T-FORS 1st Innovation Day

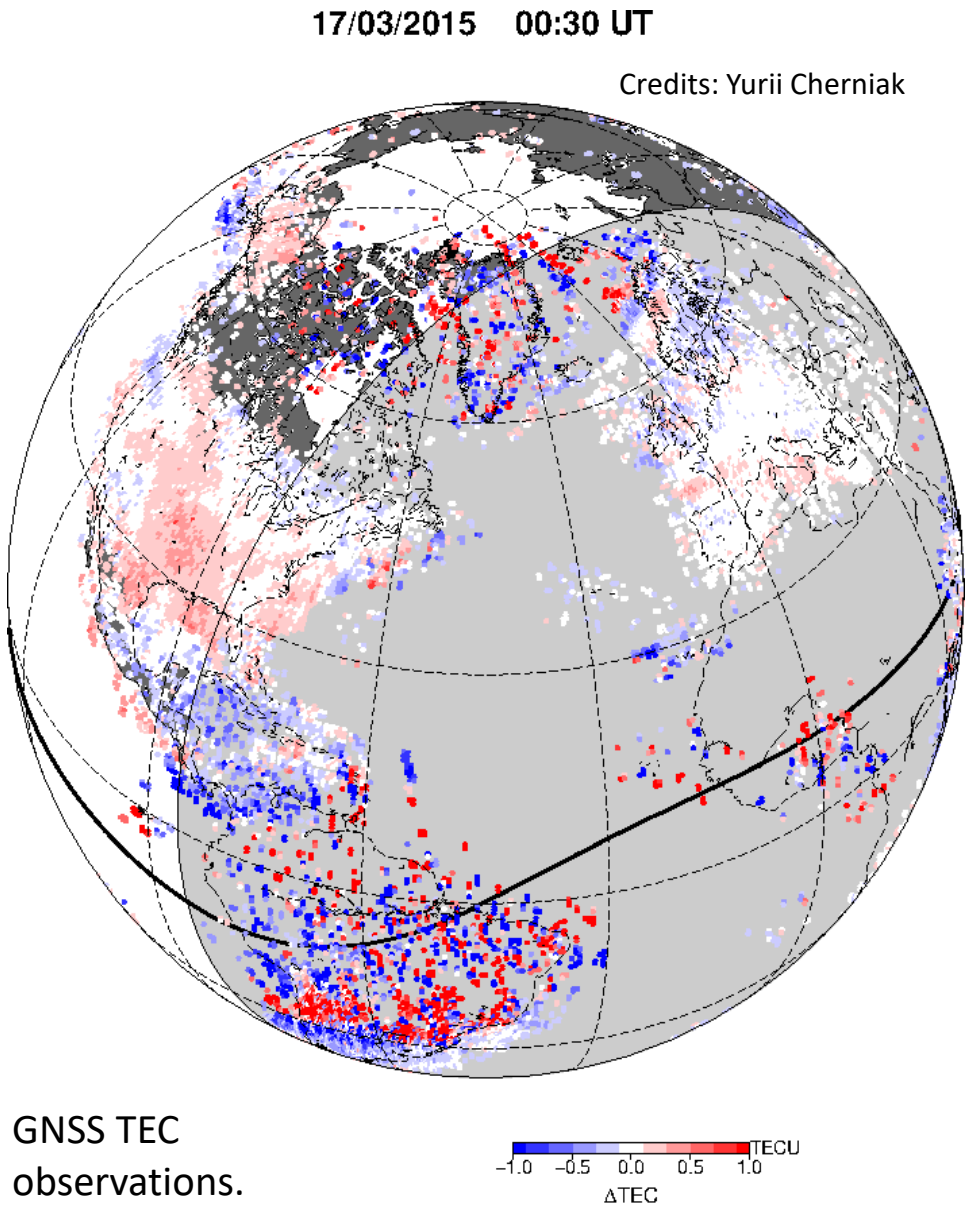
LSTID Forecasting

Luca Spogli

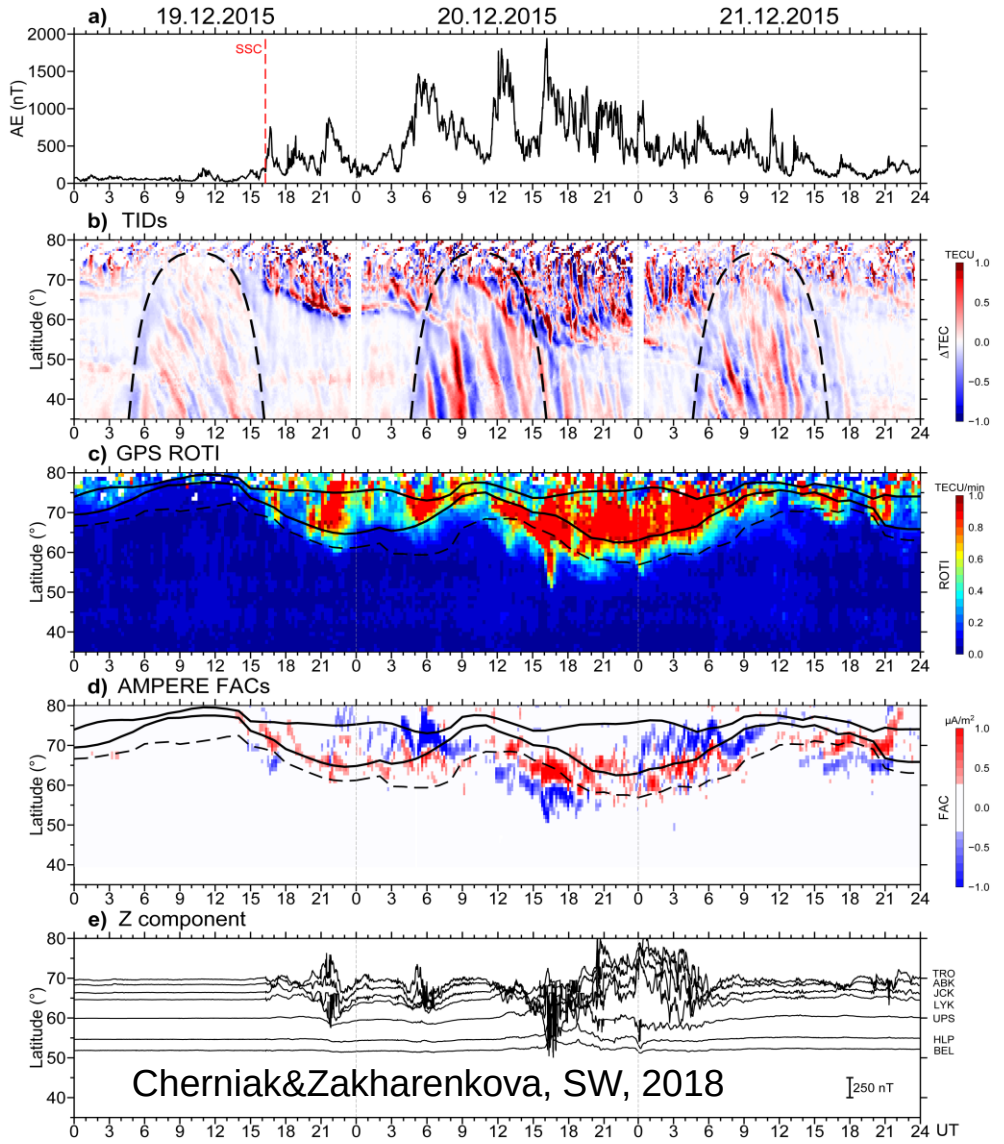
Outline

- LSTIDs occurrence chain of events
- LSTID forecasting based on ML: general strategy
- Developing the ML models
  - Features and labels
  - First approach: catalogue-based forecasting
  - Second approach: LSTID indices-based forecasting
- Remarks and way forward

Large-scale Travelling Ionospheric Disturbances



GNSS TEC observations.



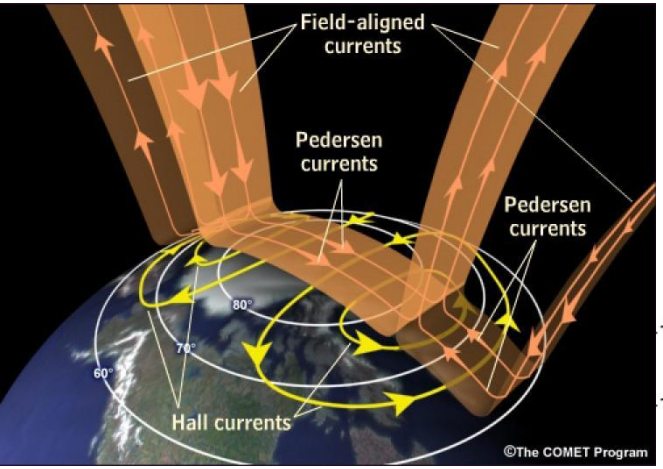
LSTIDs occurrence chain of events from the auroral oval to middle latitudes

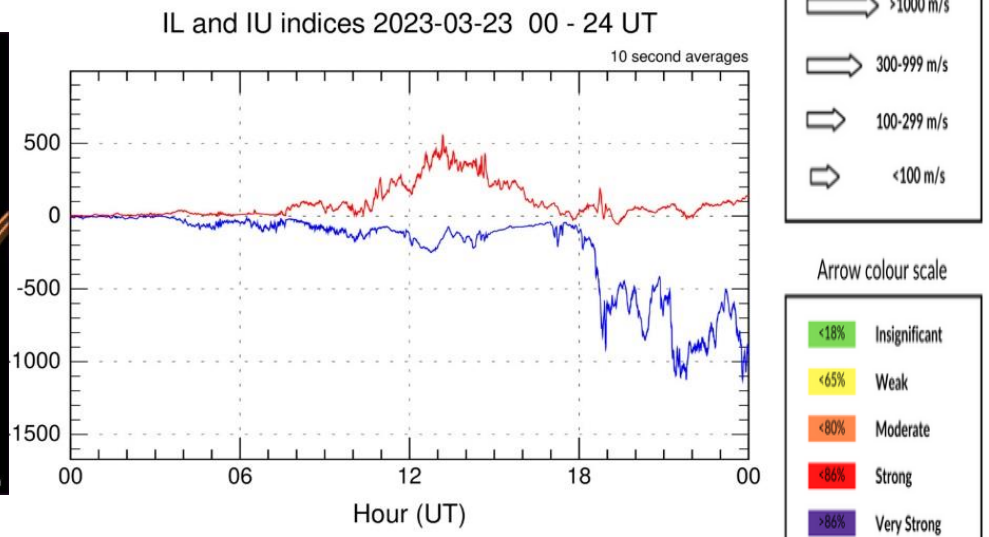
Energy injection at high latitude inducing Joule heating

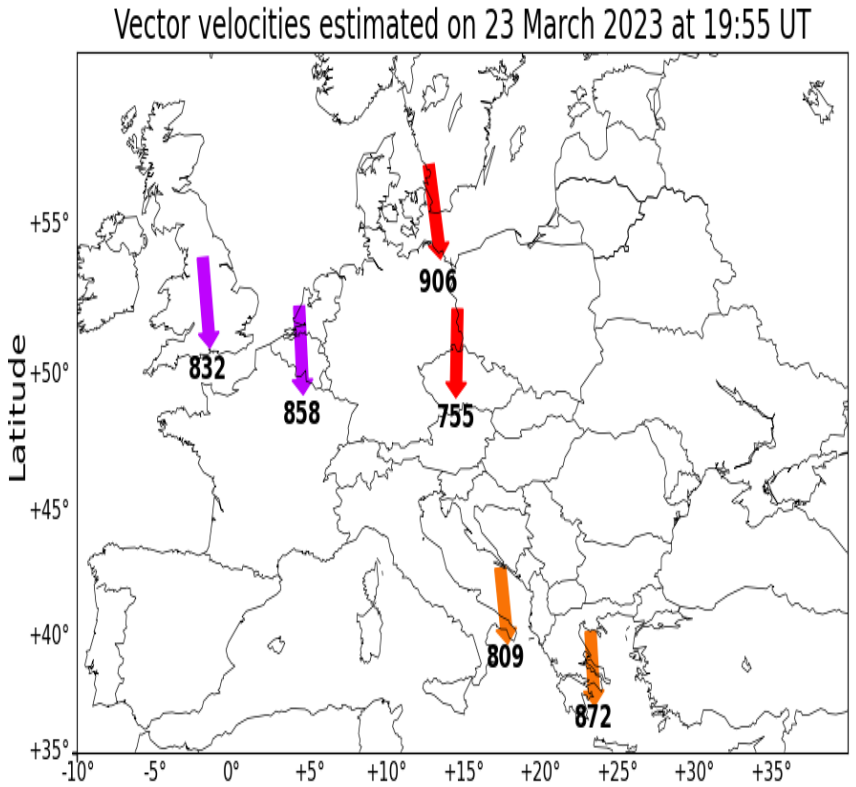
Geomagnetic field disturbances detected along a meridional chain of magnetometer at ground

Equatorward Propagation of LSTIDs from auroral latitudes

Detection of LSTIDs

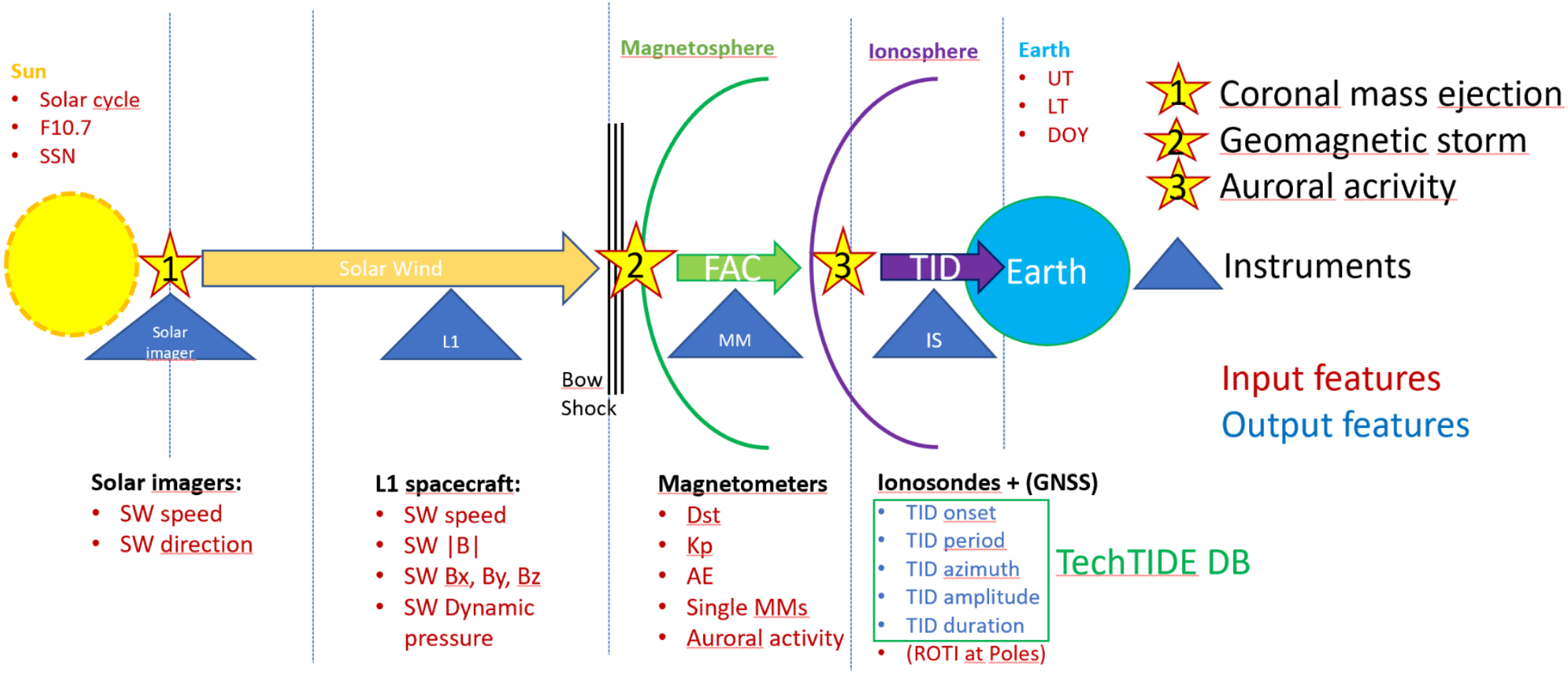




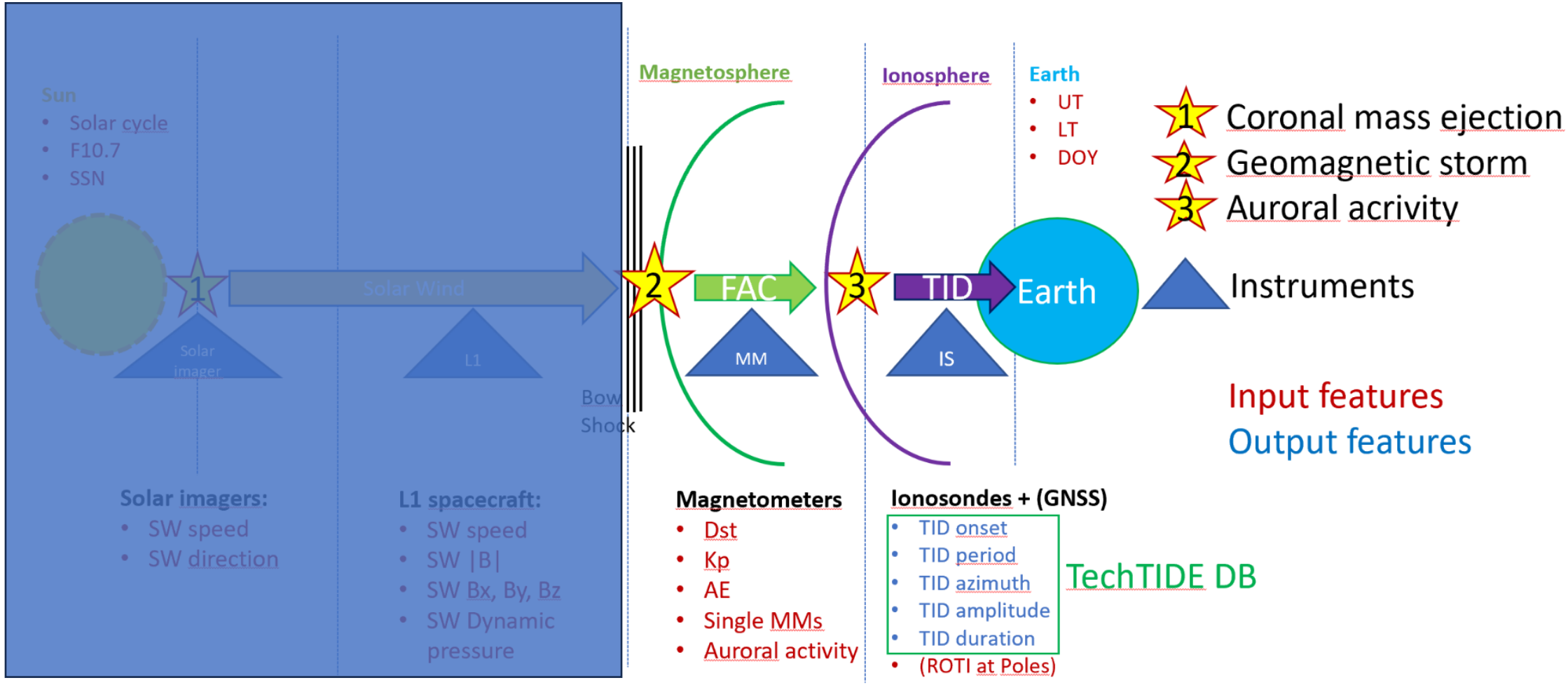


but not only geospace forcing (see remarks)

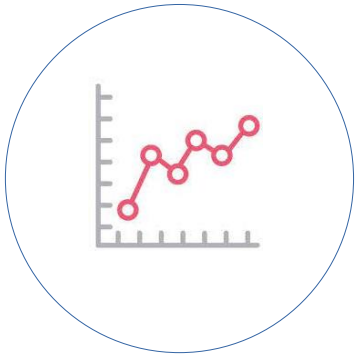
LSTID forecasting based on ML: general strategy

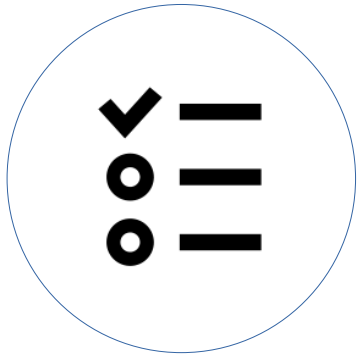


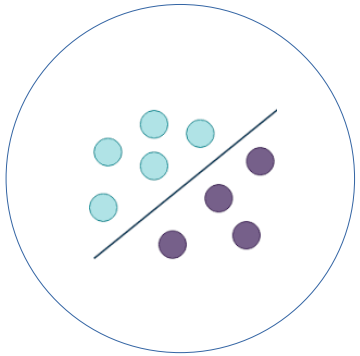
LSTID forecasting based on ML: general strategy

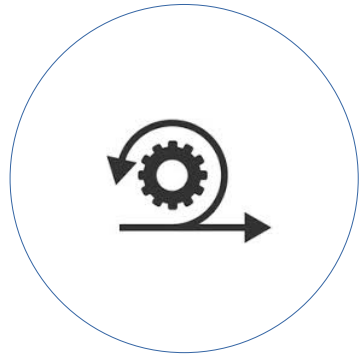


LSTID detection model development: General strategy









1.

Exploratory data analysis

Impute missing values

Sync different sources

Data visualization & inspection

2.

Feature selection

Compute covariances

Data normalization

Split training & validation sets

3.

Build and train Neural Network

Parameter tuning & cross-validation

4.

Deploy model

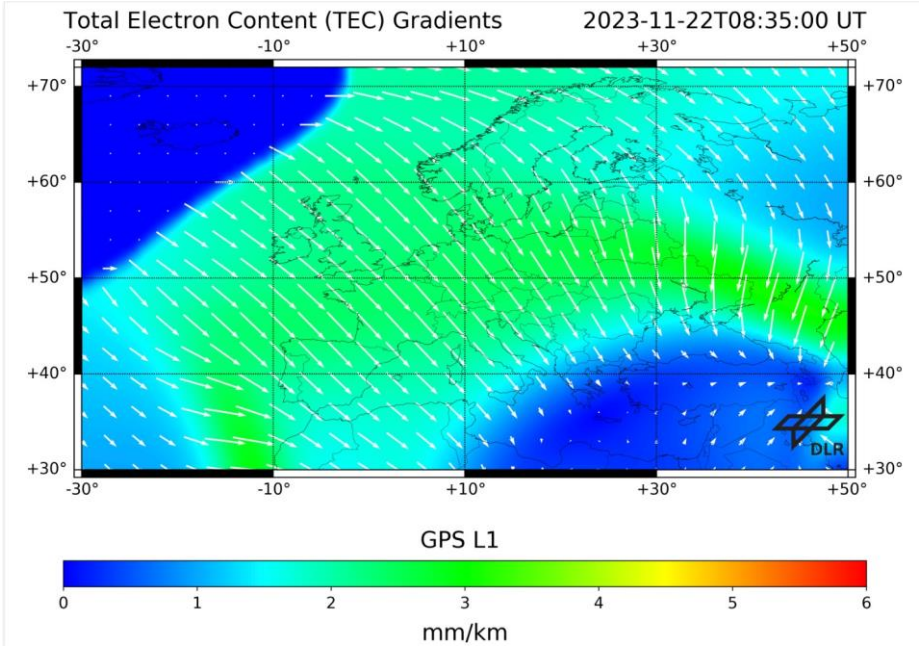
Performance evaluation

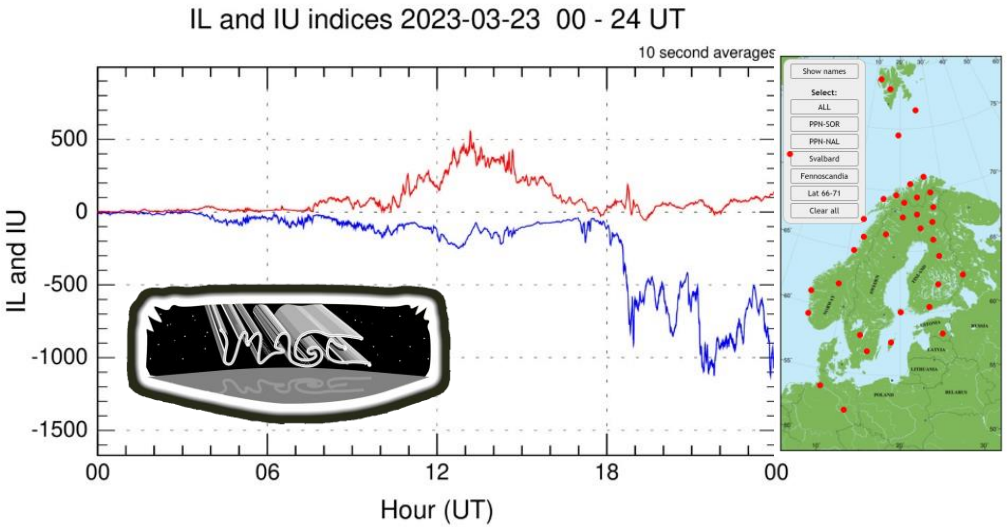
Re-train Neural Network

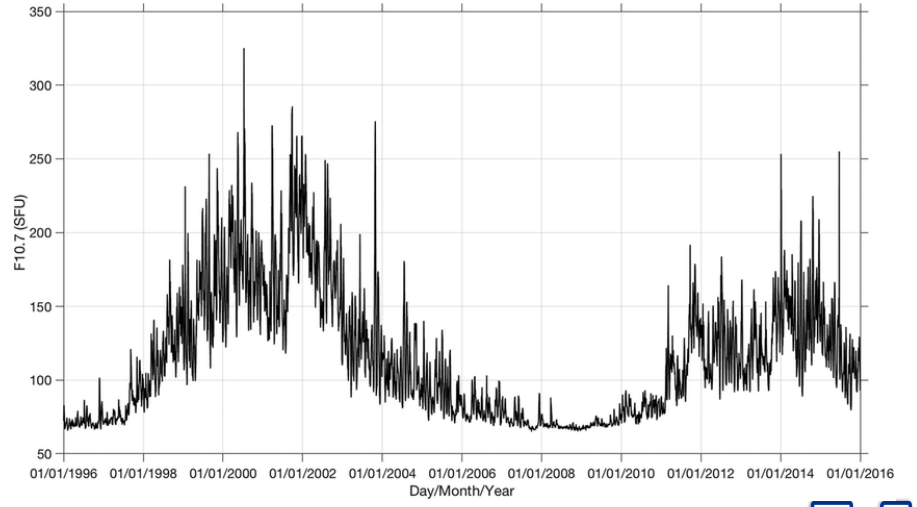
Developing the ML models: features & labels

Features

- IE (IL, IU) values
- GNSS TEC Gradients over Europe
- F10.7 value of the previous day
- current hour (ch) (cos(2pi/24 ch) + 1)/2
- current month (cm) (cos(2pi/12 cm) + 1)/2



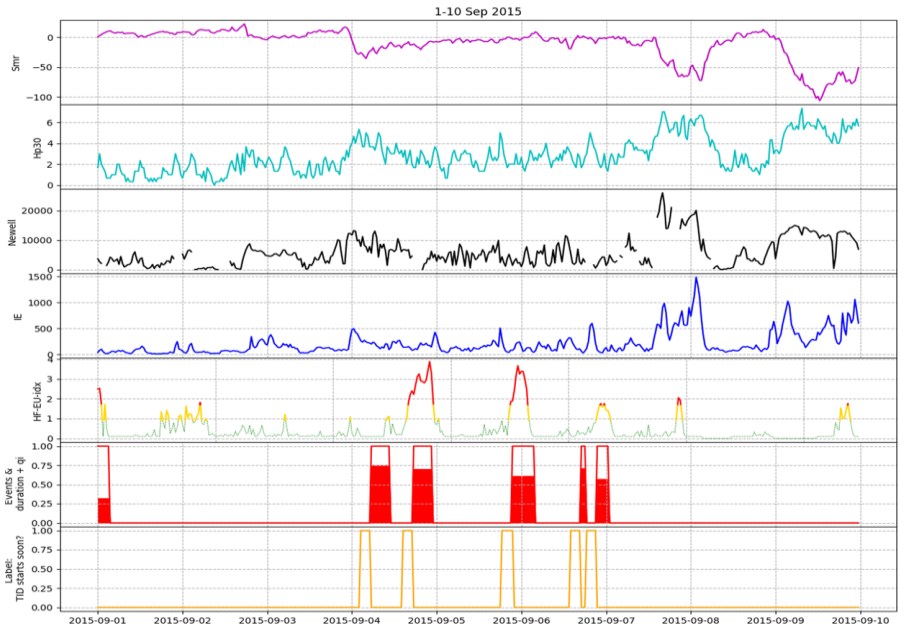




Developing the ML models: features & labels

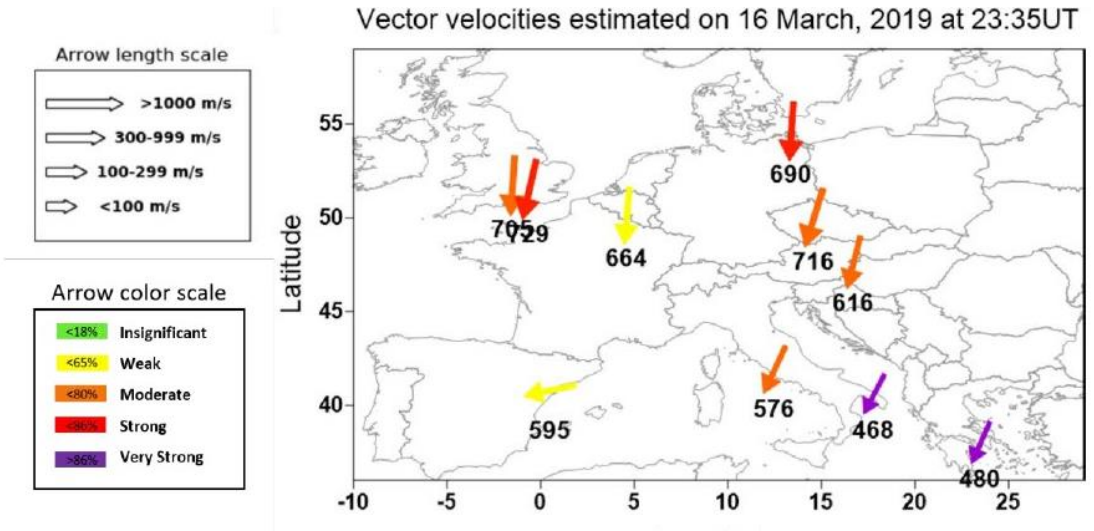
Features

- IE (IL, IU) values
- GNSS TEC Gradients over Europe
- F10.7 value of the previous day
- current hour (ch) (cos(2pi/24 ch) + 1)/2
- current month (cm) (cos(2pi/12 cm) + 1)/2



Labels

- SPCont time series (based on MUF from single ionosondes)
- LSTID catalogue provided by Ebro Observatory based on HF-INT method
- HF-EU index (Activity index at European Level)



Developing the ML models: catalogue-based forecasting

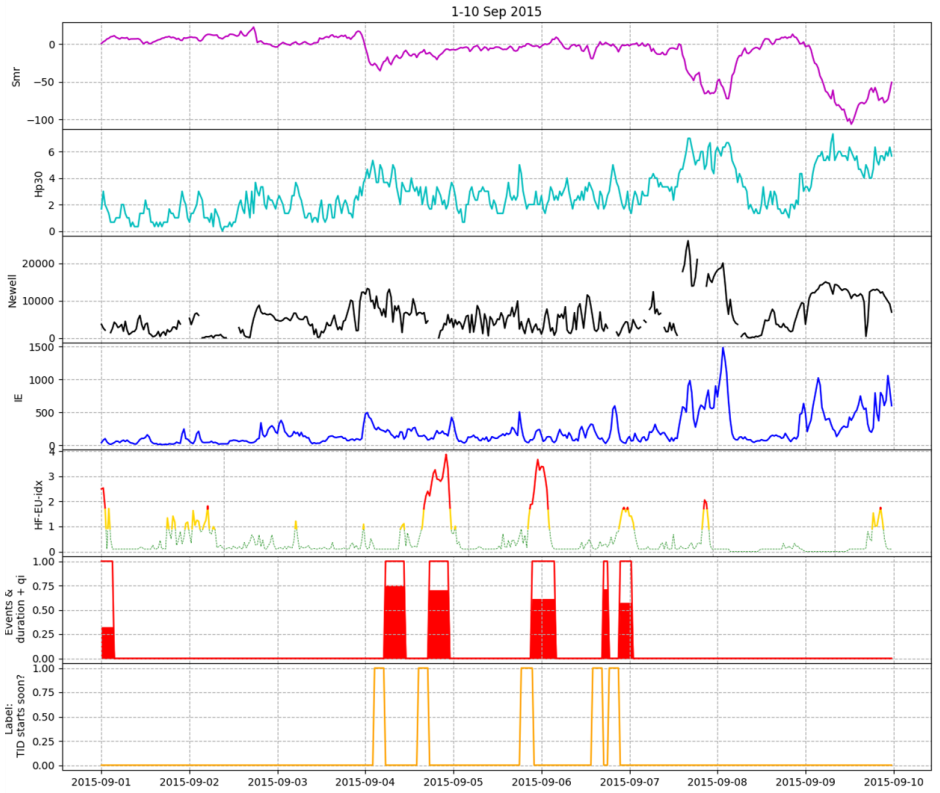
- The problem is handled as a binary classification
- **HF-INT refined LSTID catalogue** provided by Ebro Observatory
- In the catalogue, there are **760 TIDs events** detected and recorded above Europe between **FEB 2014 to DEC 2022**
- The database is generated by leveraging a **network of ionosondes** covering the European sector

Parameter	Example
Start time	2022 01 11 21:00
Duration	2.0 hrs
Period	119.74 min
Amplitude	0.72 MHz
Velocity	597.47 m/s
Azimuth	202.39°
Detection technique	D2D

Catalogue example

Developing the ML models: catalogue-based forecasting

- The dataset is incomplete: (misdetections related to the technique used to create it)
- The class are severely unbalanced: 3% of Yes and 97% of No
- The input is shifted in space and time respect to the output
- Built ML dataset made of 157.777 couples (**X**(T),y(T)) for each T every 30min between FEB 2014 and DEC 2022, where:



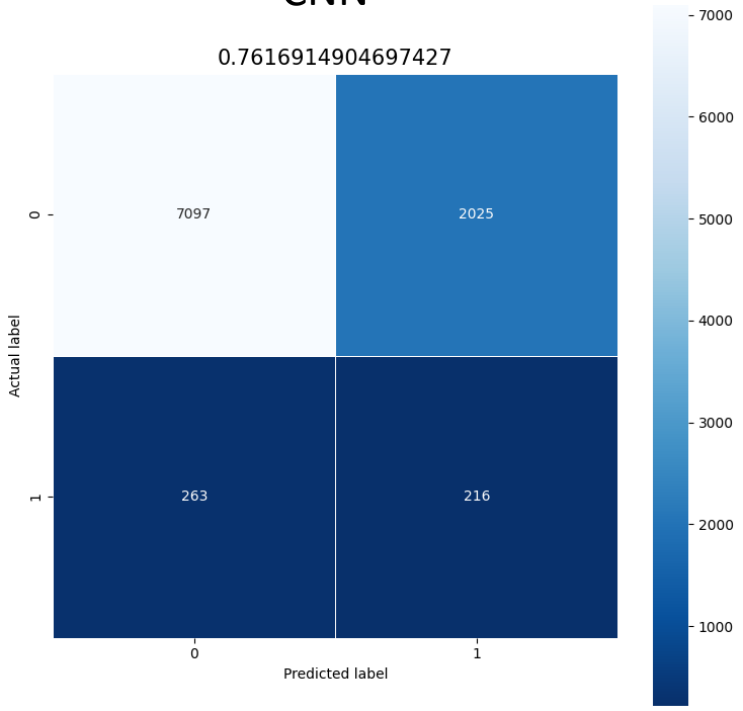
$$\mathbf{X}(T) = \begin{bmatrix} X_1(T-6.5h) & X_1(T-6h) & X_1(T-5.5h) & X_1(T-5h) & \dots & X_1(T-1h) & X_1(T-0.5h) \\ X_2(T-6.5h) & X_2(T-6h) & X_2(T-5.5h) & X_2(T-5h) & \dots & X_2(T-1h) & X_2(T-0.5h) \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ X_N(T-6.5h) & X_N(T-6h) & X_N(T-5.5h) & X_N(T-5h) & \dots & X_N(T-1h) & X_N(T-0.5h) \end{bmatrix}$$

$$y(T) = \begin{cases} 1 & \text{TID detected in 3h starting from T,} \\ 0 & \text{else.} \end{cases}$$

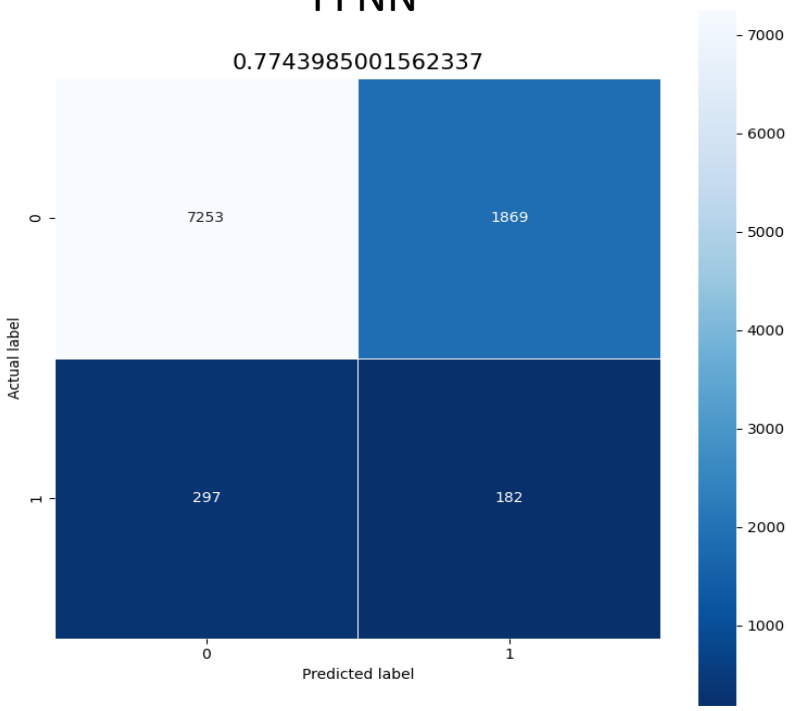
Developing the ML models: catalogue-based forecasting

- We run different configurations of the models with different hyperparameters
- Best performances are obtained using neural networks
- Results are not satisfactory: we are not able yet to classify correctly the two classes based on external drivers
- This suggests no clear correlation between the classes, given the features used

CNN



FFNN



Developing the ML models: LSTID indices-based forecasting

- The **LSTID forecasting** is treated as a **binary classification** problem
- If $\boldsymbol{SPcont} > \boldsymbol{Treshold}$ then an **LSTID** is **detected** (otherwise no LSTID is detected).
- $IL, IU, GNSS, LT$ and *SPCont* values are considered as features
- Results (so far) only for Juliusruh ionosonde

3 Different models:

- **Prediction** of LSTID based **exclusively** on the **most recent** $SPcont$ values.
- **Prediction** of LSTID based **exclusively** on the **most recent** $IL, IU, GNSS, LT$ values.
- **Prediction** of LSTID based on **both** (a) the **most recent** $SPcont$ values and (b) the **most recent** $IL, IU, GNSS, LT$ values.

Classifiers employed:

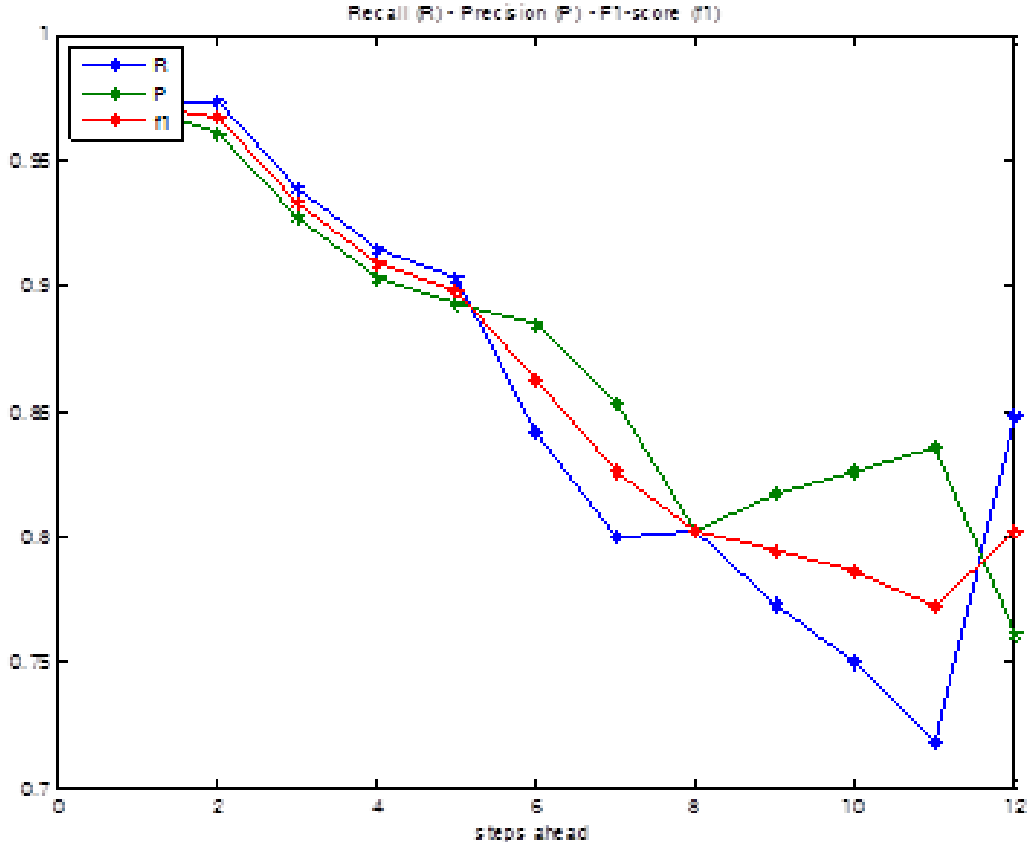
- Feedforward Neural Network classifier – **FNN**
- Block Recurrent Neural Network classifier – **RNN**

Developing the ML models: LSTID indices-based forecasting

Model 1: Prediction of LSTID based **exclusively** on the **most recent** $SPcont$ values.

Remarks:

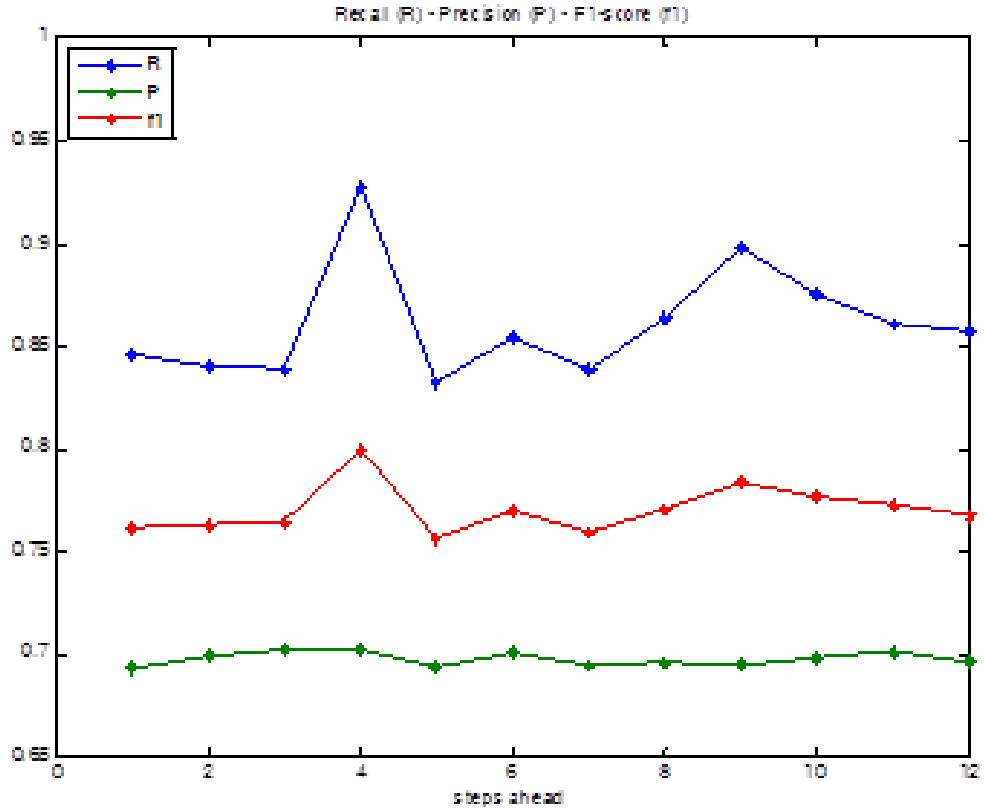
- Trained **FNN** features very high accuracy results, even for higher time horizon forecasts.
- In principle, as the time horizon increases, the **performance** degrades.



<u>Model 2:</u> Prediction of LSTID based **exclusively** on the **most recent** $IL, IU, GNSS, LT$ values.

Remarks:

- The results of the **FNN** classifier are less accurate than in the case where **only** the $\boldsymbol{SPcont}$ past values are considered (Model 2).

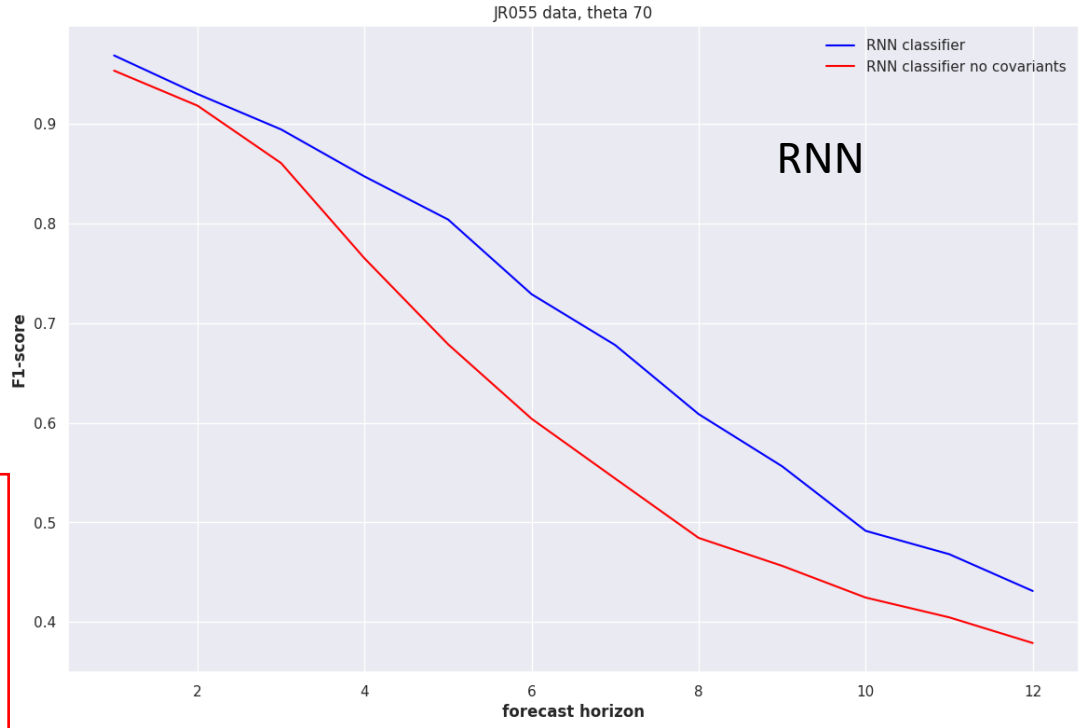


Developing the ML models: LSTID indices-based forecasting

Model 3: Prediction of LSTID based on **both** (a) the **most recent** $SPcont$ values and (b) the **most recent** $IL, IU, GNSS, LT$ values.

Remarks:

- The results of **Model 3** are more accurate than in the case where **only IL, IU, GNSS, LT** past values are exclusively considered as inputs (Model 2).
- The results of **Model 3** are less accurate than in the case where **only the $SPcont$** past values are considered (**Model 1**) but it can handle missing values



Stacked model: Exploit the three models to always provide best results

- If **all** the most recent $SPcont$ values are available, **Model 1** should be used.
- If **none** of the most recent $SPcont$ values are available, **Model 2** should be used.
- If **some** of the most recent $SPcont$ values are available, **Model 3** should be used.

Remarks and way forward

Catalogue-based forecasting:

- Add new input features
- Investigate other time-delays\input time window to be used for the catalogue-based model

Indices-based forecasting:

- Application of the model to other stations (only Juliusruh was considered).
- Utilization of larger data sets (longer time periods).
- Intensive study of the data (e.g., the time periods where LSTIDs are encountered).
- Dealing with the missing data issue (e.g., the cases where the $\boldsymbol{SPcont}$ computation fails).
- Performing classification at a specific station utilizing data from other stations that are at higher latitudes and/or develop a model leveraging the TID activity index dataset (HF-INT)

Thanks for your attention

LSTID Forecasting

Luca Spogli