ASTROFLU Seminar
8 Dec 2021

A **Machine Learning** approach for Monitoring the **Plasmaspheric Mass Density** using Ground-Based Measurements

R. Foldes[1,2], A. Del Corpo[1], E. Pietropaolo[1], M. Vellante[1], R. Marino[2], E. Lévêque[2]

[1]*Universitá de L'Aquila, Dipartimento di Scienze Fisiche e Chimiche*
[2]*École Centrale de Lyon, Laboratoire de Méchanique des Fluides et d'Acoustique*, CNRS

Summary

§ **Introduction**:

Earth's Plasmasphere and Field Line Resonances

Monitoring the Plasmasphere Dynamics: State of the art of FLRs identification

§ **Machine Learning Pipeline**:

Data set

Machine Learning Algorithms

Models Comparison

§ **Results**:

Test set

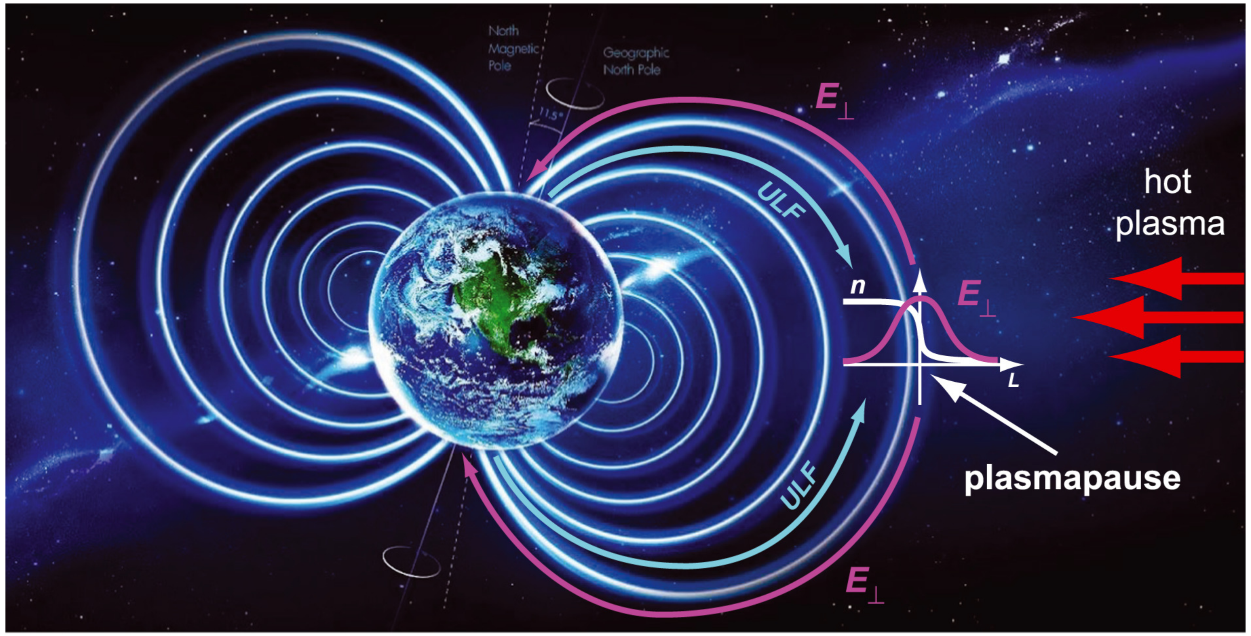
Global Results

Latitudinal Variation

Case Study: Geomagnetic Storm 1st June 2013

Earth's Plasmasphere

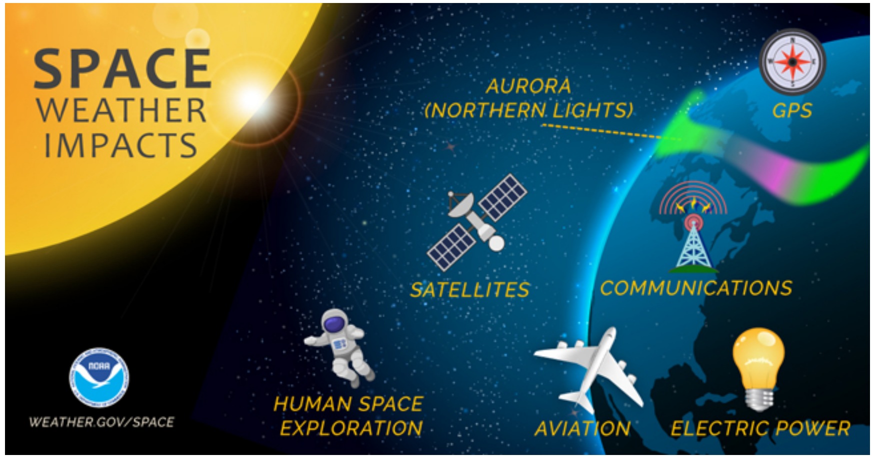
Inner region of the Earth's magnetosphere, it is composed by plasma of ionospheric origin.



Streltsov A.V. and Mishin E.V., 2018

- Composed by dense and cool plasma (E ~ 1 eV)
- Extends approx. from 1.5 to 6 Earth's radii
- Dominated by geomagnetic field, hence it co-rotates with Earth

→ It is the most important region for the **Space Weather**



Field Line Resonances

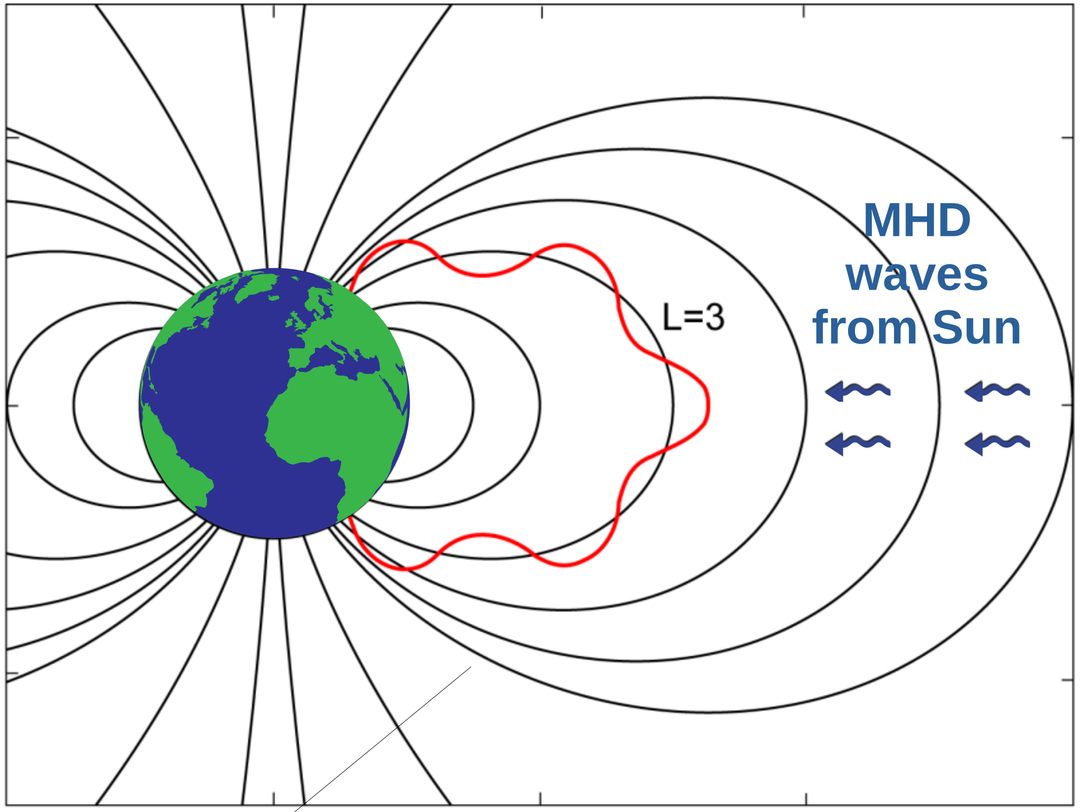
Field Line Resonance (FLR) frequencies can be excited by the interaction between shear Alfvén modes and **MHD compressive waves**. This coupling can produce **standing waves** along a specific geomagnetic field line (L).

$$\mathrm{T}_L \approx \int_{S_1}^{S_2} \frac{ds}{V_A(s)} = 2\mu_0^{1/2} \int_{S_1}^{S_2} \frac{\rho^{1/2}(s)}{B(s)} ds \qquad \text{Eigenperiod } n{=}1$$

where V_A is the Alfvén velocity and $\mathrm{S}_1(\mathrm{S}_2)$ is the initial (end) point of the field line (L).

Assuming:
1. a geomagnetic model (B)
2. a functional form for the density (ρ)
3. the field line length

FLR frequency $\longrightarrow$ **equatorial plasma density**



Dipolar approx of the Geomagnetic Field

FLR Frequencies Identification

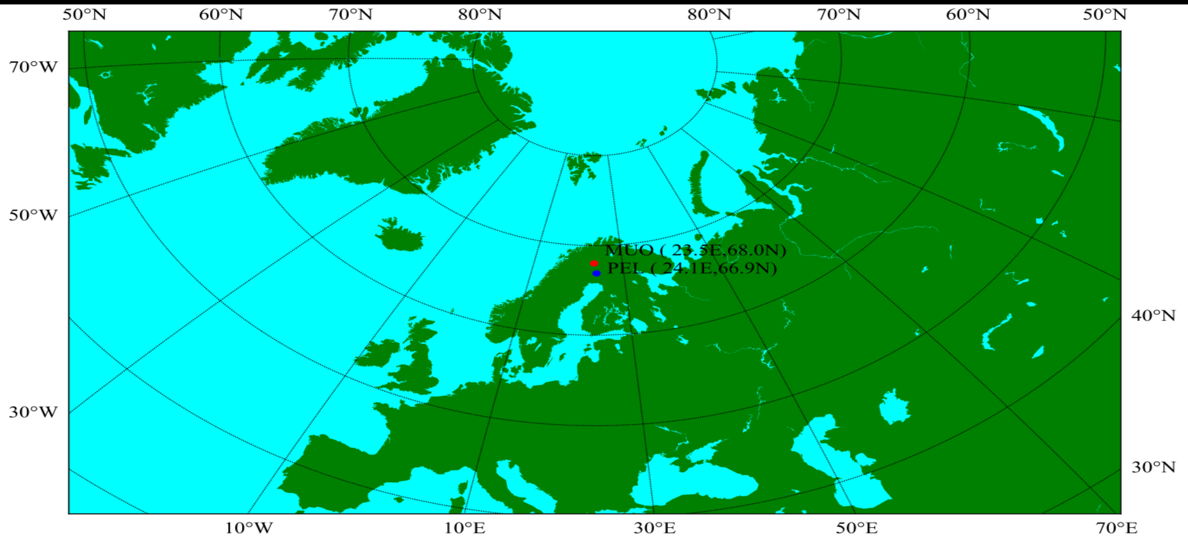
State of the art: **Gradient method** (Waters et al., 1991) from ground-based magnetometers (ULF measurements)

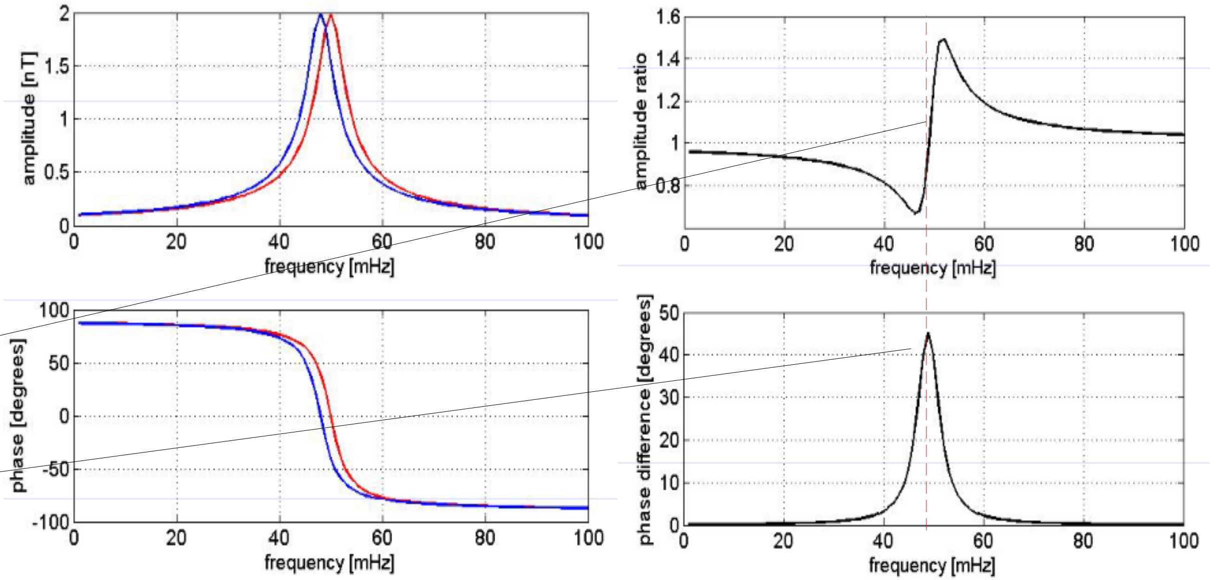
Assuming:

1. Eigenfrequency linearly decreases poleward for stations slightly separated in latitude (this is not true passing through the plasmapause)
2. Meridional aligned stations

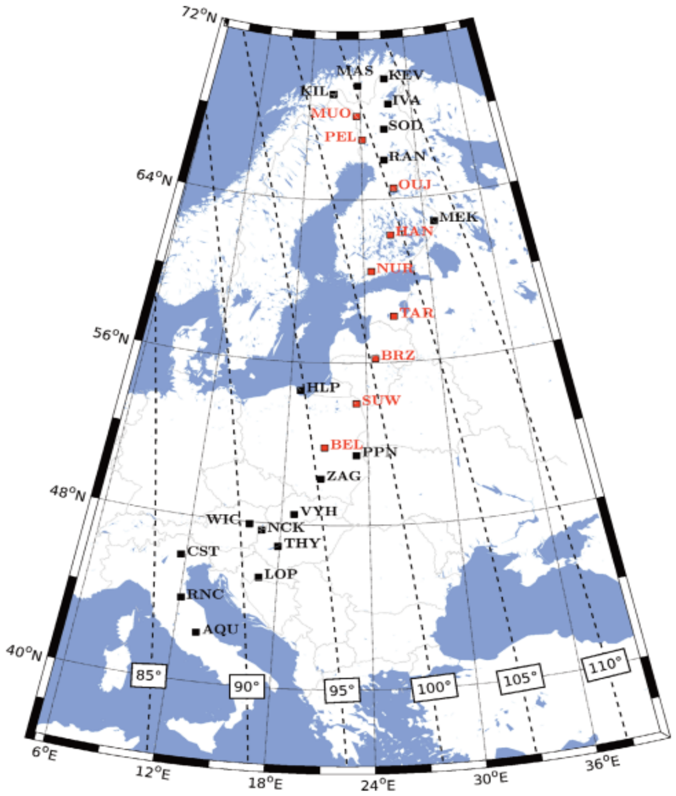
➔ Then it is possible to estimate the FLR frequency of the mid-point (MP) by computing the discrete Fourier cross-spectrum of the two signals.

1. **Cross-Amplitude** crosses unity with positive (negative) slope
2. **Cross-Phase** has its maximum (minimum) value





Field Line Resonances Monitoring



European quasi-**M**eridional **M**agnetometer **A**rray (EMMA)
(~ 30 stations)

EMMA provides magnetic measurements with a resolution of 1s.

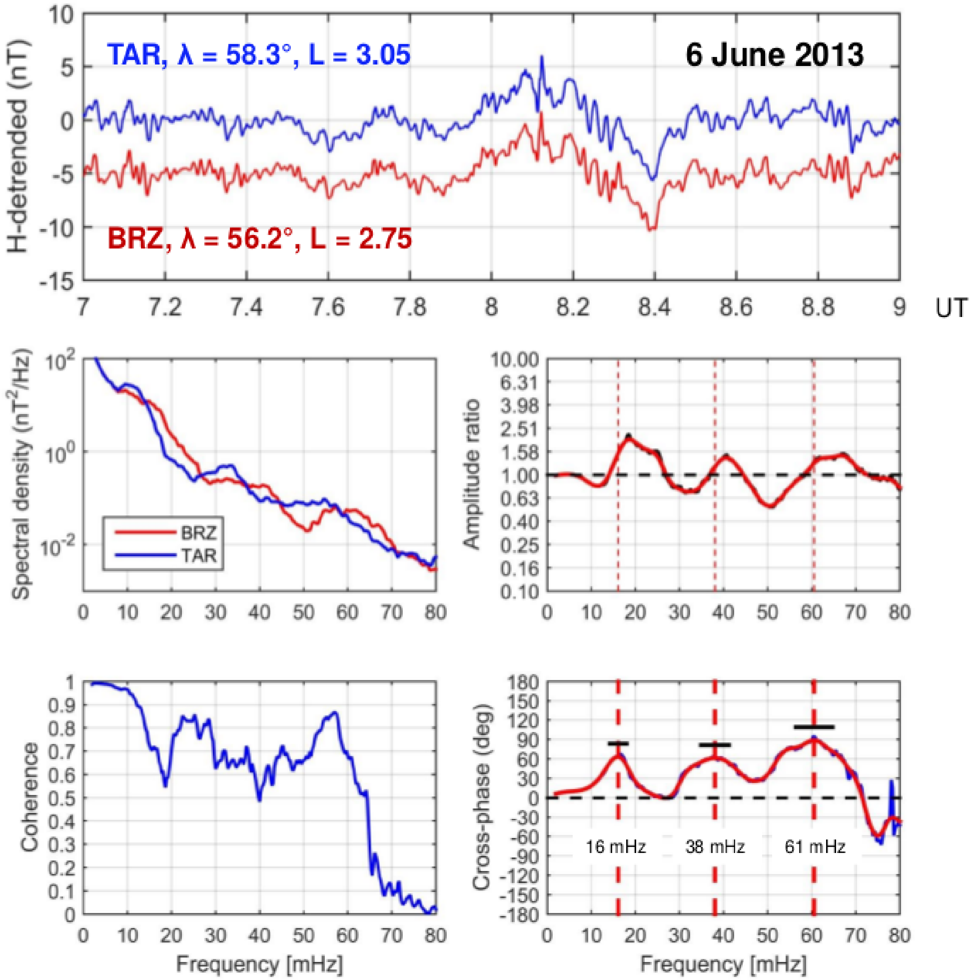
Real-time monitoring of the plasmasphere dynamics

↓

Many authors created (semi-)automated tools for monitoring 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**.

All these methods require the human intervention



Machine Learning Approach

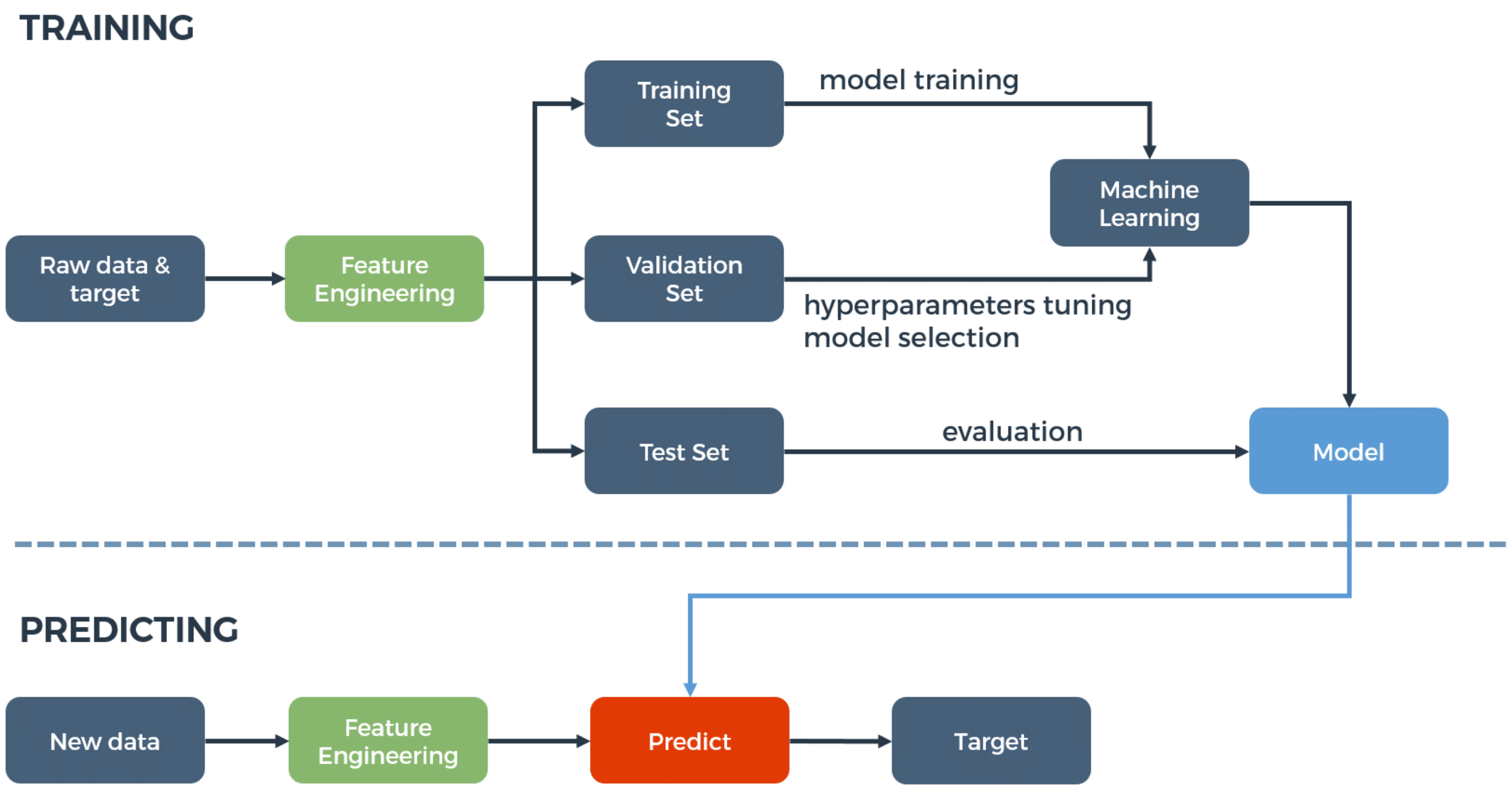
Framework:

FLR frequencies are a powerful tool to sound the cold plasma in the inner magnetosphere. Cross-phase spectra contains sufficient information for estimating FLR frequencies from ground-based ULF measurements.

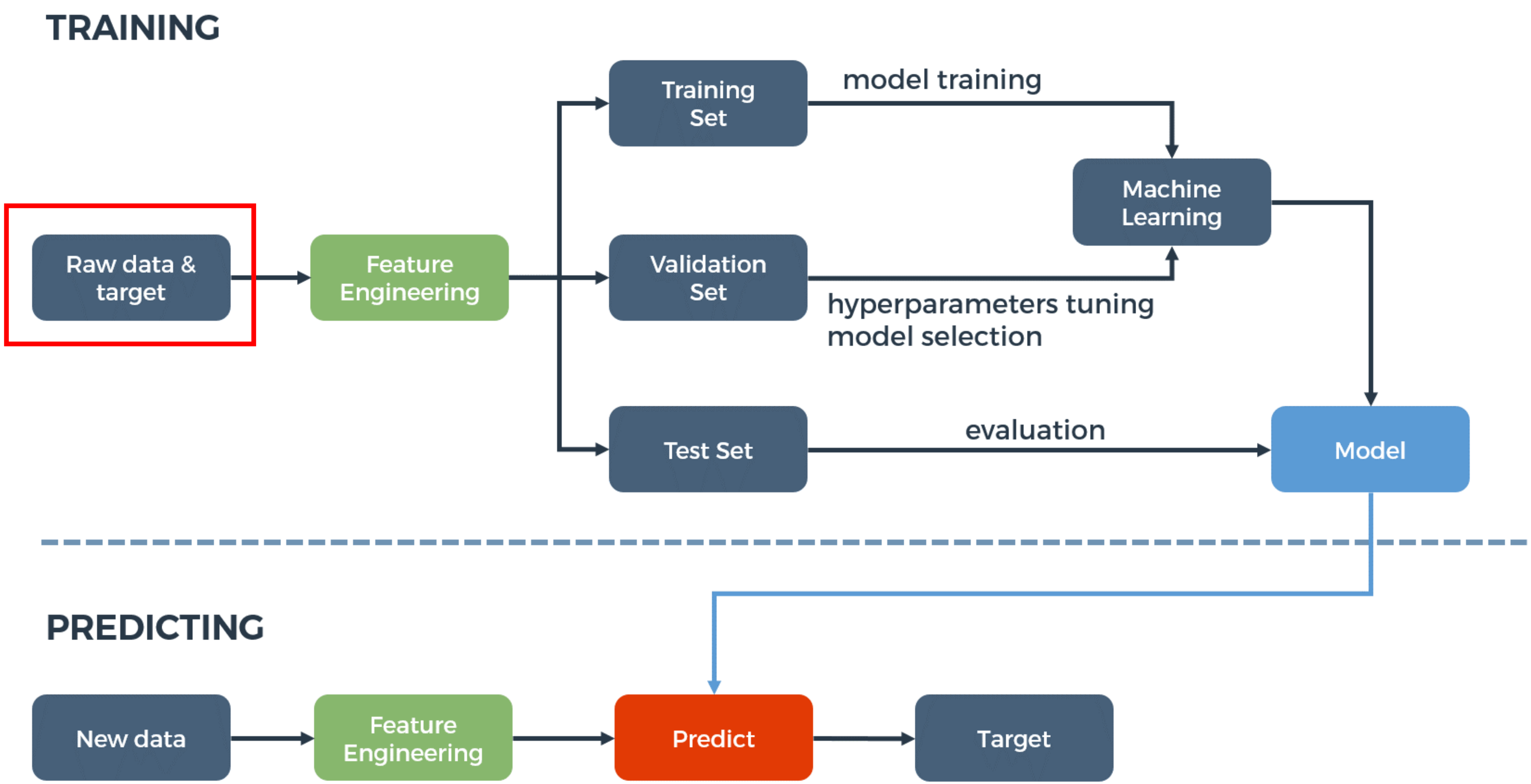
Goal:

Using Machine Learning (ML) methods to build an automated tool for estimating FLR frequencies from cross-phase spectra.

Machine Learning Pipeline



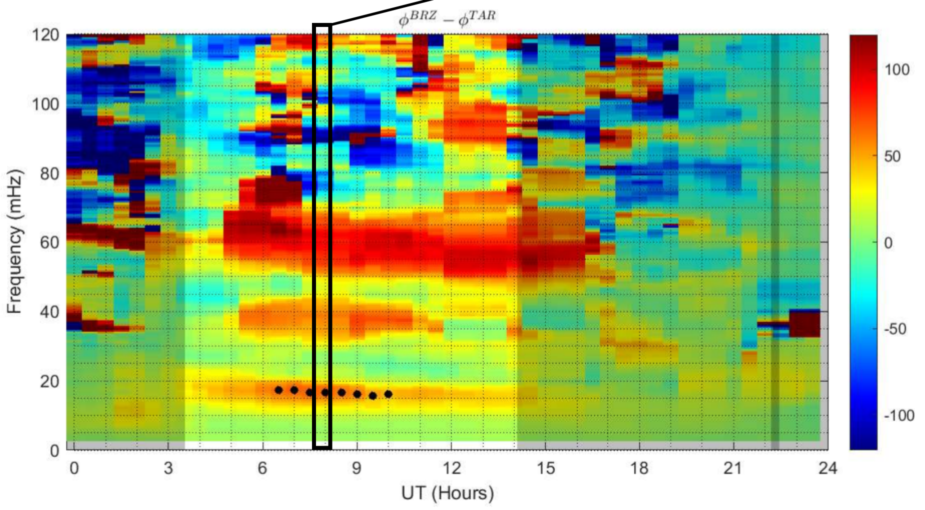
Machine Learning Pipeline



Raw Data & Target: Data Source

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 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)
- 176 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

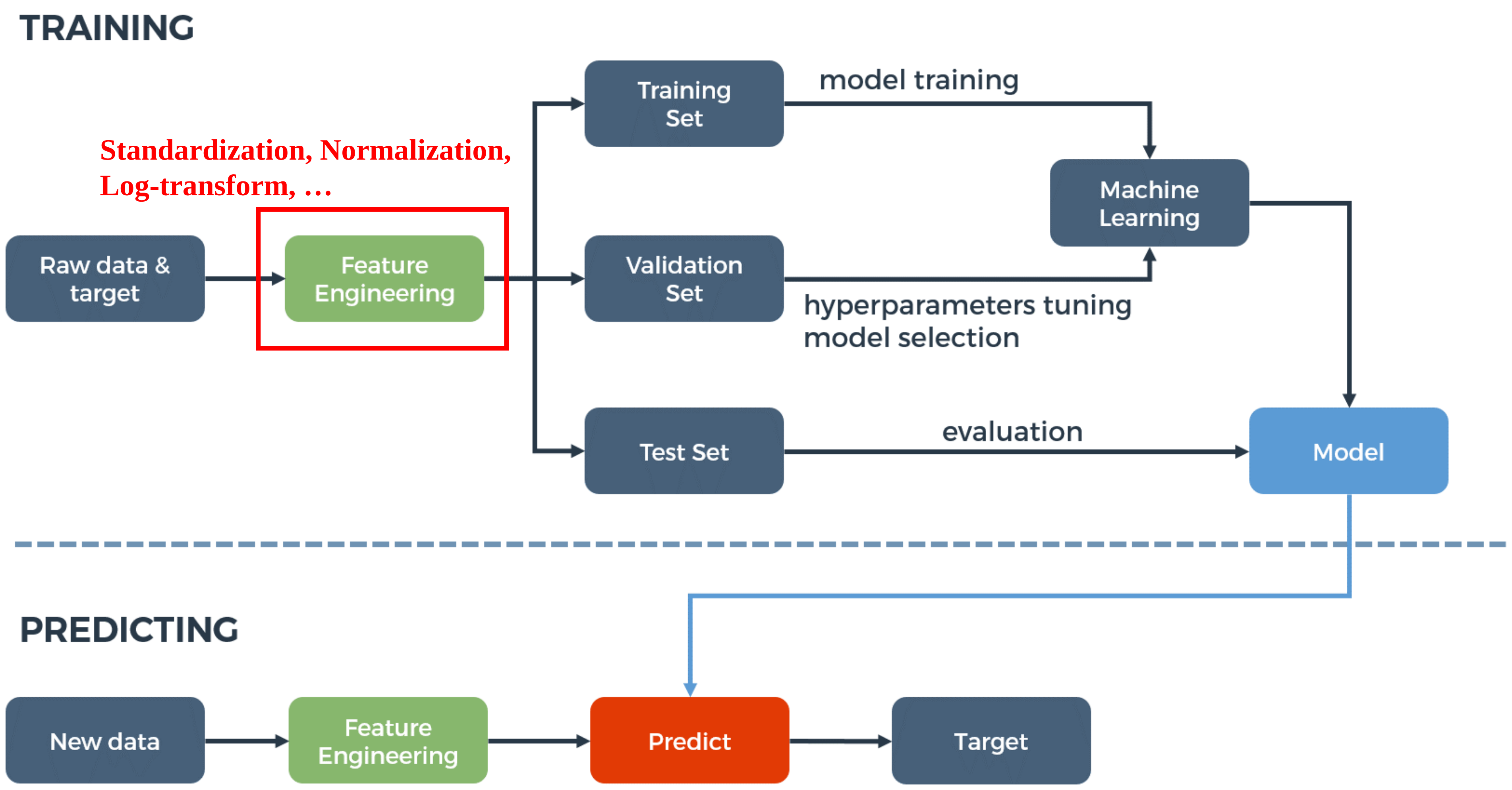


Feature Matrix

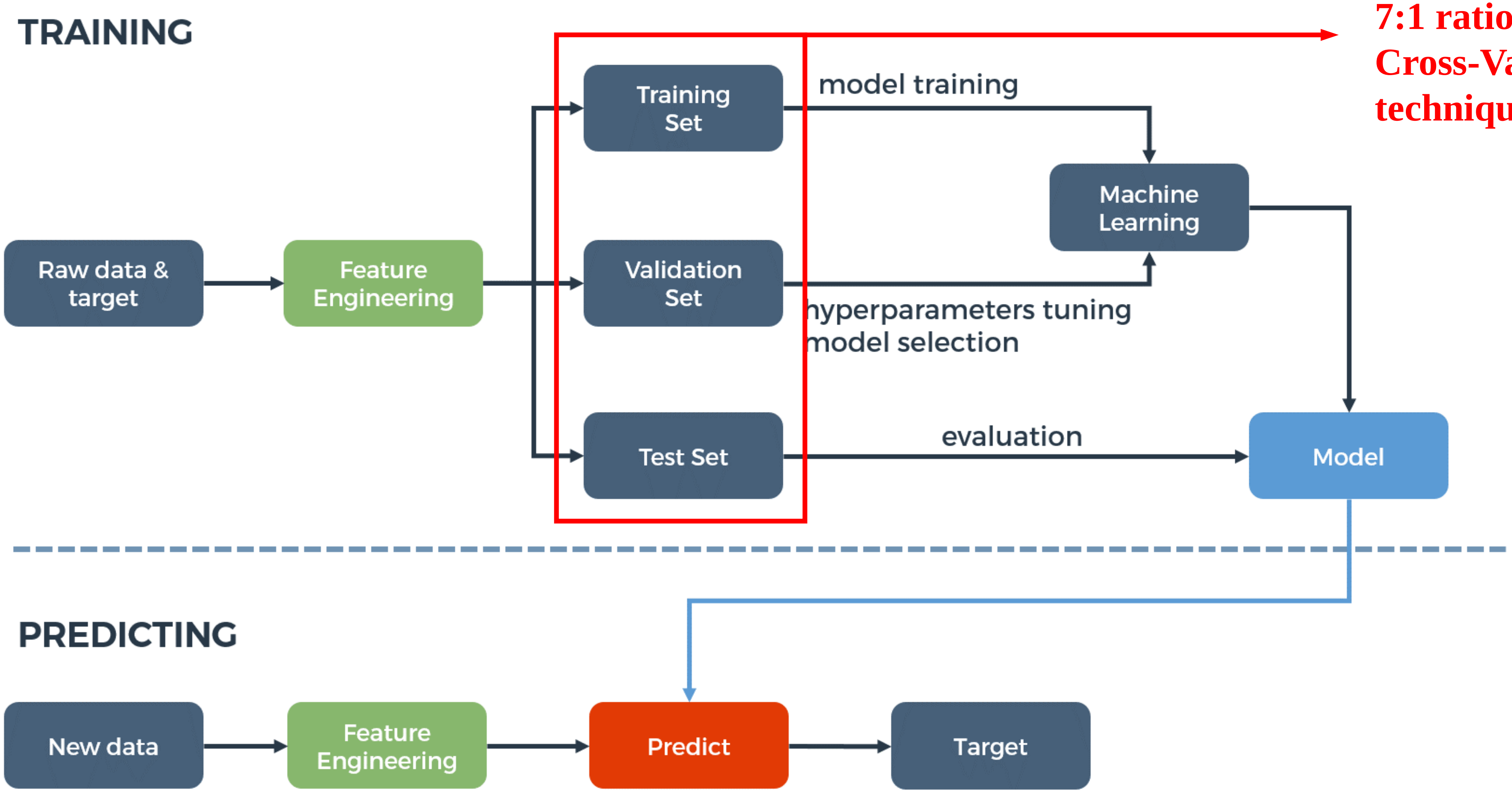
$$
\begin{array}{cccccccl}
\text{Feature-1} & \text{Feature-2} & \text{Feature-3} & \text{Feature-4} & & & \text{Feature-n} & \\
x_1^1 & x_2^1 & x_3^1 & x_4^1 & \dots & \dots & x_n^1 & \text{Sample-1} \\
x_1^2 & x_2^2 & x_3^2 & x_4^2 & \dots & \dots & x_n^2 & \text{Sample-2} \\
x_1^3 & x_2^3 & x_3^3 & x_4^3 & \dots & \dots & x_n^3 & \text{Sample-3} \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \\
x_1^m & x_2^m & x_3^m & x_4^m & \dots & \dots & x_n^m & \text{Sample-m}
\end{array}
$$

INPUT: **Cross-phase spectra**
OUTPUT: **validated FLR frequencies**

Machine Learning Pipeline

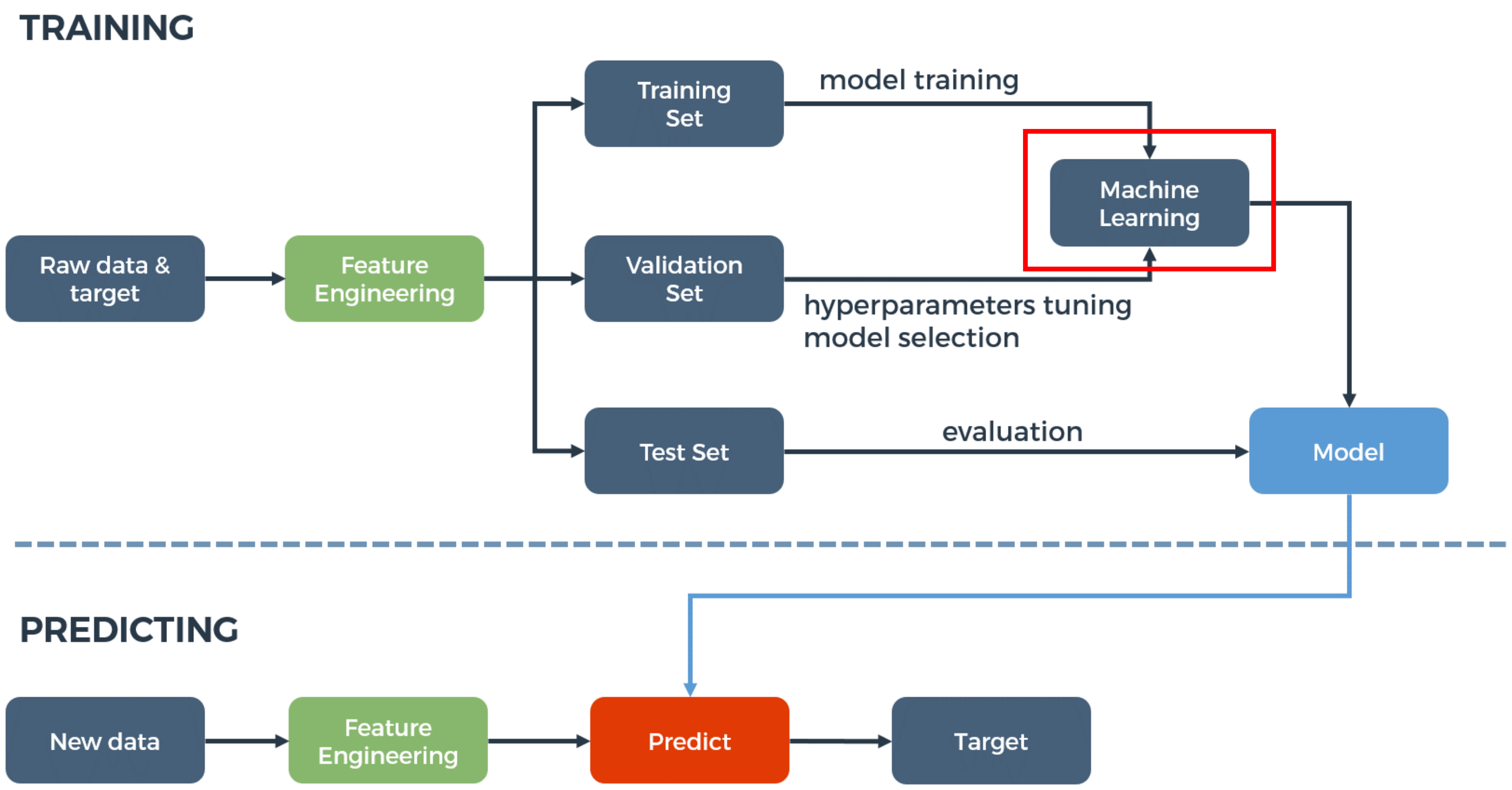


Machine Learning Pipeline



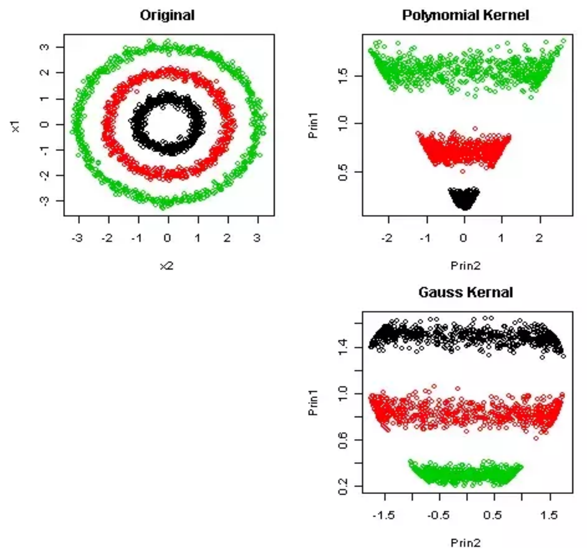
7:1 ratio Train/Test Cross-Validation technique

Machine Learning Pipeline

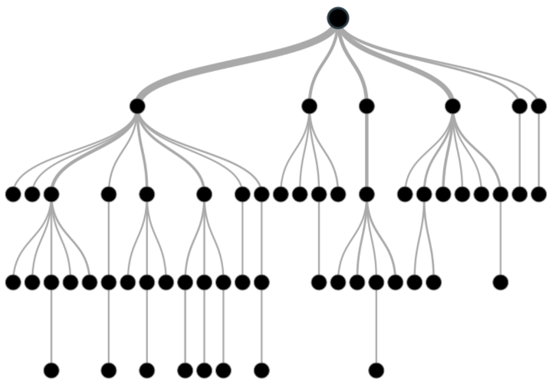


Machine Learning Algorithms

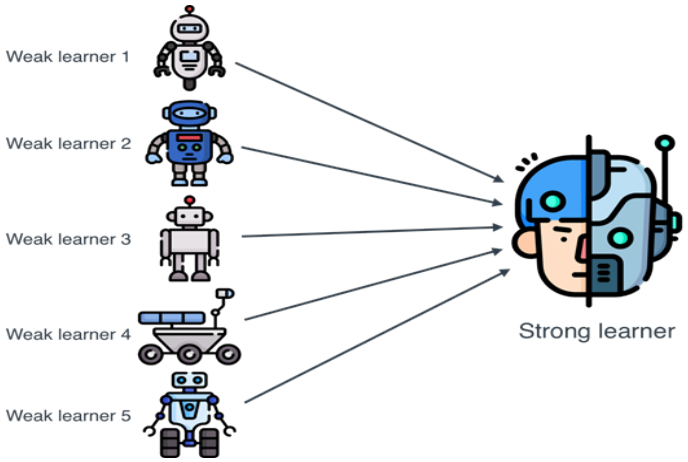
For each pair of stations we evaluate 6 different ML algorithms typical used for regression problems.



1. **Kernel Methods**:
Kernel Ridge (KRR) and *Support Vector Machine* (SVR)

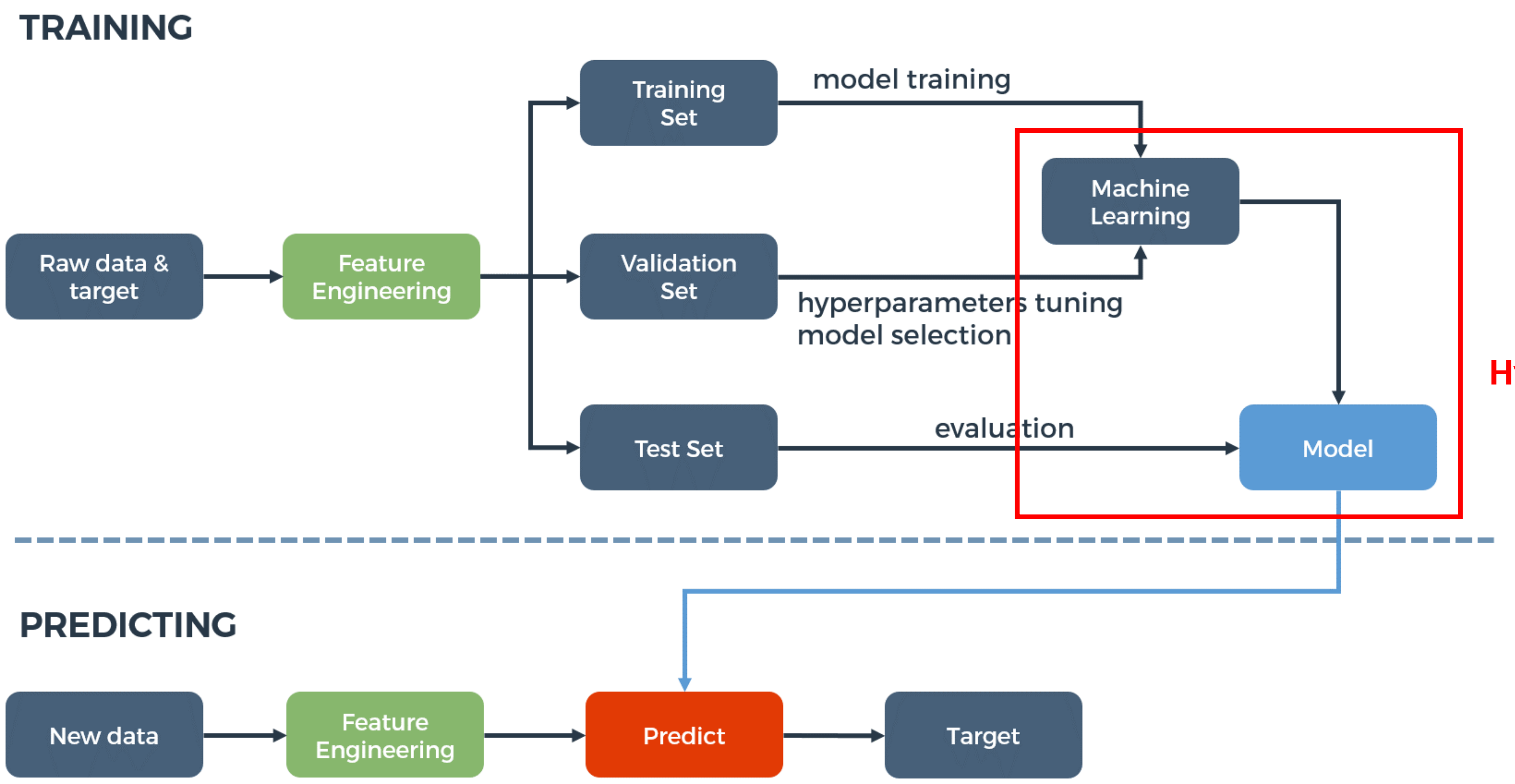


2. **Tree-based Methods**:
Decision Tree (DTR)



3. **Ensemble Methods**:
Random Forest (RF), *Light Gradient Boosting Machine* (LGBM), *Extreme Gradient Boost* (XGB)

Machine Learning Pipeline



CV
+
Hyperp. Tuning

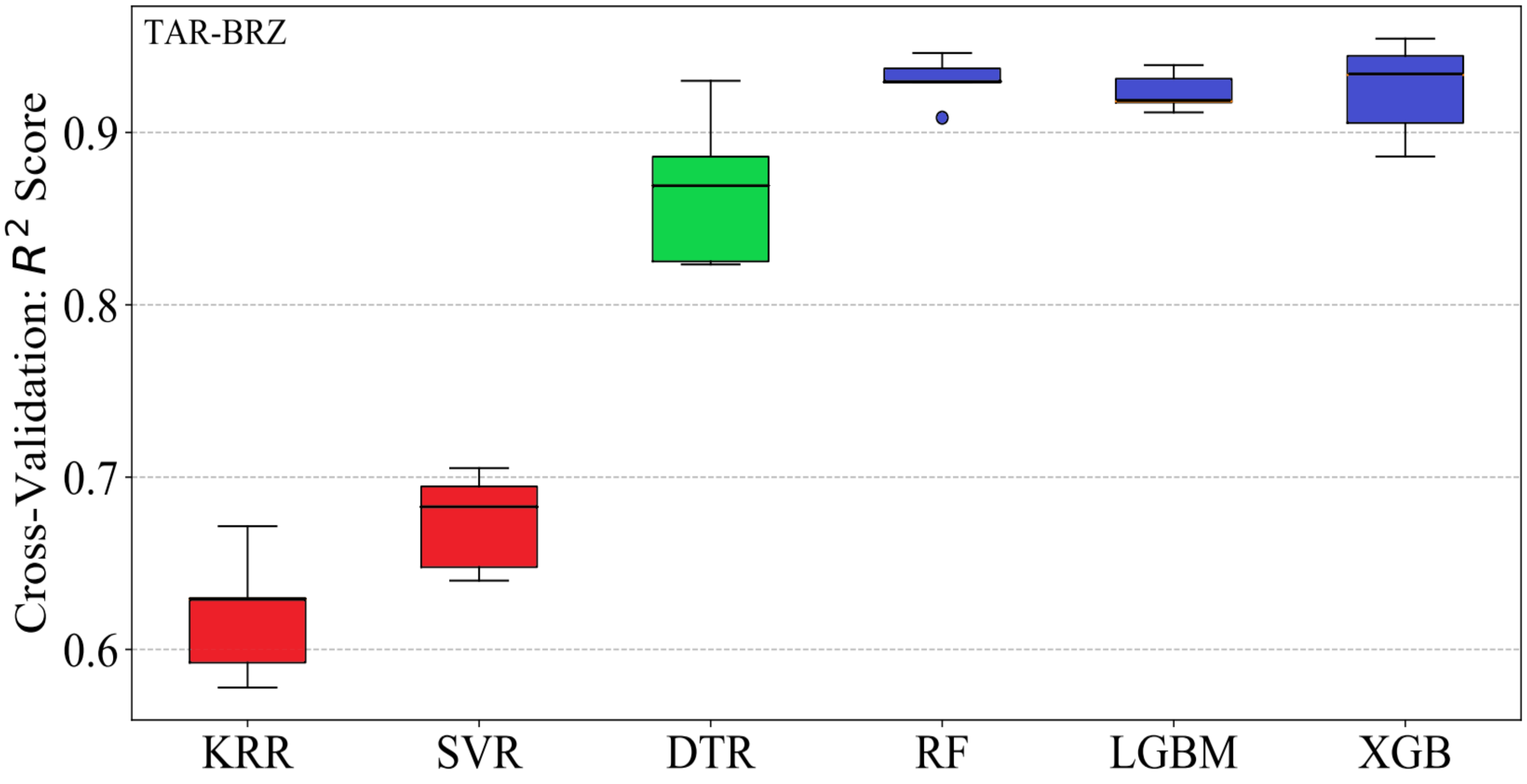
Model Comparison

Results of the cross-validation procedure on the training set:

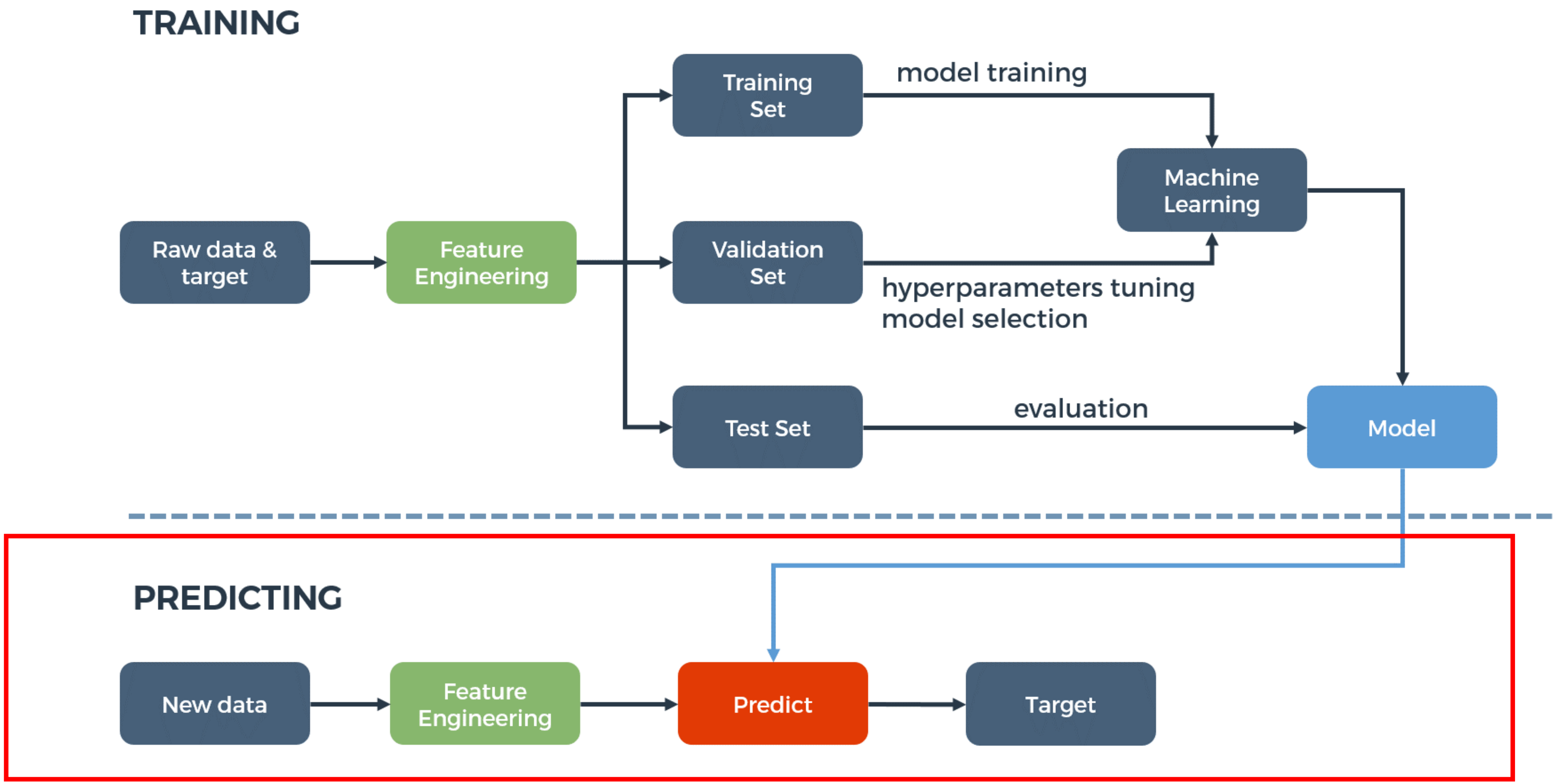
- **Kernel**, **Tree-based** and **Ensemble** methods have significantly different performances

- **Tree-based** methods result better in handling discrete-like data

- **Ensemble** methods are the most suitable with data set with a large number of features (200) wrt the number of samples (4000)



Machine Learning Pipeline

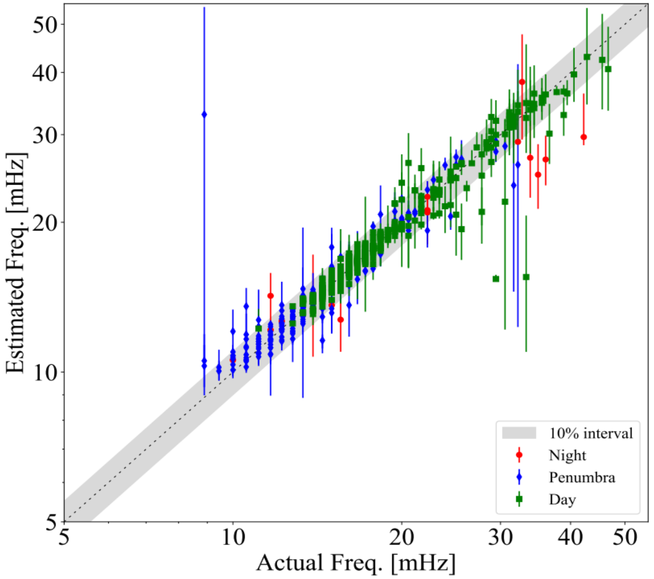


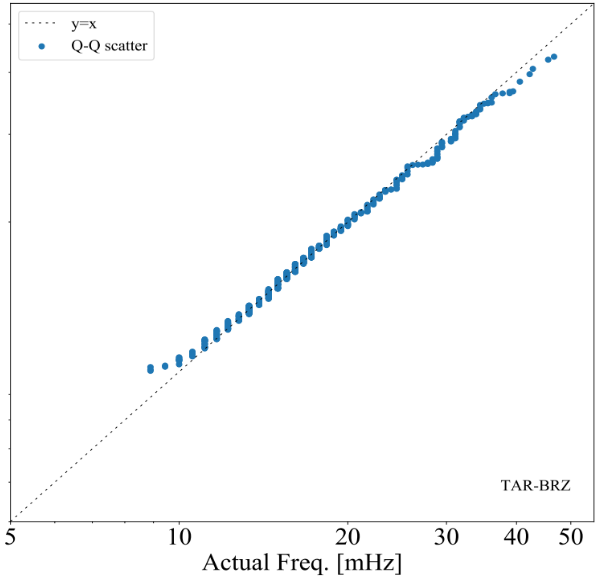
Best Model

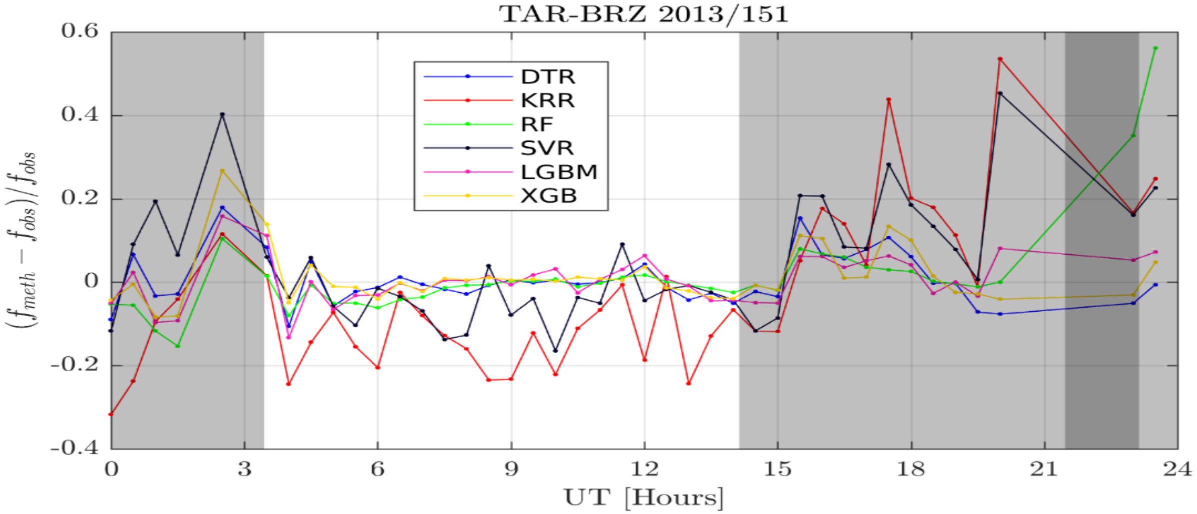
Results 2: Global Results

	R^2	MAE (*mHz*)	MAPE	RMSE (*mHz*)	CV Time (s)
KRR	0.613	2.42	0.128	3.9	2.83
SVR	0.688	2.00	0.106	3.5	17.8
DTR	0.828	1.00	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

- Estimation error does not increase with increasing frequency (top panel)
- All models have higher estimation errors during nighttime (dark-grey area), or when one of the two footprints is nightside (light-grey area)



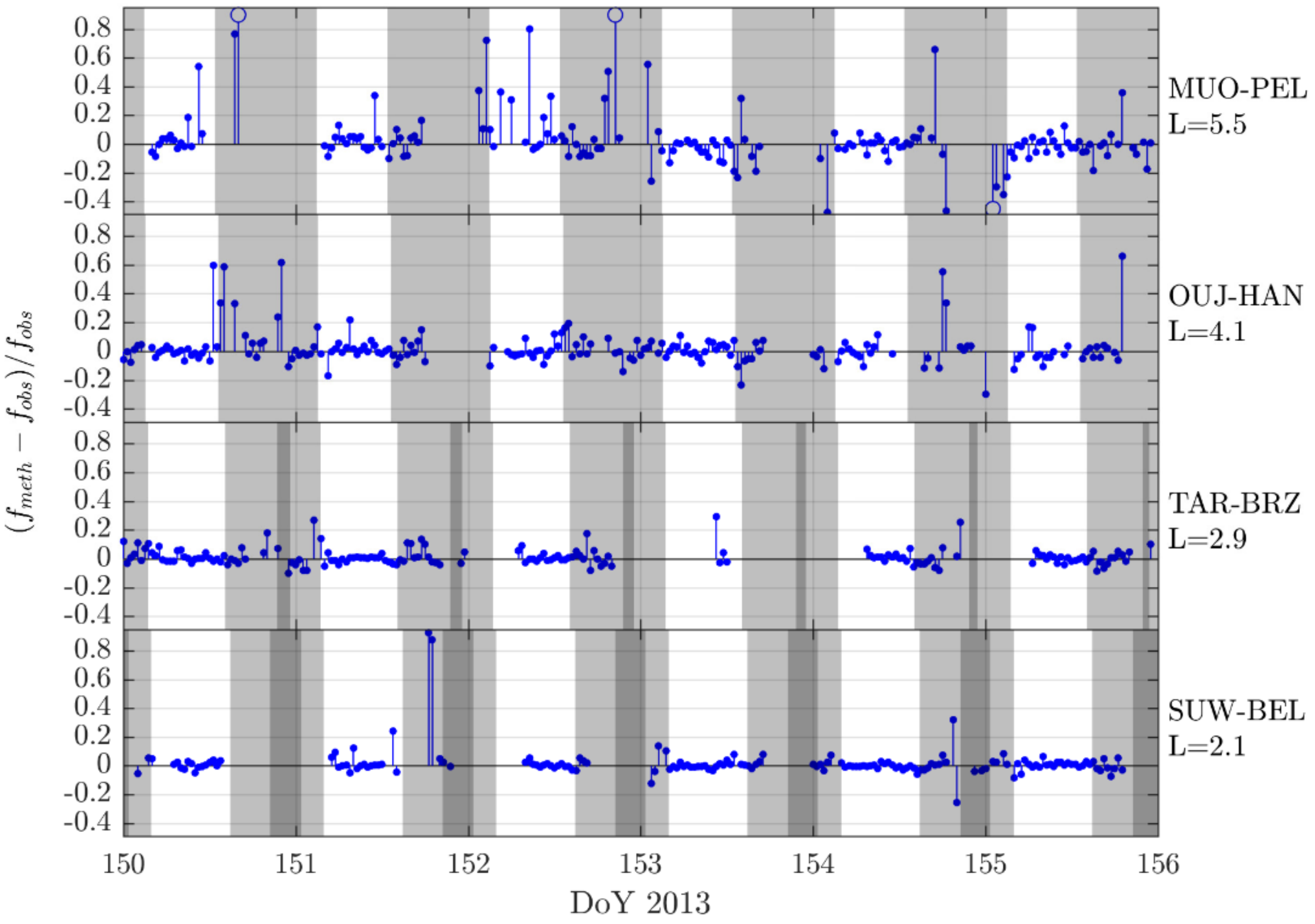




Results 3: Latitudinal Variation

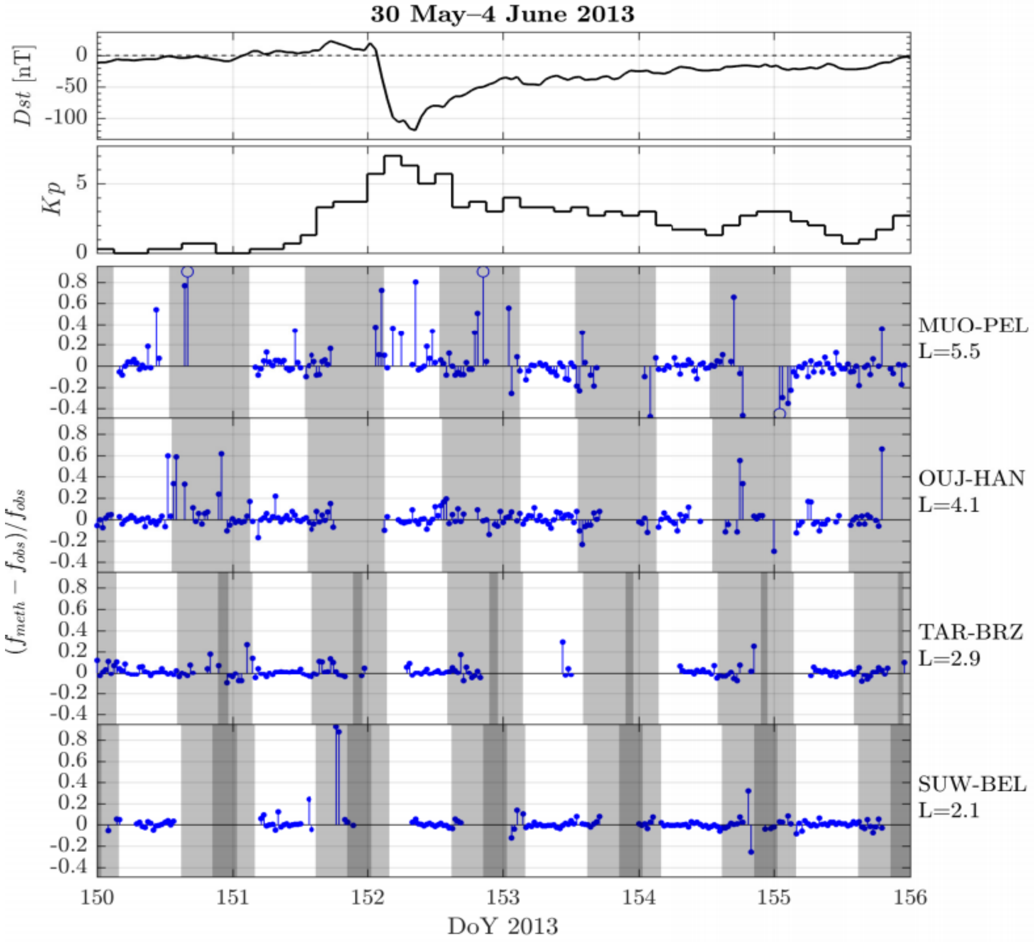
Relative estimation error of the four pairs of stations for six consecutive days.

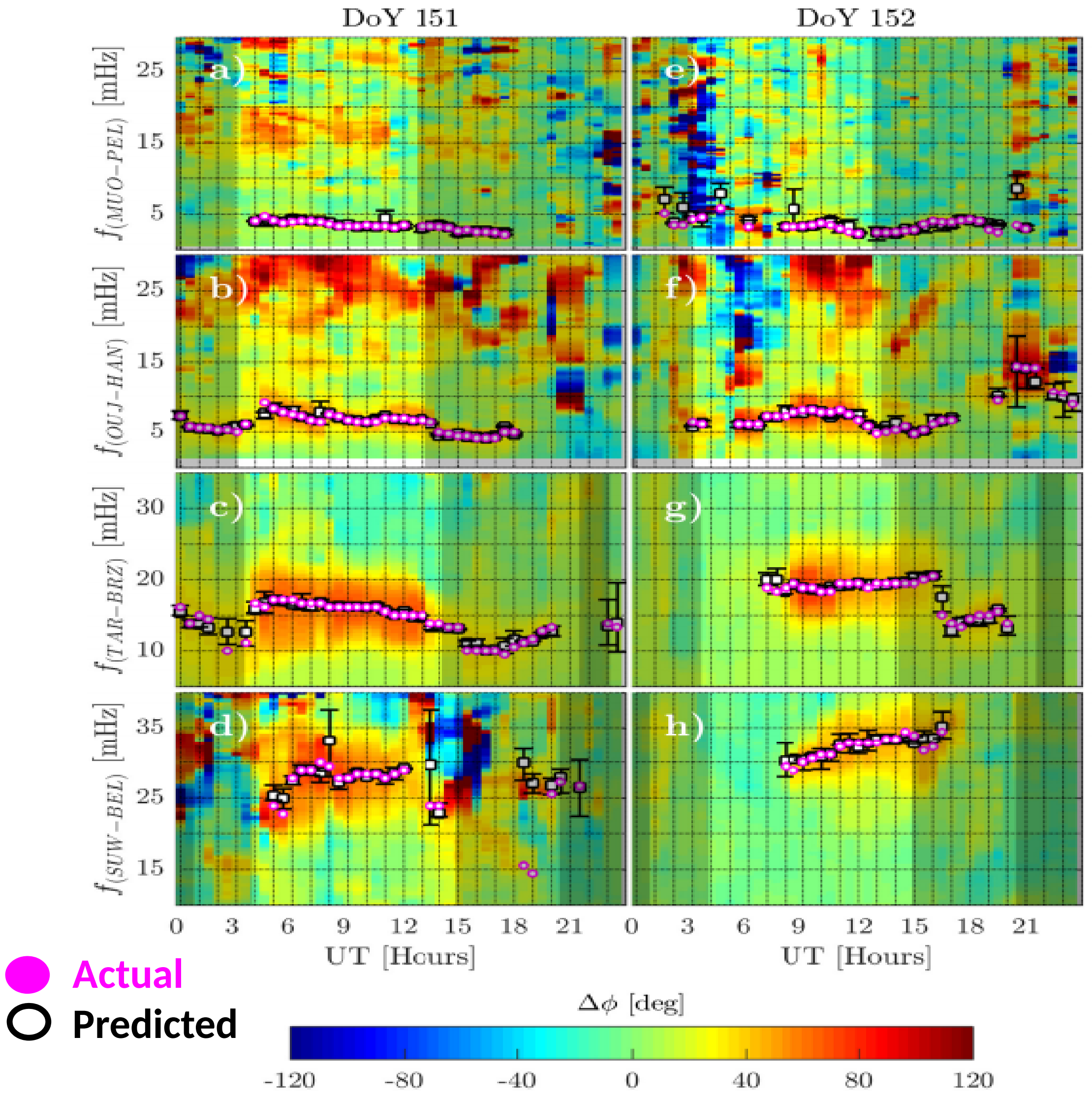
- Error slightly increases with increasing L probably because of fuzzier cross-phase spectra
- Average relative error is +1-2% from L=2.1 to L=4.1, for MUO-PEL is 4.5% meaning that overestimation errors have a heavier weight.
- At every L we can observe that the error is higher during nighttime



Results 4: Case Study

Except for MUO-PEL, the estimation error **does not show** any **dependence from the geomagnetic activity level**.





Conclusions 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 determines when FLRs can be observed from signals.

- Our final goal is to create a single ML tool which includes all the EMMA stations (even other magnetometer array) and which is able to determine FLR frequency directly from spectrograms (CNN)

References

- R. Foldes, A. Del Corpo, E. Pietropaolo and M. Vellante, *Assessing Machine Learning Techniques for Identifying Field Line Resonance Frequencies From Cross-Phase Spectra*, 2021, *JGR: Space Physics*

- J. Lichtenberger, M. A. Clilverd, B. Heilig, M. Vellante, J. Manninen, C. J. Rodger, A. B. Collier, A. M. Jørgensen, Jan Reda, R. H. Holzworth, R. Friedel and M. Simon-Wedlund, *The plasmasphere during a space weather event: first results from the PLASMON project*, 2013, *JSWSC*

- P. J. Chi, M. J. Engebretson, M. B. Moldwin, C. T. Russell, I. R. Mann, M. R. Hairston, M. Reno, J. Goldstein, L. I. Winkler, J. L. Cruz-Abeyro, D.-H. Lee, K. Yumoto, R. Dalrymple, B. Chen, and J. P. Gibson, *Sounding of the plasmasphere by Mid-continent MAgnetoseismic Chain (McMAC) magnetometers*, 2013, *AGU JGR*

- A. Del Corpo, M. Vellante, B. Heilig, E. Pietropaolo, J. Reda, J. Lichtenberger, *Observing the cold plasma in the Earth's magnetosphere with the EMMA network*, 2018, *Ann. of Geop.*

- S. J. Wharton, D. M. Wright, T. K. Yeoman, M. K. James, J. K. Sandhu, *Cross-phase determination of ultralow frequency wave harmonic frequencies and their associated plasma mass density distributions*, 2018, *AGU JGR*

- A.V. Streltsov and E.V. Mishin, *On the existence of ionospheric feedback instability in the Earth's magnetosphere-ionosphere system*, 2018, *AGU JGR*