the algorithm, even if this larger value might be due to a preponderance of positive deviations on DoY 152 during which the main phase of the storm took place. To further evaluate the performance of the algorithm we can look at the average deviation

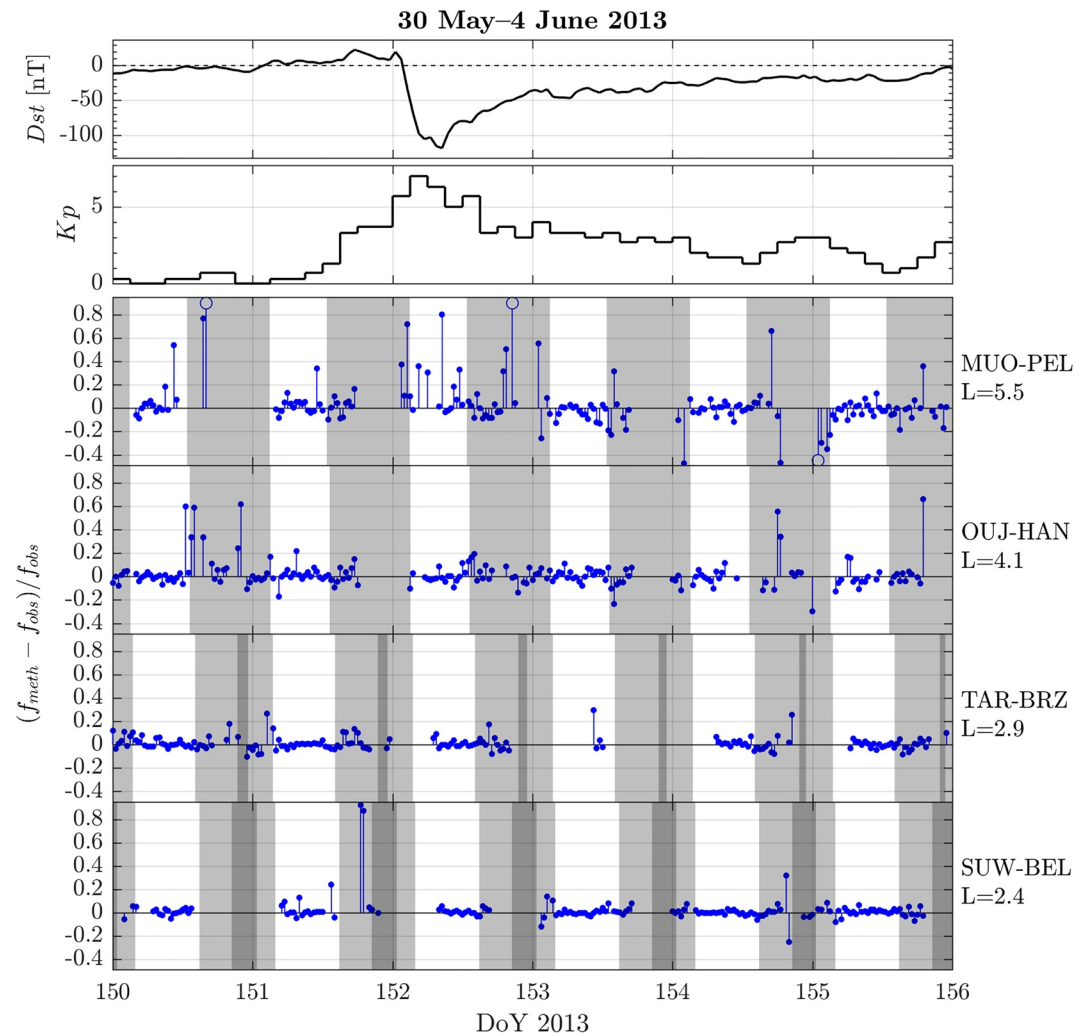


Figure 11. From top to bottom. First two panels represent the Kp and Dst indices respectively during DoY 150–155 of 2013 (30 May to 4 June). It is possible to clearly distinguish between the pre-storm phase (DoY 150 and 151), the main phase (DoY 152) and the recovery phase (DoY 153 and subsequent). The other four panels represent the evolution of the relative error during these six days with the station latitude in descending order. In this representation the difference between stations, the behavior with variable geomagnetic conditions and various data gaps are well evidenced. Points out of graph limits are represented as open circles.

from zero (the MAPE) that does not exceed 13%, although it increases from 4% to 13% with increasing latitude. As already pointed out, this could be due to the signal-to-noise ratio in the cross-phase spectra that becomes lower as the latitude increases. During daytime hours the MAPE assumes the values 2.4%, 2.7%, 4.7%, and 6.9% at $L = 2.4, 2.9, 4.1,$ and $5.5,$ respectively, suggesting that the method performs significantly better in these conditions. Again, there is the exception on DoY 152 at $L = 5.5$ for which the daytime MAPE is 17.5%, indicating that, at least for high latitude pairs, the performance of the method deteriorates during highly dynamical geomagnetic conditions. Also during the extreme quiet conditions on DoY 150 there are some important deviations at $L = 4.1$ and $5.5,$ mainly during nighttime. Other than that, there is no evidence of any dependence of the relative error from the geomagnetic activity. It is important to note that almost all the situations in which the algorithm does not perform well are associated to fuzzy or low signal-to-noise ratio cross-phase spectra for which also the supervised selection can be sometimes questionable. These issues can be better understood looking at Figure 12 which shows the dynamic cross-phase spectra of the four pairs on DoY 151 (panels a–d) and 152 (panels e–h). Black open squares are the frequencies estimated by the algorithm while magenta open circles are the frequencies validated by Del Corpo et al. (2019). The shadowed regions represent dark and light hours as in Figures 9 and 11. The main phase of the storm starts at the beginning of DoY 152, however the magnetosphere is already perturbed in the

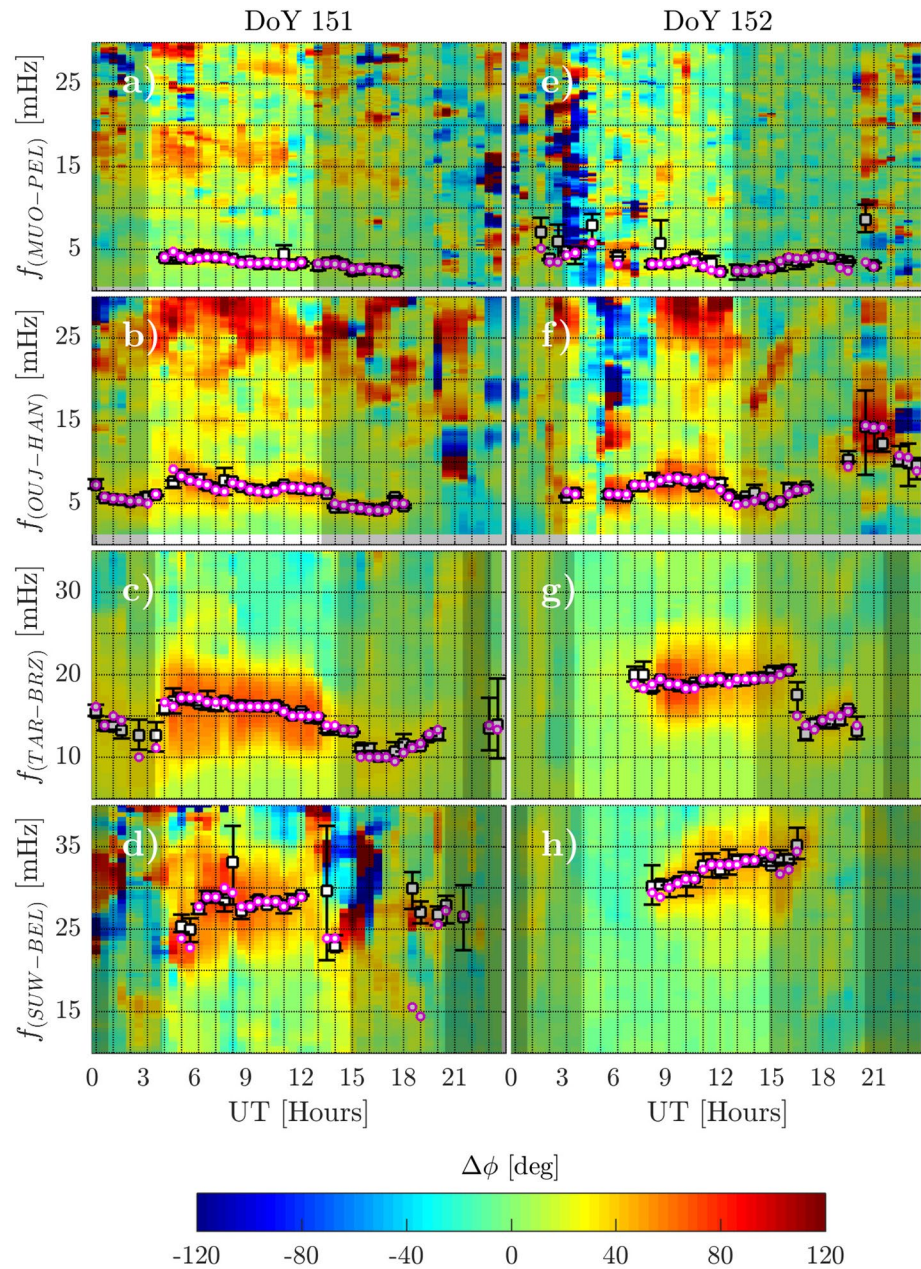


Figure 12. Panel of results obtained with eXtreme Gradient Boost for DoY 151 and 152 (2013). Black points with error bars are the predicted frequencies while the magenta ones are the observed frequencies. DoY 151 is a quiescent day while during DoY 152 (1 June) the main phase of a geomagnetic storm occurs. Such a difference is highlighted by slightly different resonance frequencies between the two days. Predictions show high accordance with observed frequencies in most cases, especially during daytime.

afternoon of DoY 151 as clearly visible from the growing of the K_p index (see Figure 11). During this period the spectra are fuzzy for all the pairs but TAR-BRZ, and important mismatches between the estimated and validated frequencies occur at SUW-BEL on DoY 151 and at MUO-PEL on DoY 152. Such mismatches could not depend exclusively from the algorithm efficacy, but also from the presence of questionable points in the data set by Del Corpo et al. (2019), as is the case for the above mentioned points at SUW-BEL. They created the data set taking particular care to daytime hours when the physical assumptions made in the entire procedure are expected to be applicable, but keeping also nighttime frequencies to study the efficiency of the procedure during these conditions; occasionally, they kept also some unclear points as the ones just described. In this preliminary work we

used the entire data set as it was, analyzing also nocturnal frequencies, but to obtain an operative space weather tool the reliability of the selected night frequencies should be investigated and all questionable data points should be excluded. It is worth noting that, also during nighttime or disturbed geomagnetic conditions, the algorithm performs well when the spectra have smooth and clear traces of the resonance peak. A good example are the spectra of TAR-BRZ (panels c and g) in which the only notable deviations from the validated frequencies arise during discontinuities as in the interval 02:30–03:30 of DoY 151 and at 16:30 on DoY 152, when a quarter wave event and a PBL crossing probably happened. In general, the good performances of the algorithm, mainly during daytime hours, are confirmed.

6. Conclusions

Our work represents a preliminary result in the direction of building a tool for monitoring the plasmasphere dynamics using the whole EMMA network.

We have evaluated several ML algorithms, commonly adopted for regression problems, to estimate FLR frequencies in cross-phase spectra among magnetic field measurements. This survey has proved that ML techniques can be successfully applied for addressing this challenging task; in our particular case tree-based methods have resulted the most reliable algorithms of those considered.

Results have shown a slight dependence of the estimation error with the station latitude, that is, MAPE increases from 4% to 13% passing from $L = 2.4$ to $L = 5.5$, so it remains acceptable even at high latitude. In fact, the percentage of points with an absolute relative error ≤ 0.1 is about 70% for MUO-PEL ($L = 5.5$), and it increases rapidly at OUJ-HAN ($L = 4.1$), 82%, reaching 90% for TAR-BRZ ($L = 2.9$) and SUW-BEL ($L = 2.4$). An interesting aspect evidenced in this analysis is that results generally do not depend on geomagnetic activity, except for high latitude pairs for which extreme geomagnetic conditions (i.e., storm main phase, prolonged quiet conditions) could produce a higher estimation error. We have considered the June 2013 geomagnetic storm as case study; except for $L = 5.5$, when the MAPE reaches values greater than 15% in some points, during this event the estimation error remains around its average value passing from quiescent days to the main phase of the geomagnetic storm, and eventually to the recovery phase. This feature makes our ML procedure particularly reliable for space weather purposes within the framework of the plasmaspheric density nowcast.

However, since we have trained our models only on times with observed FLR frequencies, our procedure is unable to distinguish whether a resonance is triggered. To create a complete automated ML procedure suitable for real-time monitoring, it may be required an additional ML step placed before the one we have already built. To classify times when the FLR frequency can be observed or not, we could follow an approach similar to Balasis et al. (2019), who adopted a fuzzy-logic neural network for classifying periods with ULF activity in geomagnetic field time series.

The present analysis could be mainly improved by extending our training set. First, a larger data set means a more balanced proportion between the number of samples, N , and the number of features, M , that in our case is represented by the number of frequencies in the cross-phase spectra. Then, by extending the data set, we could remove from the training data those FLR frequencies which present an unclear trace on the cross-phase spectrum and hence are considered less reliable (e.g., nighttime frequencies, quarter-waves). Lastly, a larger data set would allow to test our procedure on most varied geomagnetic conditions.

Our approach may be extended to all the EMMA station pairs, and in principle to any magnetometer network, like CARISMA (Mann et al., 2008) and McMAC (Chi et al., 2005). Testing this method on other pairs of stations, even longitudinally separated from EMMA, would allow to obtain more information about the accuracy and the robustness of our technique.

Data Availability Statement

Cross-phase spectra and FLR resonances used to train the ML methods are available from Zenodo (<https://doi.org/10.5281/zenodo.4304662>). The output of the various ML methods are available from Zenodo (<https://doi.org/10.5281/zenodo.4304911>).

Acknowledgments

This work is partially supported by Italian MIUR-PRIN grant 2017APKP7T_004 on *Circumterrestrial Environment: Impact of Sun-Earth Interaction*. We thank the Finnish Meteorological Institute (FMI), the University of Oulu (Finland), the Institute of Geophysics of the Polish Academy of Sciences (IGF-PAS), the Mining and Geological Survey of Hungary (MBFSZ) and the University of L'Aquila for contributing to EMMA. We also acknowledge use of NASA/GSFC's Space Physics Data Facility's OMNIWeb service, and OMNI data. Open Access Funding provided by Università degli Studi dell'Aquila within the CRUI-CARE Agreement.

References

Allan, W., & Knox, F. B. (1979). A dipole field model for axisymmetric Alfvén waves with finite ionosphere conductivities. *Planetary and Space Science*, 27(1), 79–85. [https://doi.org/10.1016/0032-0633\(79\)90149-1](https://doi.org/10.1016/0032-0633(79)90149-1)

Armstrong, J. A., & Fletcher, L. (2019). Fast solar image classification using deep learning and its importance for automation in solar physics. *Solar Physics*, 294(80). <https://doi.org/10.1007/s11207-019-1473-z>

Balasis, G., Ainalragia-Giamini, S., Papadimitriou, C., Daglis, I. A., Anastasiadis, A., & Haegmans, R. (2019). A machine learning approach for automated ulf wave recognition. *Journal of Space Weather and Space Climate*, 9, A13. <https://doi.org/10.1051/swsc/2019010>

Bao, X., Jiang, D., Yang, X., & Wang, H. (2021). An improved deep belief network for traffic prediction considering weather factors. *Alexandria Engineering Journal*, 60, 413–420. <https://doi.org/10.1016/j.aej.2020.09.003>

Baransky, L. N., Borovkov, J. E., Gokhberg, M. B., Krylov, S. M., & Troitskaya, V. A. (1985). High resolution method of direct measurement of the magnetic field lines' Eigen frequencies. *Planetary and Space Science*, 33(12), 1369–1374. [https://doi.org/10.1016/0032-0633\(85\)90112-6](https://doi.org/10.1016/0032-0633(85)90112-6)

Berube, D., Moldwin, M. B., & Weygand, J. M. (2003). An automated method for the detection of field line resonance frequencies using ground magnetometer techniques. *Journal of Geophysical Research*, 108(A9), 1348. <https://doi.org/10.1029/2002JA009737>

Boynton, R., Balikhin, M., Wei, H.-L., & Lang, Z.-Q. (2018). Applications of NARMAX in space weather. In E. Camporeale, J. Johnson, & S. Wing (Eds.), *Machine learning techniques for space weather* (p. 203 - 236). Elsevier. <https://doi.org/10.1016/B978-0-12-811788-0.00008-1>

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123–140. <https://doi.org/10.1023/A:1018054314350>

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth. <https://doi.org/10.1201/9781315139470>

Breuillard, H., Dupuis, R., Retino, A., Le Contel, O., Amaya, J., & Lapenta, G. (2020). Automatic classification of plasma regions in near-Earth space with supervised machine learning: Application to magnetospheric multi scale 2016–2019 observations. *Frontiers in Astronomy and Space Sciences*, 7, 55. <https://doi.org/10.3389/fspas.2020.00055>

Buhlmann, P. (2012). *Bagging, boosting and ensemble learning*. Springer-Verlag. https://doi.org/10.1007/978-3-642-21551-3_33

Buntine, W., & Niblett, T. (1992). A further comparison of splitting rules for decision-tree induction. *Machine Learning*, 8, 75–85. <https://doi.org/10.1007/BF00994006>

Camporeale, E., Wing, S., & Johnson, J. R. (2018). Introduction. In E. Camporeale, J. Johnson, & S. Wing (Eds.), *Machine learning techniques for space weather* (pp. XIII–XVIII). Elsevier. <https://doi.org/10.1016/b978-0-12-811788-0.09987-x>

Carpenter, D. L., & Anderson, R. R. (1992). An ISEE/whistler model of equatorial electron density in the magnetosphere. *Journal of Geophysical Research*, 97(A2), 1097–1108. <https://doi.org/10.1029/91JA01548>

Carpenter, D. L., & Lemaire, J. (2004). The plasmasphere boundary layer. *Annales Geophysicae*, 22(12), 4291–4298. <https://doi.org/10.5194/angeo-22-4291-2004>

Chandorkar, M., & Camporeale, E. (2018). Probabilistic forecasting of geomagnetic indices using Gaussian process models. In E. Camporeale, J. Johnson, & S. Wing (Eds.), *Machine learning techniques for space weather* (pp. 237–258). Elsevier. <https://doi.org/10.1016/B978-0-12-811788-0.00009-3>

Chen, T., & Guestrin, C. (2016). XGBoost. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. <https://doi.org/10.1145/2939672.2939785>

Chi, P. J., Engebretson, M. J., Moldwin, M. B., Russell, C. T., Mann, I. R., Hairston, M. R., et al. (2013). Sounding of the plasmasphere by Mid-continent MAGnetoseismic Chain (McMAC) magnetometers. *Journal of Geophysical Research: Space Physics*, 118(6), 3077–3086. <https://doi.org/10.1002/jgra.50274>

Chi, P. J., Engebretson, M. J., Moldwin, M. B., Russell, C. T., Mann, I. R., Samson, J. C., & Lee, D. H. (2005). Mid-continent magnetoseismic chain (McMAC): A meridional magnetometer chain for magnetospheric sounding. In *Proceedings of the environment modeling workshop* (pp. 17–22).

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>

Del Corpo, A., & Vellante, M. (2020). *Field line resonances observed by EMMA*. ver. 1.0.0 [Data set] Zenodo. <https://doi.org/10.5281/zenodo.4304662>

Del Corpo, A., Vellante, M., Heilig, B., Pietropaolo, E., Reda, J., & Lichtenberger, J. (2019). Observing the cold plasma in the Earth's magnetosphere with the EMMA network. *Annals of Geophysics*, 61(4). <https://doi.org/10.4401/ag-7751>

Del Corpo, A., Vellante, M., Heilig, B., Pietropaolo, E., Reda, J., & Lichtenberger, J. (2020). An empirical model for the dayside magnetospheric plasma mass density derived from EMMA magnetometer network observations. *Journal of Geophysical Research: Space Physics*, 125(2), e2019JA027381. <https://doi.org/10.1029/2019JA027381>

Esposito, F., Malerba, D., Semeraro, G., & Kay, J. (1997). A comparative analysis of methods for pruning decision trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5), 476–491. <https://doi.org/10.1109/34.589207>

Ferri, F., Pudil, P., Hatef, M., & Kittler, J. (1994). Comparative study of techniques for large-scale feature selection. In E. S. Gelsema & L. S. Kanal (Eds.), *Pattern recognition in practice iv* (Vol. 16, pp. 403–413). North-Holland. <https://doi.org/10.1016/B978-0-444-81892-8.50040-7>

Foldes, R., Del Corpo, A., Pietropaolo, E., & Vellante, M. (2020). *Field line resonances estimated using machine learning methods*. ver. 1.0.0 [Data set] Zenodo. <https://doi.org/10.5281/zenodo.4304911>

Fraser, B., Horwitz, J., Slavin, J., Dent, Z., & Mann, I. (2005). Heavy ion mass loading of the geomagnetic field near the plasmapause and ULF wave implications. *Geophysical Research Letters*, 32(4), L04102. <https://doi.org/10.1029/2004GL021315>

Fujimoto, A., Obana, Y., & Saita, S. (2019). *Seasonal variation in plasmaspheric mass density in the New Zealand meridian*. In (Vol. 2019, pp. SM51B–3186).

Gallagher, D., Craven, P., & Comfort, R. (1988). An empirical model of the Earth's plasmasphere. *Advances in Space Research*, 8(8), 15–24. [https://doi.org/10.1016/0273-1177\(88\)90258-X](https://doi.org/10.1016/0273-1177(88)90258-X)

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157–1182. <https://doi.org/10.1162/153244303322753616>

Heidrich-Meisner, V., & Wimmer-Schweingruber, R. F. (2018). Chapter 16—Solar wind classification via k-means clustering algorithm. In E. Camporeale, J. Johnson, & S. Wing (Eds.), *Machine learning techniques for space weather* (pp. 397–424). Elsevier. <https://doi.org/10.1016/B978-0-12-811788-0.00016-0>

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634>

Hua, J., Xiong, Z., Lowey, J., Suh, E., & Dougherty, E. R. (2004). Optimal number of features as a function of sample size for various classification rules. *Bioinformatics*, 21(8), 1509–1515. <https://doi.org/10.1093/bioinformatics/bti171>

- Ivanov, S., Tsizh, M., Ullmann, D., Panos, B., & Voloshynovskiy, S. (2020). Solar activity classification based on mg ii spectra: Toward classification on compressed data. *arXiv*. Retrieved From <https://arxiv.org/abs/2009.07156>
- Kapetanios, G. (2007). Variable selection in regression models using nonstandard optimisation of information criteria. *Computational Statistics & Data Analysis*, 52(1), 4–15. <https://doi.org/10.1016/j.csda.2007.04.006>
- Ke, G., Meng, Q., Finley, T., Wang, T., Wei, C., Ma, W., et al. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Nips'17: Proceedings of the 31st international conference on neural information processing systems* (pp. 3149–3157).
- Kumar, B., Vyas, O. P., & Vyas, R. (2019). A comprehensive review on the variants of support vector machines. *Modern Physics Letters B*, 33(25). <https://doi.org/10.1142/S0217984919503032>
- Kunte, A., & Panicker, S. (2020). Personality prediction of social network users using ensemble and xgboost. In H. Das, P. K. Pattnaik, S. S. Rautaray, & K.-C. Li (Eds.), *Progress in computing, analytics and networking* (pp. 133–140). Springer Singapore. https://doi.org/10.1007/978-981-15-2414-1_14
- Lal, T. N., Chapelle, O., Weston, J., & Elisseeff, A. (2006). Embedded methods. In I. Guyon, M. Nikravesh, S. Gunn, & L. A. Zadeh (Eds.), *Feature extraction: Foundations and applications* (pp. 137–165). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-35488-8_6
- Li, R.-H., & Belford, G. G. (2002). Instability of decision tree classification algorithms. In *Proceedings of ACM SIGKDD international conference on knowledge discovery and data mining*. <https://doi.org/10.1145/775047.775131>
- Lichtenberger, J., Clilverd, M. A., Heilig, B., Vellante, M., Manninen, J., Rodger, C. J., et al. (2013). The plasmasphere during a space weather event: First results from the PLASMON project. *Journal of Space Weather and Space Climate*, 3, A23. <https://doi.org/10.1051/swsc/2013045>
- Liemohn, M. W. (2006). Introduction to special section on “results of the national science foundation geospace environment modeling inner magnetosphere/storms assessment challenge”. *Journal of Geophysical Research*, 111, A11S01. <https://doi.org/10.1029/2006JA011970>
- Liu, J., Ye, Y., Shen, C., Wang, Y., & Erdélyi, R. (2018). A new tool for CME arrival time prediction using machine learning algorithms: CAT-PUMA. *The Astrophysical Journal*, 855(2), 109. <https://doi.org/10.3847/1538-4357/aaac69>
- Luckner, M., Topolski, B., & Mazurek, M. (2017). Application of xgboost algorithm in fingerprinting localisation task. https://doi.org/10.1007/978-3-319-59105-6_57
- Mann, I., Milling, D., Rae, I. J., Ozeke, L., Kale, A., Kale, Z., et al. (2008). The upgraded CARISMA magnetometer array in the THEMIS era. *Space Science Reviews*, 141, 413–451. <https://doi.org/10.1007/s11214-008-9457-6>
- Massone, A. M., Piana, M., & Consortium, F. (2018). Chapter 14—Machine learning for flare forecasting. In E. Camporeale, J. Johnson, & S. Wing (Eds.), *Machine learning techniques for space weather* (pp. 355–364). Elsevier. <https://doi.org/10.1016/B978-0-12-811788-0.00014-7>
- Menk, F. W., Mann, I. R., Smith, A. J., Waters, C. L., Clilverd, M. A., & Milling, D. K. (2004). Monitoring the plasmopause using geomagnetic field line resonances. *Journal of Geophysical Research*, 109, A04216. <https://doi.org/10.1029/2003JA010097>
- Milling, D. K., Mann, I. R., & Menk, F. W. (2001). Diagnosing the plasmopause with a network of closely spaced ground-based magnetometers. *Geophysical Research Letters*, 28(1), 115–118. <https://doi.org/10.1029/2000GL011935>
- Moldwin, M. B., Downward, L., Rassoul, H., Amin, R., & Anderson, R. (2002). A new model of the location of the plasmopause: CRRES results. *Journal of Geophysical Research*, 107(A11), 1339. <https://doi.org/10.1029/2001JA009211>
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. The MIT Press. <https://doi.org/10.1109/iroso.2012.6386118>
- Obana, Y., Menk, F. W., Sciffer, M. D., & Waters, C. L. (2008). Quarter-wave modes of standing Alfvén waves detected by cross-phase analysis. *Journal of Geophysical Research*, 113(A8). <https://doi.org/10.1029/2007JA012917>
- Obana, Y., Waters, C. L., Sciffer, M. D., Menk, F. W., Lysak, R. L., Shiokawa, K., et al. (2015). Resonance structure and mode transition of quarter-wave ULF pulsations around the dawn terminator. *Journal of Geophysical Research: Space Physics*, 120(6), 4194–4212. <https://doi.org/10.1002/2015JA021096>
- Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), 1226–1238. <https://doi.org/10.1109/TPAMI.2005.159>
- Schulz, M. (1996). Eigenfrequencies of geomagnetic field lines and implications for plasma-density modeling. *Journal of Geophysical Research*, 101(A8), 17385–17397. <https://doi.org/10.1029/95JA03727>
- Sheeley, B., Moldwin, M., Rassoul, H., & Anderson, R. (2001). An empirical plasmasphere and trough density model: CRRES observations. *Journal of Geophysical Research*, 106(A11), 25631–25641. <https://doi.org/10.1029/2000JA000286>
- Singer, H., Southwood, D., Walker, R., & Kivelson, M. (1981). Alfvén wave resonances in a realistic magnetospheric magnetic field geometry. *Journal of Geophysical Research*, 86(A6), 4589–4596. <https://doi.org/10.1029/JA086iA06p04589>
- Stepanova, M., Antonova, E. E., Moya, P. S., & Pinto, J. A. (2019). Multisatellite analysis of plasma pressure in the inner magnetosphere during the 1 June 2013 geomagnetic storm. *Journal of Geophysical Research: Space Physics*, 124. <https://doi.org/10.1029/2018JA025965>
- Takahashi, K., Denton, R. E., Anderson, R. R., & Hughes, W. J. (2006). Mass density inferred from toroidal wave frequencies and its comparison to electron density. *Journal of Geophysical Research*, 111, A01201. <https://doi.org/10.1029/2005JA012886>
- Takahashi, K., Vellante, M., Del Corpo, A., Claudepierre, S. G., Kletzing, C., Wygant, J., & Koga, K. (2020). Multiharmonic toroidal standing Alfvén waves in the midnight sector observed during a geomagnetically quiet period. *Journal of Geophysical Research: Space Physics*, 125(3), e2019JA027370. <https://doi.org/10.1029/2019JA027370>
- Thaller, S. A., Wygant, J. R., Dai, L., Breneman, A. W., Kersten, K., Cattell, C. A., et al. (2015). Van Allen probes investigation of the large-scale duskward electric field and its role in ring current formation and plasmasphere erosion in the 1 June 2013 storm. *Journal of Geophysical Research: Space Physics*, 120(6), 4531–4543. <https://doi.org/10.1002/2014JA020875>
- Theodoridis, S., & Koutroumbas, K. (2008). *Pattern recognition* (4th ed.). Academic Press, Inc. <https://doi.org/10.1016/B978-1-59749-272-0.X0001-2>
- Thorne, R. M. (2010). Radiation belt dynamics: The importance of wave-particle interactions. *Geophysical Research Letters*, 37, L22107. <https://doi.org/10.1029/2010GL044990>
- Tsyganenko, N. A., & Sitnov, M. I. (2005). Modeling the dynamics of the inner magnetosphere during strong geomagnetic storms. *Journal of Geophysical Research*, 110(A3). <https://doi.org/10.1029/2004JA010798>
- Vellante, M., Förster, M., Villante, U., Zhang, T. L., & Magnes, W. (2007). Solar activity dependence of geomagnetic field line resonance frequencies at low latitudes. *Journal of Geophysical Research: Space Physics*, 112, A02205. <https://doi.org/10.1029/2006JA011909>
- Waters, C. L., Menk, F. W., & Fraser, B. J. (1991). The resonance structure of low latitude Pc3 geomagnetic pulsations. *Geophysical Research Letters*, 18(12), 2293–2296. <https://doi.org/10.1029/91GL02550>
- Wharton, S. J., Wright, D. M., Yeoman, T. K., James, M. K., & Sandhu, J. K. (2018). Cross-phase determination of ultralow frequency wave harmonic frequencies and their associated plasma mass density distributions. *Journal of Geophysical Research: Space Physics*, 123(8), 6231–6250. <https://doi.org/10.1029/2018JA025487>

- Wing, S., Johnson, J. R., Camporeale, E., & Reeves, G. D. (2018). Chapter 6 - Untangling the solar wind drivers of the radiation belt: An information theoretical approach. In E. Camporeale, J. Johnson, & S. Wing (Eds.), *Machine learning techniques for space weather* (pp. 149–175). Elsevier. <https://doi.org/10.1016/B978-0-12-811788-0.00006-8>
- Yang, H., Huang, K., King, I., & Lyu, M. R. (2009). Localized support vector regression for time series prediction. *Neurocomputing*, 72(10), 2659–2669. (Lattice Computing and Natural Computing (JCIS 2007)/Neural Networks in Intelligent Systems Designn (ISDA 2007)) <https://doi.org/10.1016/j.neucom.2008.09.014>
- Zhelavskaya, I. S., Shprits, Y. Y., & Spasojević, M. (2017). Empirical modeling of the plasmasphere dynamics using neural networks. *Journal of Geophysical Research: Space Physics*, 122(11), 11227–11244. <https://doi.org/10.1002/2017JA024406>
- Zhelavskaya, I. S., Spasojevic, M., Shprits, Y. Y., & Kurth, W. S. (2016). Automated determination of electron density from electric field measurements on the van Allen probes spacecraft. *Journal of Geophysical Research: Space Physics*, 121(5), 4611–4625. <https://doi.org/10.1002/2015JA022132>
- Zou, S., Ridley, A., Jia, X., Boyd, E., Nicolls, M., Coster, A., et al. (2017). PFISR observation of intense ion upflow fluxes associated with an SED during the 1 June 2013 geomagnetic storm. *Journal of Geophysical Research: Space Physics*, 122(2), 2589–2604. <https://doi.org/10.1002/2016JA023697>