Identification of Electron Diffusion Regions with an AI approach

Quentin Lenouvel

P. Garnier, V. Génot, S. Toledo-Redondo, B. Lavraud and MMS team

IRAP, Toulouse

May 2020, EGU

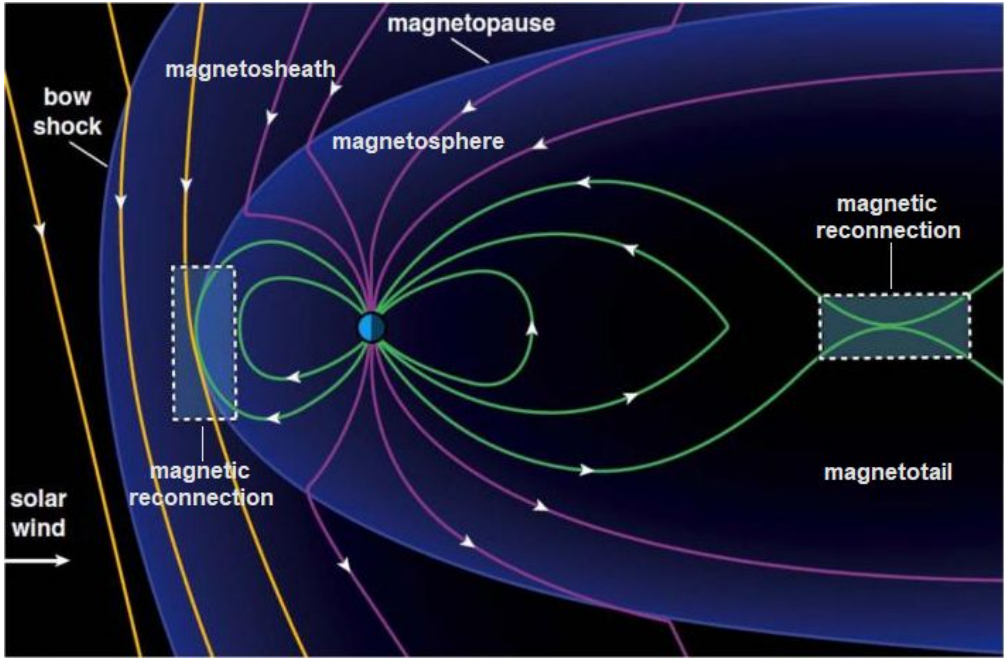
Introduction

- **Purpose of the work** : Automatic detection of **Electron Diffusion Regions** (EDR) and other plasma regions of interest with **Machine Learning**
- **How** : Training of a **Neural Network** on *in situ* MMS data from **phase 1** to study and understand complex relationships between several physical parameters. **Predictions** of the algorithm **on magnetopause crossings intervals** (listed in the ISSI team's magnetopause crossings database) **from phase 1a**
- **Why** : The identification of EDR events is hard (**32 dayside reported events** at the moment)

Find **more EDR events** => **Better understanding of magnetic reconnection**

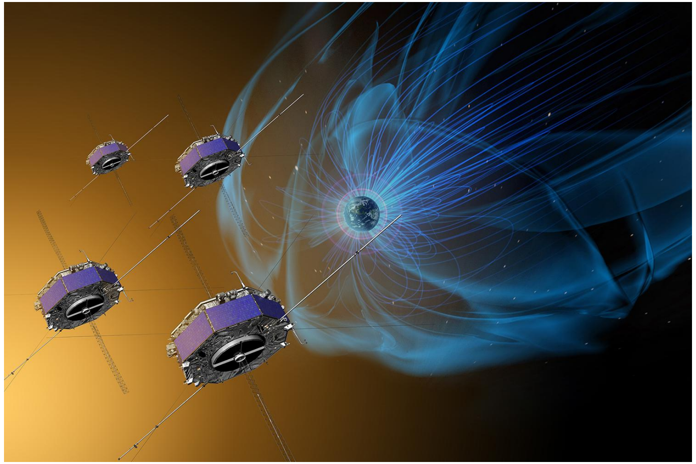
Earth’s plasma environment

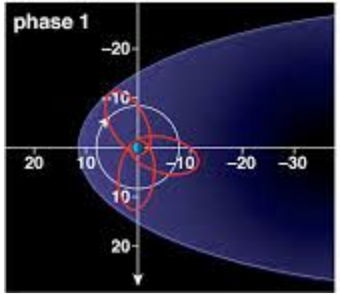
- **Plasma regions :**
  - Solar wind
  - Bow shock
  - Magnetosheath
  - Magnetopause
  - Magnetosphere
  - Magnetotail
- **Magnetic Reconnection** is a major energy transfer process that can happen around the **magnetopause** and in the **magnetotail**

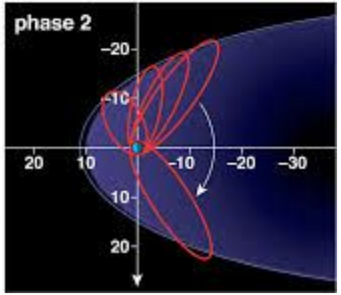


Magnetic reconnection : MMS

- Magnetospheric Multiscale (MMS) mission launched by NASA in **March, 2015**
- **Study Magnetic Reconnection** near Earth's magnetosphere
- Resolution of the instruments allowing for the first time the study of Electron Diffusion Regions
- Use of 4 identical spacecraft able to study :
  - **Electric and Magnetic fields**
  - **Particles** (electrons and ions)
- **Phase 1 orbit** : Mainly dayside magnetopause

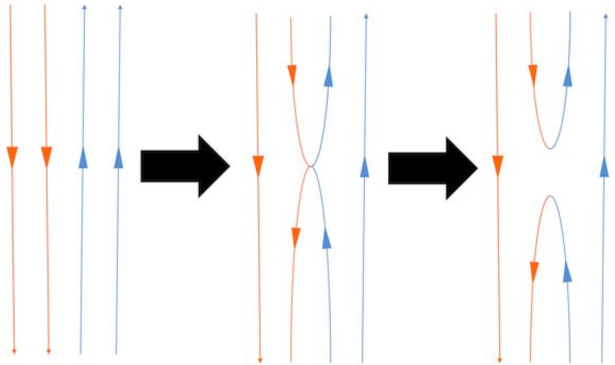


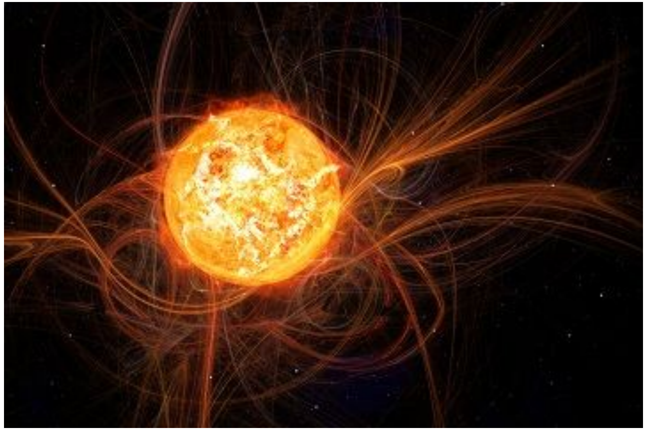




Magnetic reconnection : Concept

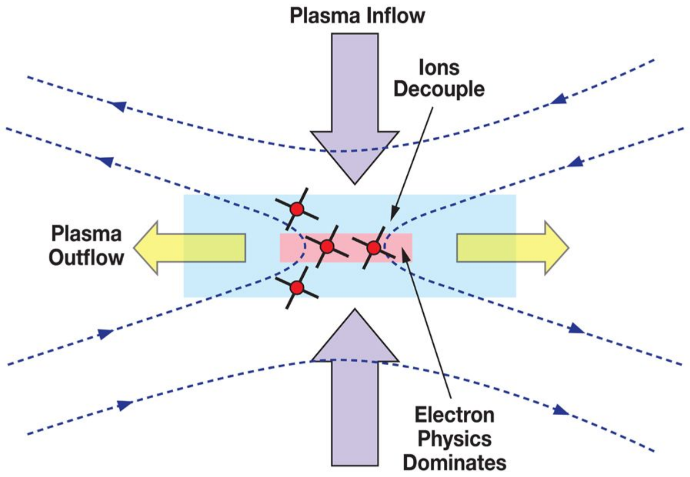
- **Magnetic Reconnection** : **Modification of the topology** of magnetic field lines
  ⇒ **Conversion of magnetic energy into kinetic energy** for particles
- **Recent physical concept** introduced first in the 50's, first model from **Sweet** and **Parker**
- Ubiquitous in many other astrophysical context such as **Solar Flares** or certain models of **Gamma Ray Bursts emissions**, and it is also one of the main problems to achieve **nuclear fusion**

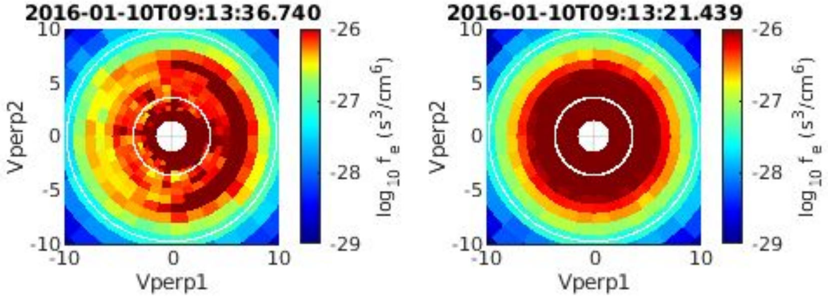


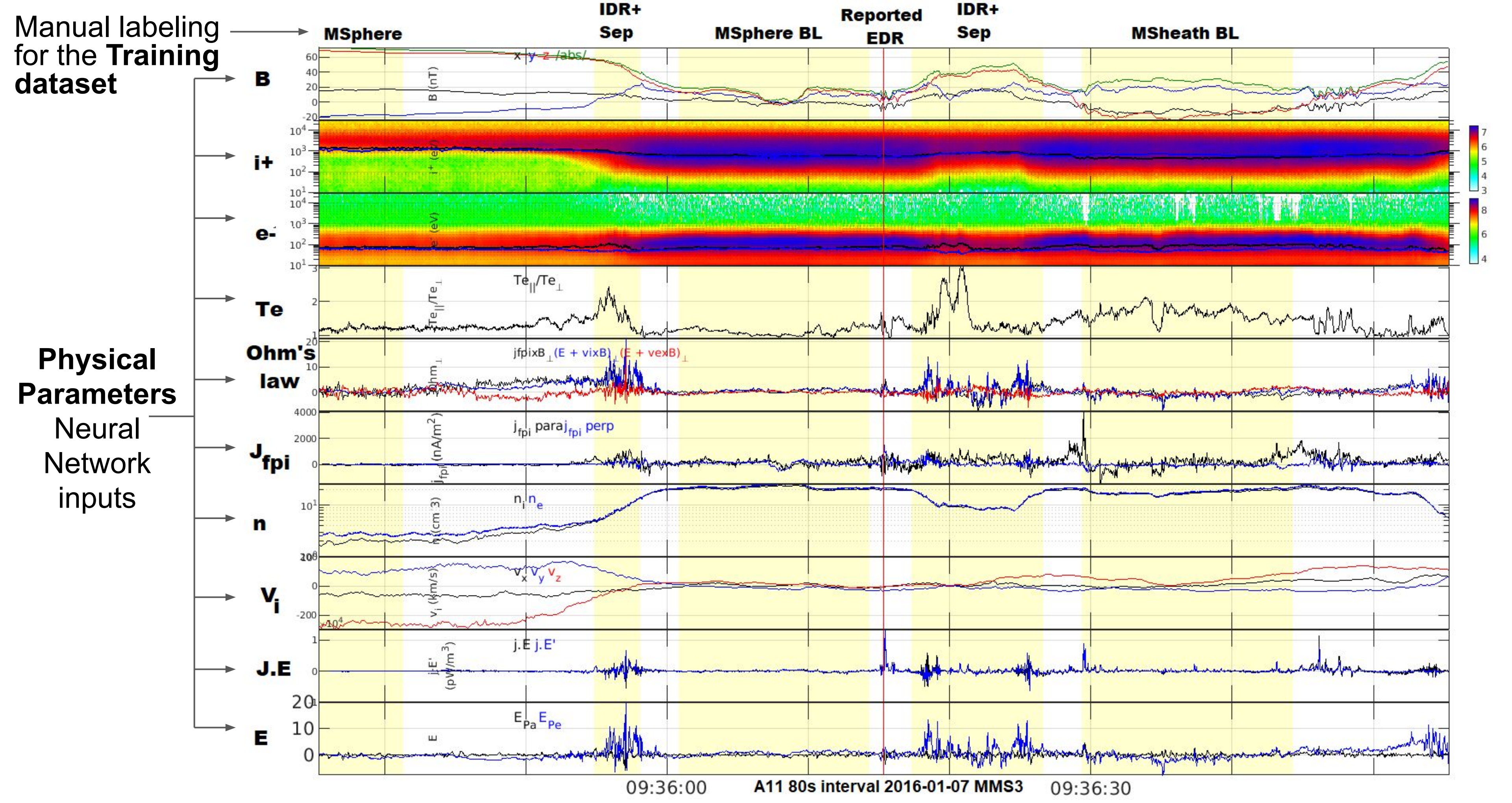


Magnetic reconnection : EDR

- The **Electron Diffusion Region** is the heart of the magnetic reconnection process
- **32 reported EDR encounters (dayside)**, listed in **Webster et al. [2018]**
- Reconnection scale larger than the diffusion region :
  - Ion Diffusion Region [**IDR**] : 50 to 500 km
  - Electron Diffusion Region [**EDR**] : 1 to 10 km (observable for the 1st time thanks to MMS)
- Presence of **crescents** (agyrotropy) in the **electron distribution functions** of **EDRs** (Hesse et al. [2014], Burch et al. [2016])







Manual labeling for the **Training dataset**
MSphere
IDR+ Sep
MSphere BL
Reported EDR
IDR+ Sep
MSheath BL
B
i+
e-
Te
Ohm's law
Jfpi
n
Vi
J.E
E
Physical Parameters
Neural Network inputs
x y z /abs/
B (nT)
Te||/Te⊥
jfpixB⊥ (E + vixB)⊥ (E + vexB)⊥
jfpi para jfpi perp
ni ne
vx vy vz
j.E j.E'
EPa EPe
09:36:00
A11 80s interval 2016-01-07 MMS3
09:36:30

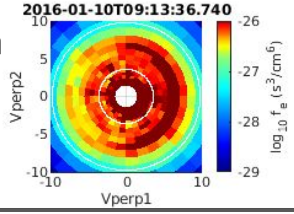
Neural Network for Supervised Learning

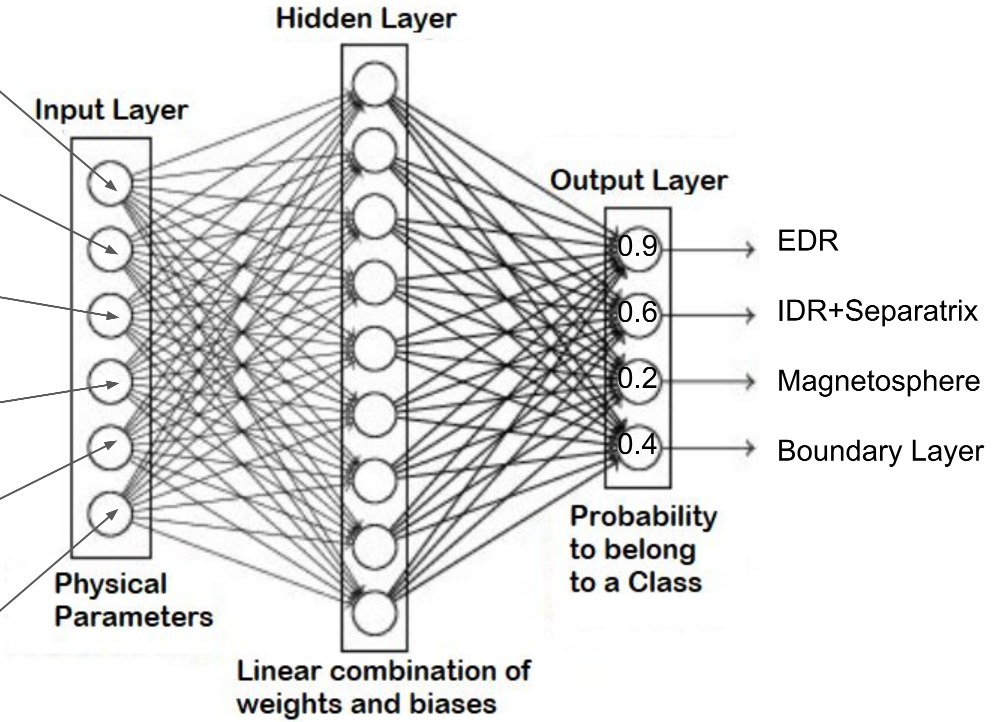
- **16 classical plasma parameters** :

B_z	$E_{\parallel}$	$(E+v_e xB)_{\perp}$	$V_{e\,x,y,z}$	$j_{fpi\,\parallel}$	$j_{fpi}xB_{\perp}$	$Te_{\parallel}$
|B|	$E_{\perp}$	$(E+v_i xB)_{\perp}$	n_e	$j_{fpi\,\perp}$	J.E	$Te_{\perp}$

- **+1 artificial parameter : Mean(R/L)**

Scalar measuring the **electron agyrotropy** (asymmetry) in the electron distribution functions



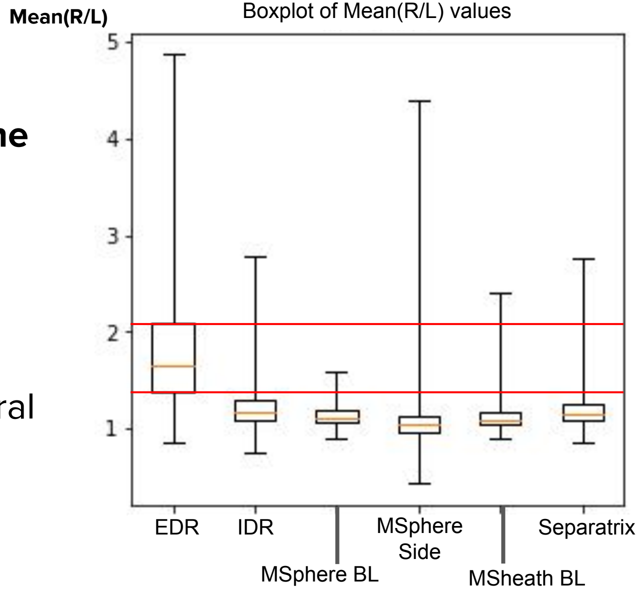


- **Architecture** : Use of a Feedforward **Multilayer Perceptron** (MLP)

Input physical parameters for the Neural Network

- **<u>Mean(R/L)</u>** : Scalar measuring the **asymmetry of the electron distribution function** => $\frac{\text{Average intensity of Right pixels}}{\text{Average intensity of Left pixels}}$ between [40 eV, 275 eV] after normalisation of each pixel by pixel ring intensity

- **Mean(R/L) values** generally **higher for the EDR** class than for the others
  => Very important parameter for the neural network !



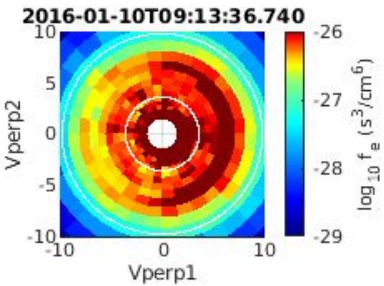


Image with no treatment

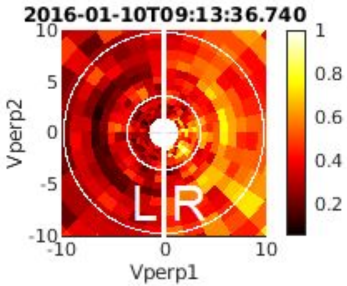


Image after pixel ***normalisation by ring***

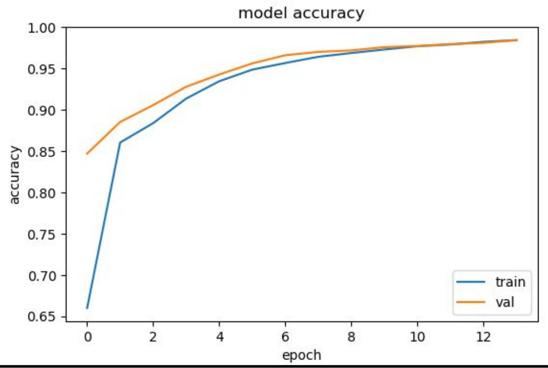
- **Electron agyrotropic index** $\sqrt{Q}$ (Swisdak et al. [2016]) found **inefficient :** worst results for the training and for the predictions with this parameter included

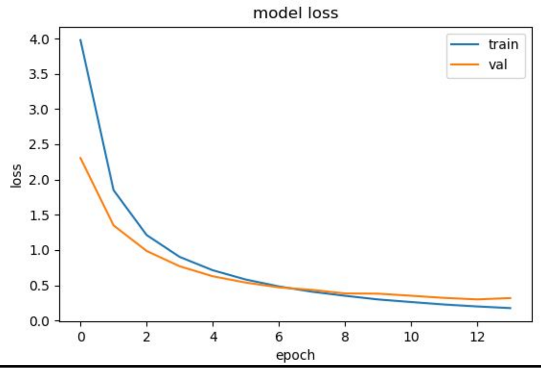
Training of the neural network

- **Manual labeling** on 32 events from Webster et al. (~80s burst data each), **to build our training dataset,** each of the **4 spacecraft is considered independently**

- **<u>Splitting of the data into 3 sets</u>** :
    - **Training** (60% of each class) : Data points that the algorithm will **learn** from
    - **Validation** (20% of each class) : Data points used to **control** the training
    - **Testing** (20% of each class) : Data points used to **evaluate** the performance of the algorithm

- **<u>Training curves</u> :** Show the accuracy and the learning curve for each epoch during the training (no overfitting here using early stopping technique !)





Evaluation of the performance of the algorithm

- **Evaluation of the algorithm with different metrics** :
  - Precision = Tp/(Tp+Fp) ⇒ Low Precision = overestimation of the number of instances of the class
  - Recall = Tp/(Tp+Fn) ⇒ Low recall = overlooking of a lot of instances of the class
  - F1-score = 2*(P*R)/(P+R) ⇒ Harmonic mean of Precision and Recall

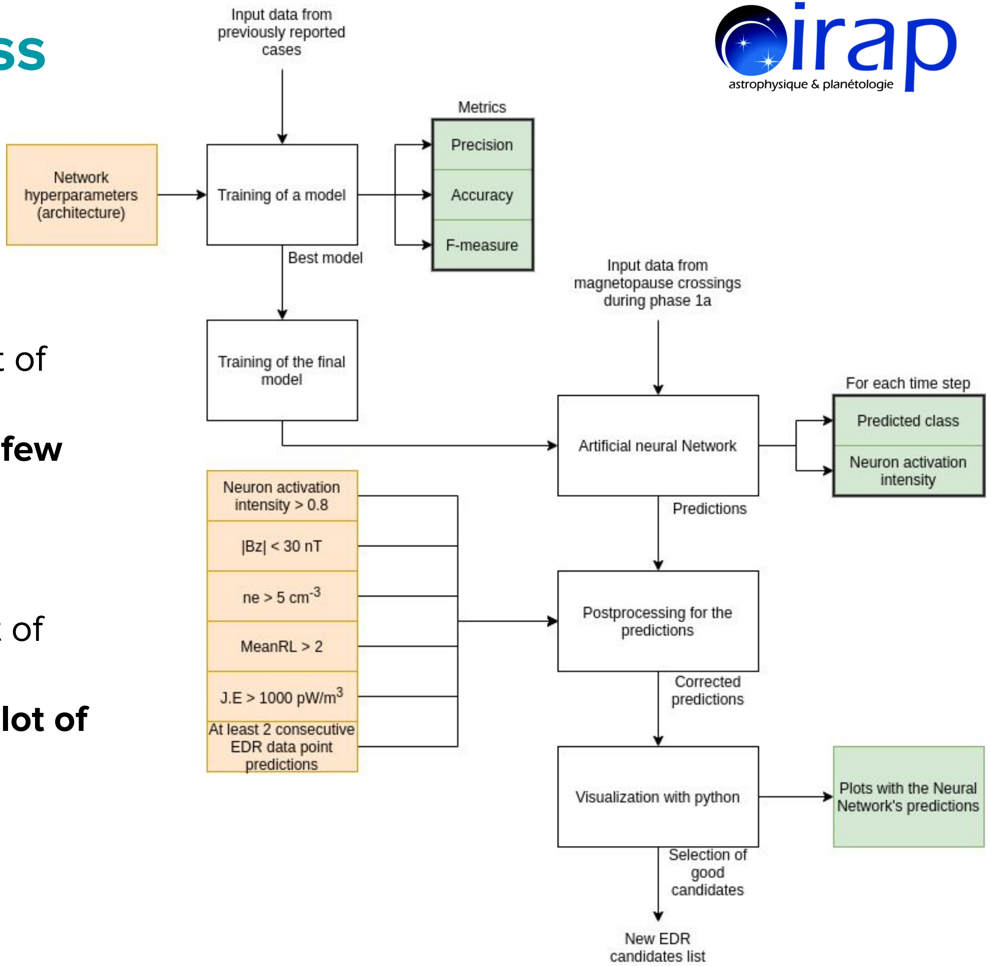
Class	Population	Precision	Recall	F1-score
EDR	**40**	**83%**	**95%**	**88%**
IDR + Separatrix Region	1008	97%	97%	97%
Magnetosphere	840	98%	100%	99%
Boundary Layer	2574	99%	99%	99%

Confusion Matrix	P. EDR	P. IDR	P. MSp	P. BL
T. EDR	**38**	**2**	0	0
T. IDR	**8**	**974**	**8**	**18**
T. MSp	0	0	**840**	0
T. BL	0	**29**	**6**	**2539**

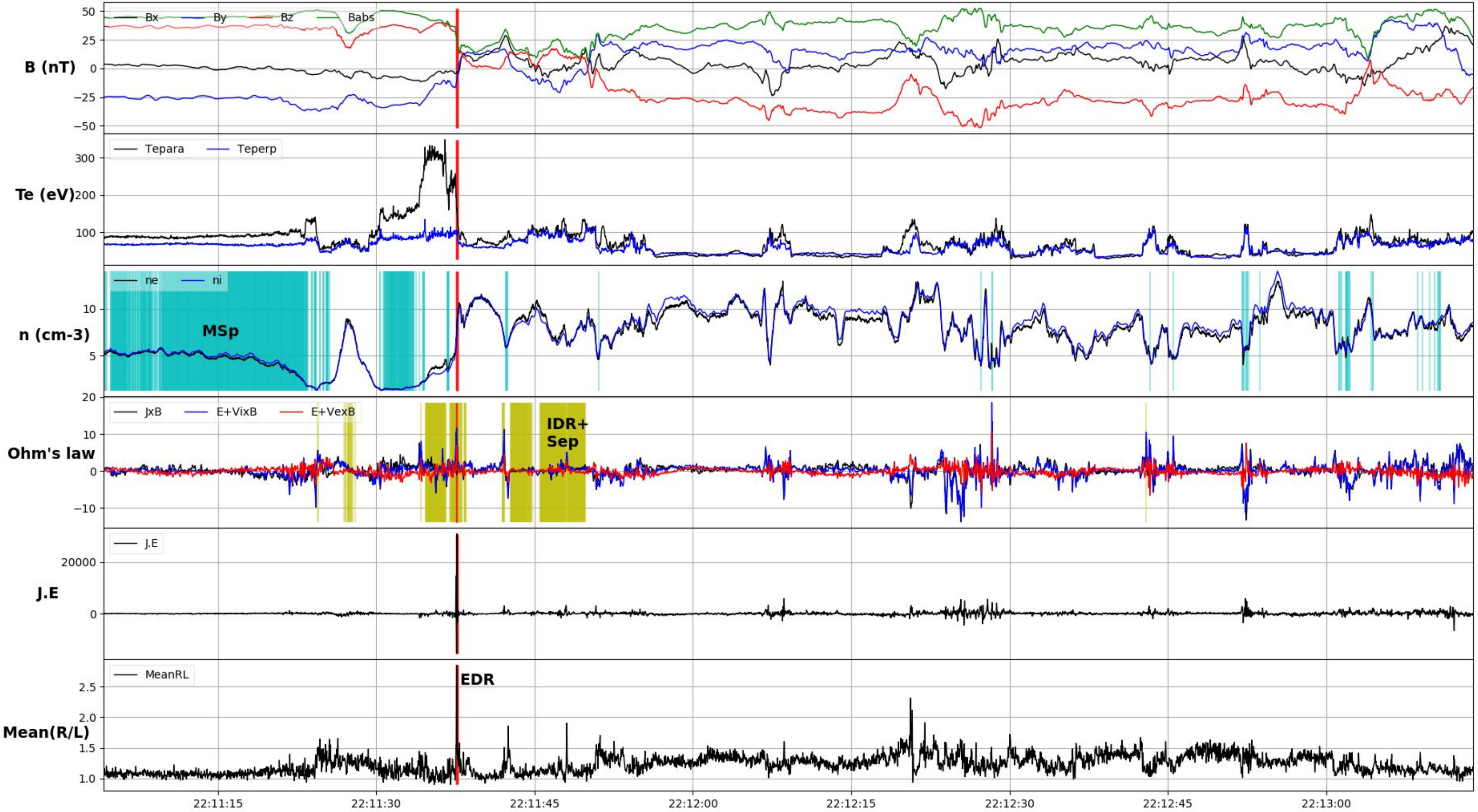
- The algorithm tends to **overestimate the number of EDR data points without missing many of them**
- High performance for the rest of plasma regions (F1-score > 95%) ⇒ Training shows **Promising results**

Flowchart of the whole process

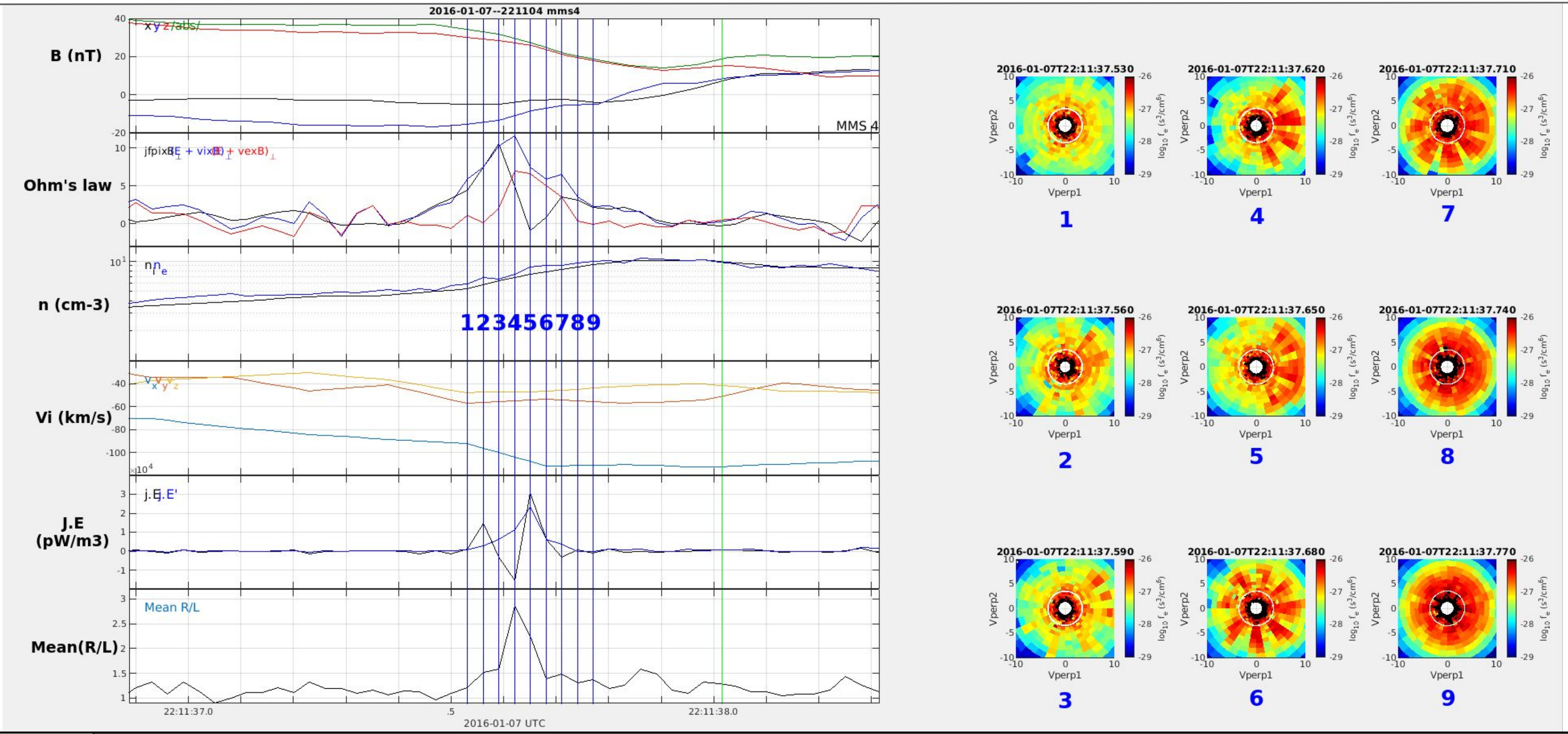
- **<u>Importance of post-processing</u>** : Different configurations of criteria for **different approach** :
  - **Very restrictive parameters** => Reduced list of possible EDR candidates with **potentially a few false positives**
  - **Not too restrictive parameters** => Large list of possible EDR candidates with **potentially a lot of false positives**
  - **Currently going for option number one**



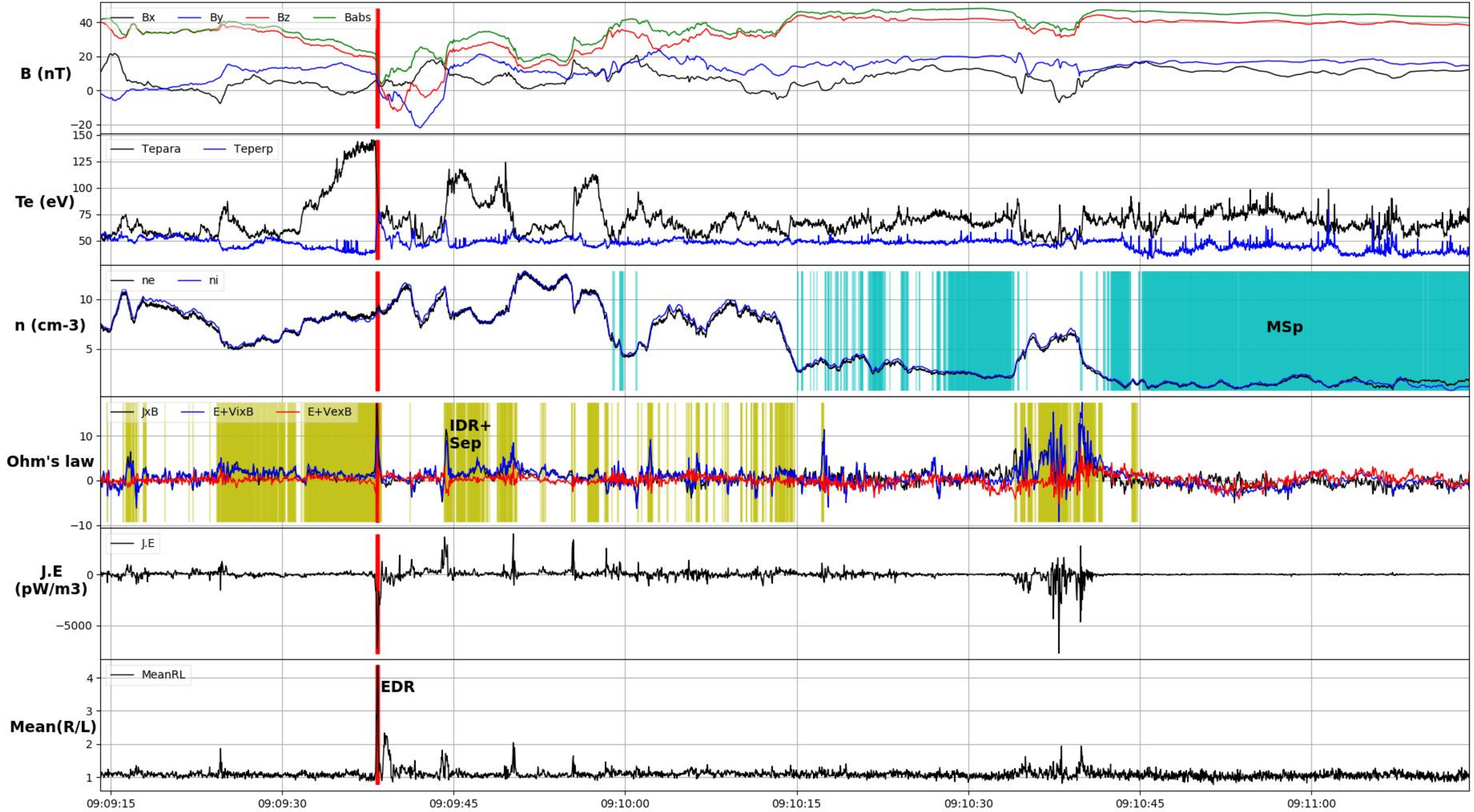
Example of potential new EDR candidate found (1)



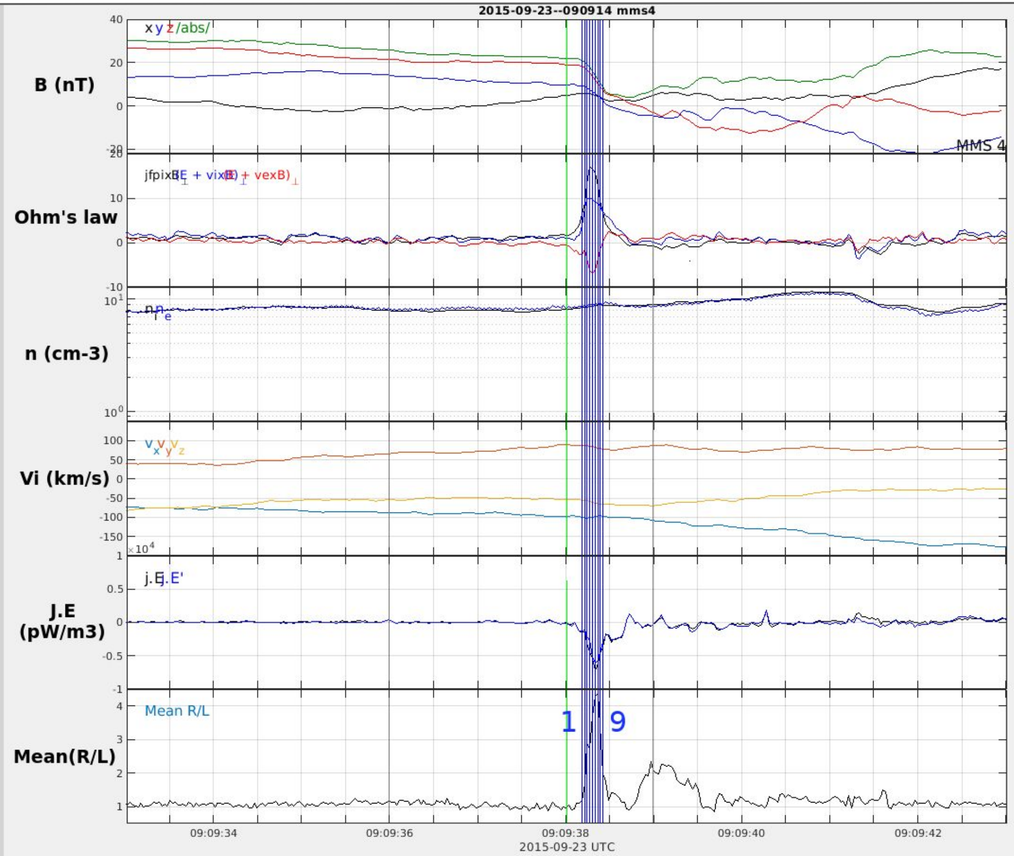
Example of potential new EDR candidate found (2)

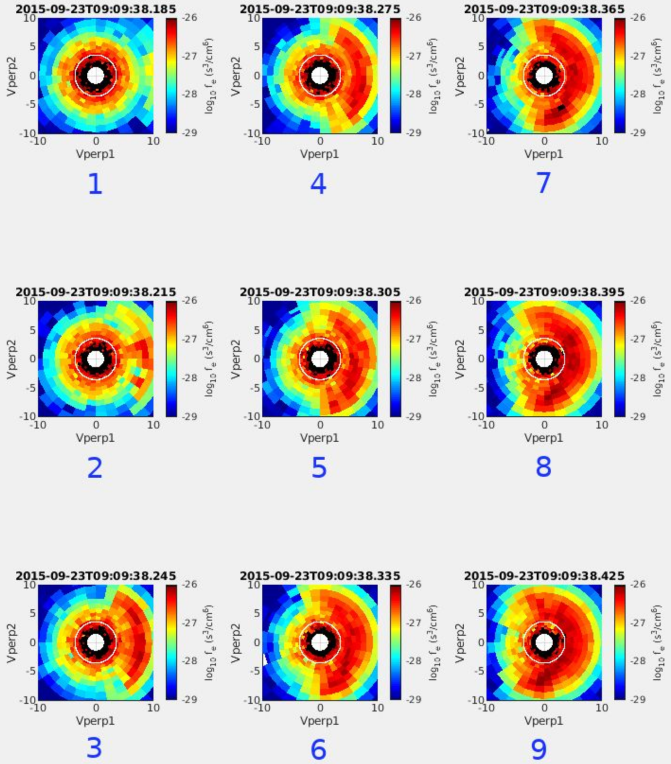


Outer EDR candidates with J.E < 0 (1)



Outer EDR candidates with J.E < 0 (2)





Conclusion

- **Summary** :
  - **Training of a Machine Learning algorithm** on manually labeled data from the phase 1 of MMS, Predictions of the algorithm on **Magnetopause Crossings intervals** from phase 1a **from ISSI team's database**
  - **Use of a special parameter** : *Mean(R/L)* to better identify EDR crescents on time series
  - **Good results** (F1-score > 95%) for the detection of plasma regions during the training, even though the **number of EDRs seem to be overestimated** by the algorithm
  - Possibility to produce either a big list of possible EDR candidats with many false positives or a "clean" reduced one depending on the chosen post-processing parameters, but visual inspection still needed at this stage to check EDR candidates
  - **Paper in preparation** with list of possible EDRs

End

Thank you for reading !