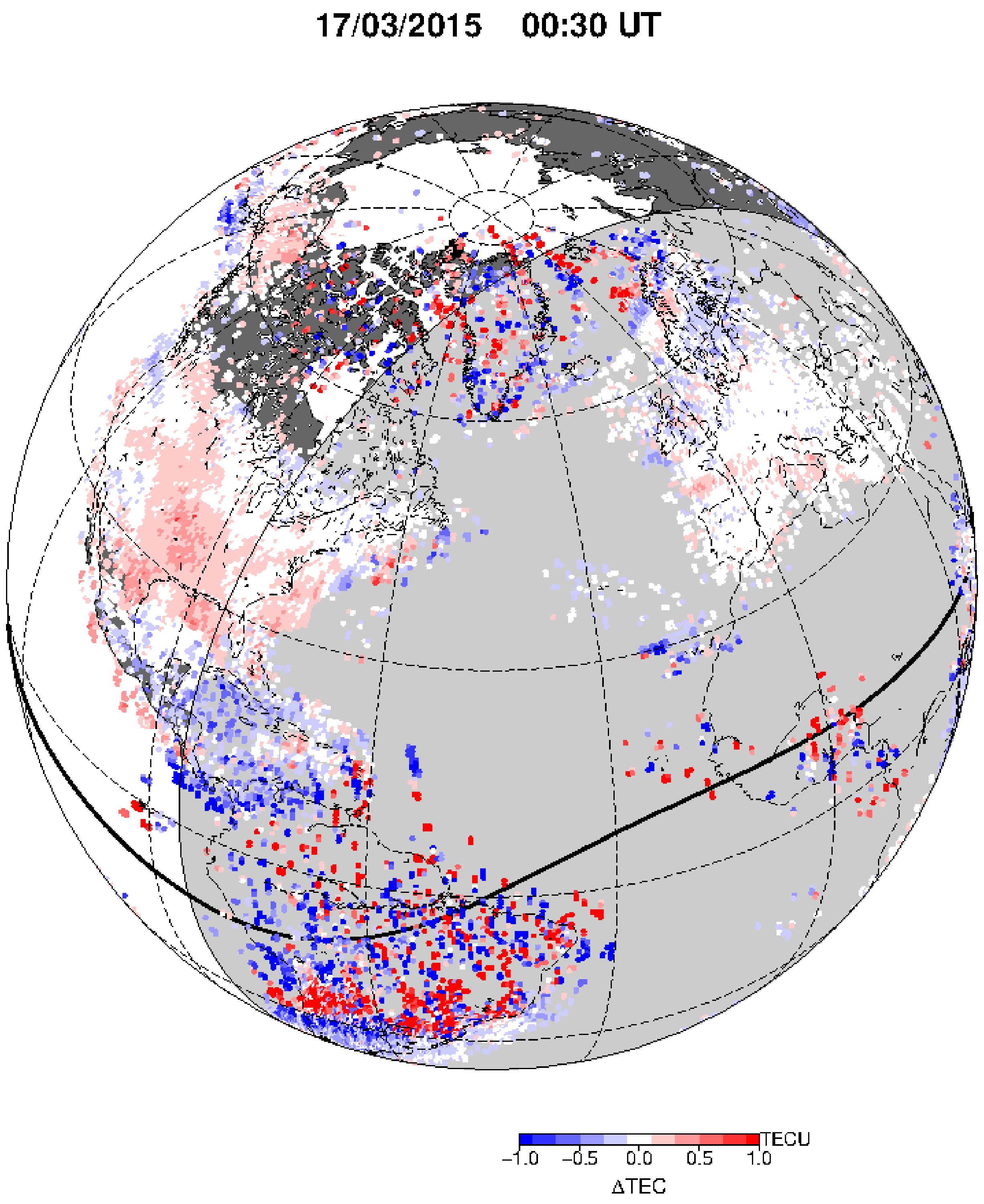
LSTIDs ML forecasting models

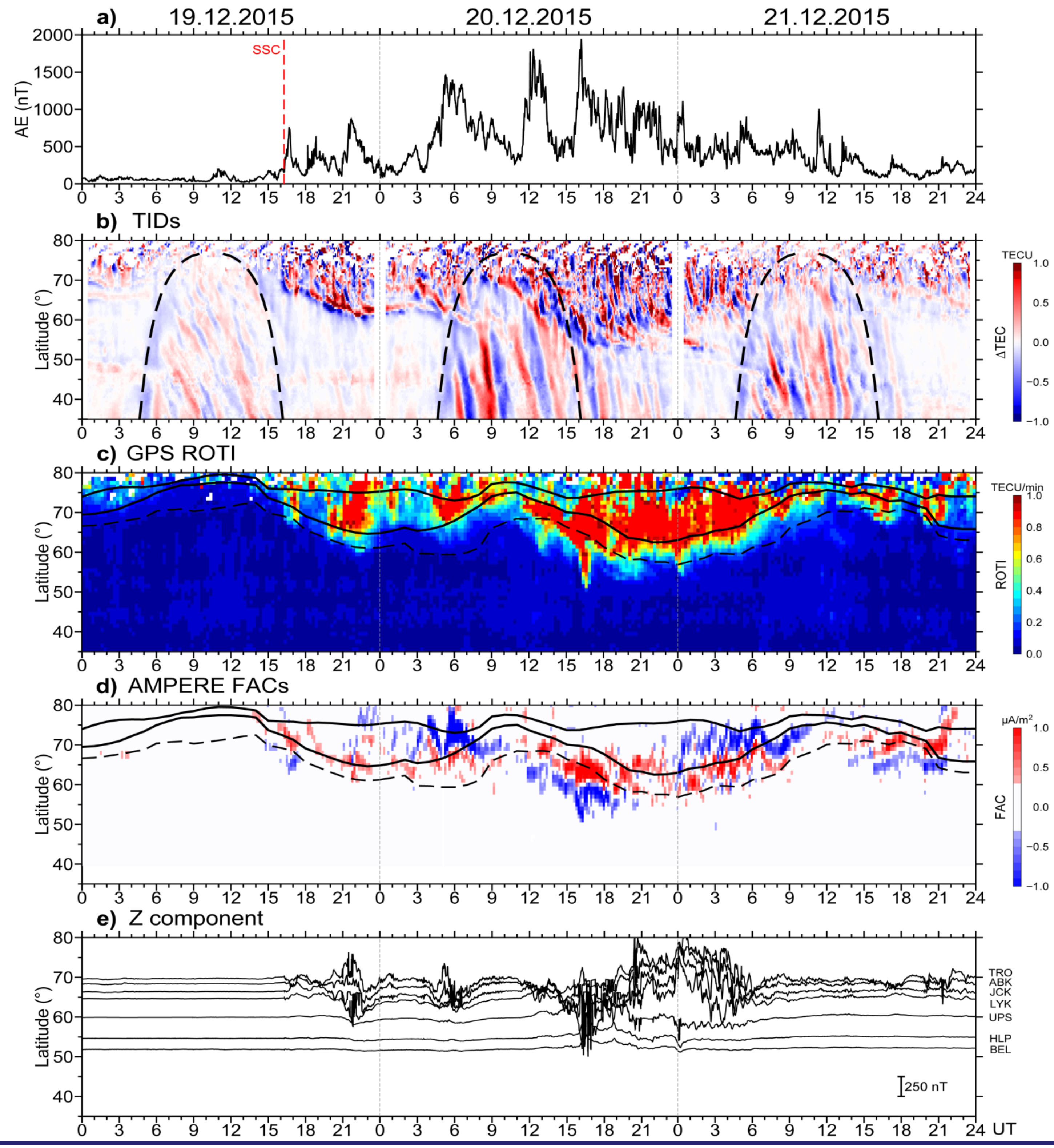
Claudio Cesaroni[(1)] on behalf of T-FORS WP2 participants

Luca Spogli, Marco Guerra, Vincenzo Ventriglia, **Istituto Nazionale di Geofisica e Vulcanologia (INGV)**
Kostas Themelis, Konstantinos Koutroumbas, Anna Belehaki, **National Observatory of Athens (NOA)**
David Altadill, Toni Segarra, **Ebro Observatory (EO)**
Tobias Verlhust**, Royal Meteorological Institute of Belgium (RMI)**
Ivan Galkin**, Borealis Global Designs Ltd. (BGD)**
Veronika Barta, **Institute of Earth Physics and Space Science (FI)**

LSTID in a nutshell

Large-scale Travelling Ionospheric Disturbances

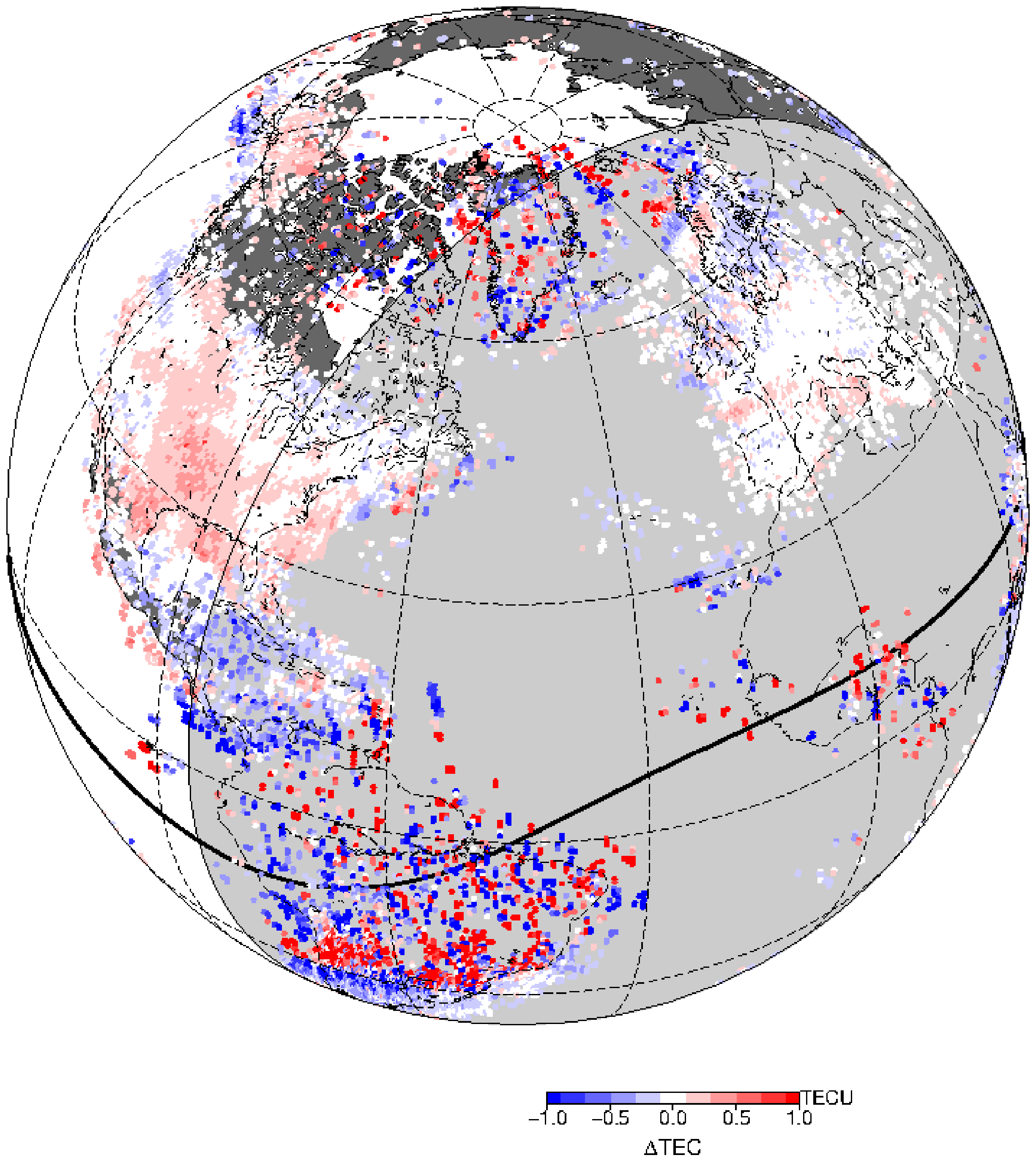




LSTID in a nutshell

Large-scale Travelling Ionospheric Disturbances

17/03/2015 00:30 UT



TIDs constitute a threat for operational systems using predictable ionospheric characteristics as they can impose disturbances with:

- amplitudes of up to ~20% of the ambient electron density

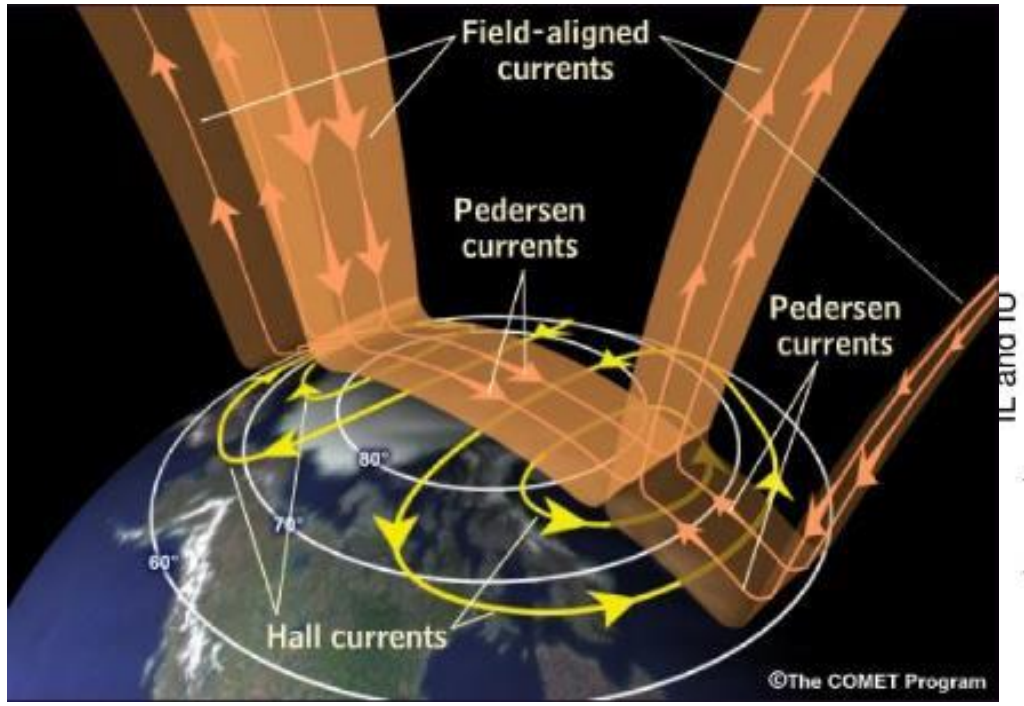
- a Doppler frequency shift of the order of 0.5 Hz on HF signals

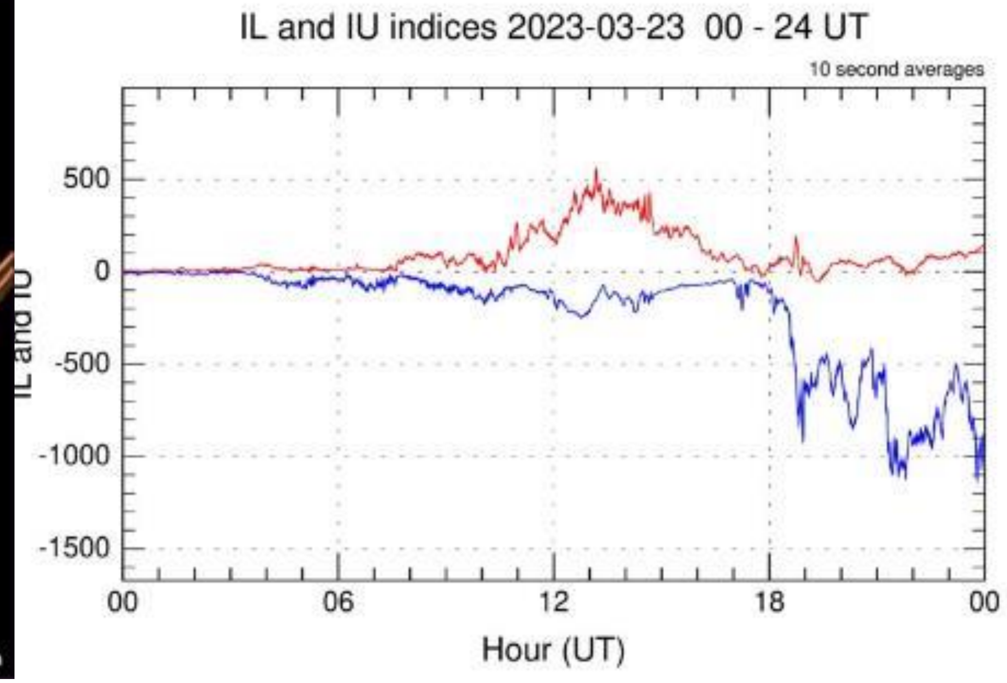
- perturbations in the Total Electron Content (TEC) from less than 1 TEC unit (TECU) up to 10 TECU.

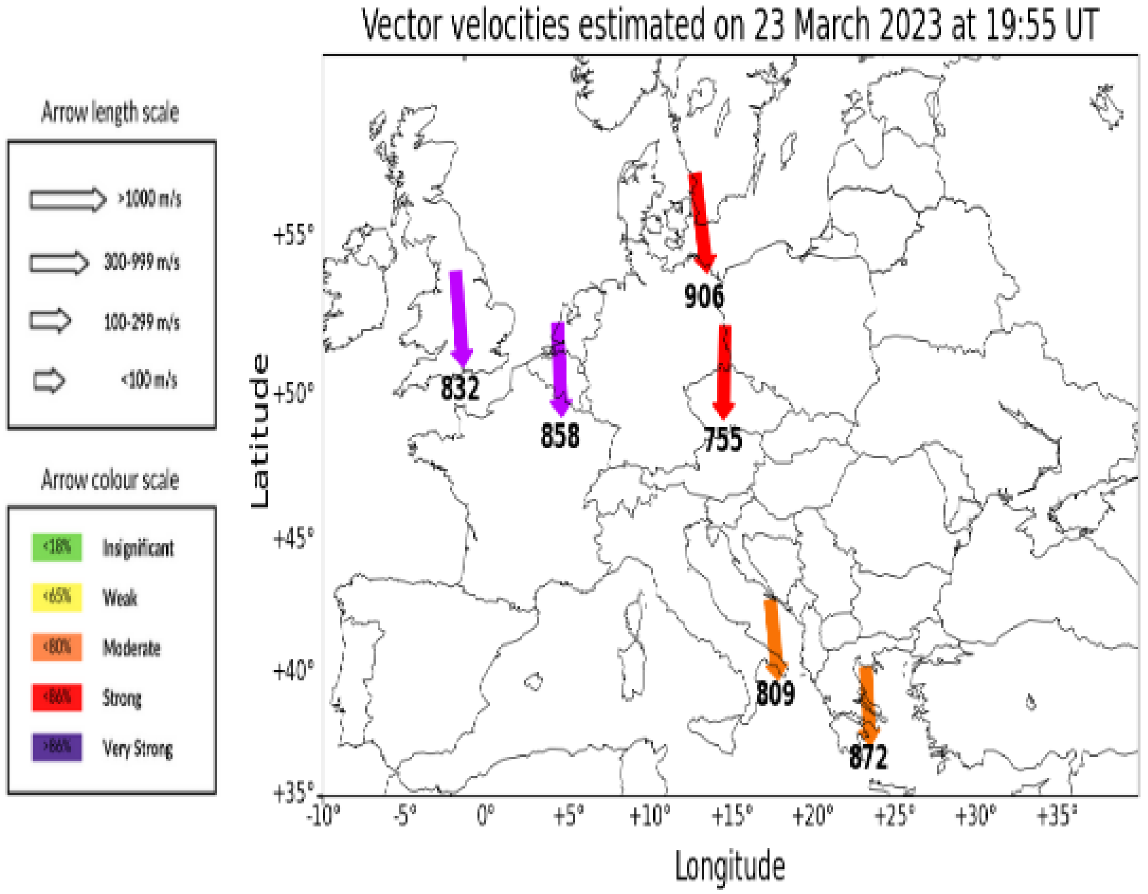
LSTID in a nutshell

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

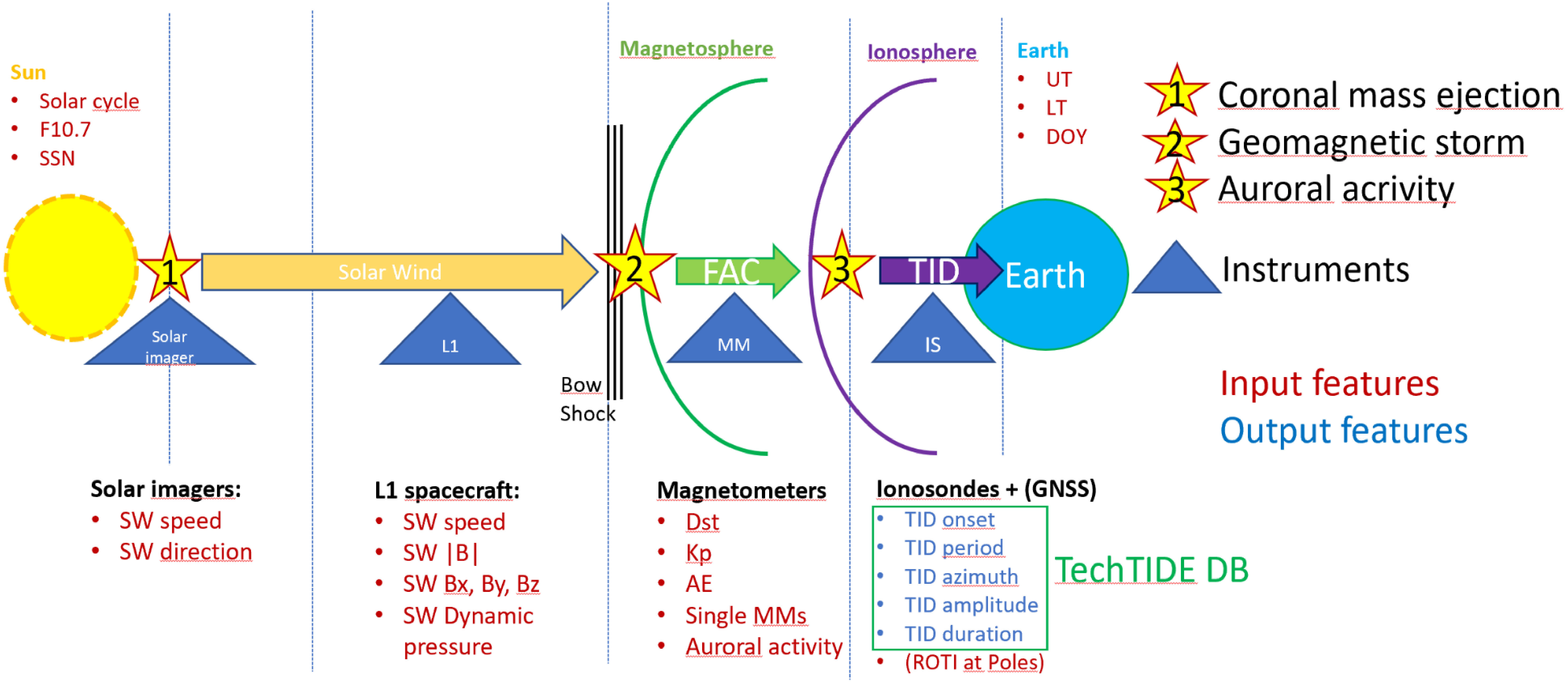






LSTIDs forecasting

LSTID forecasting: general strategy



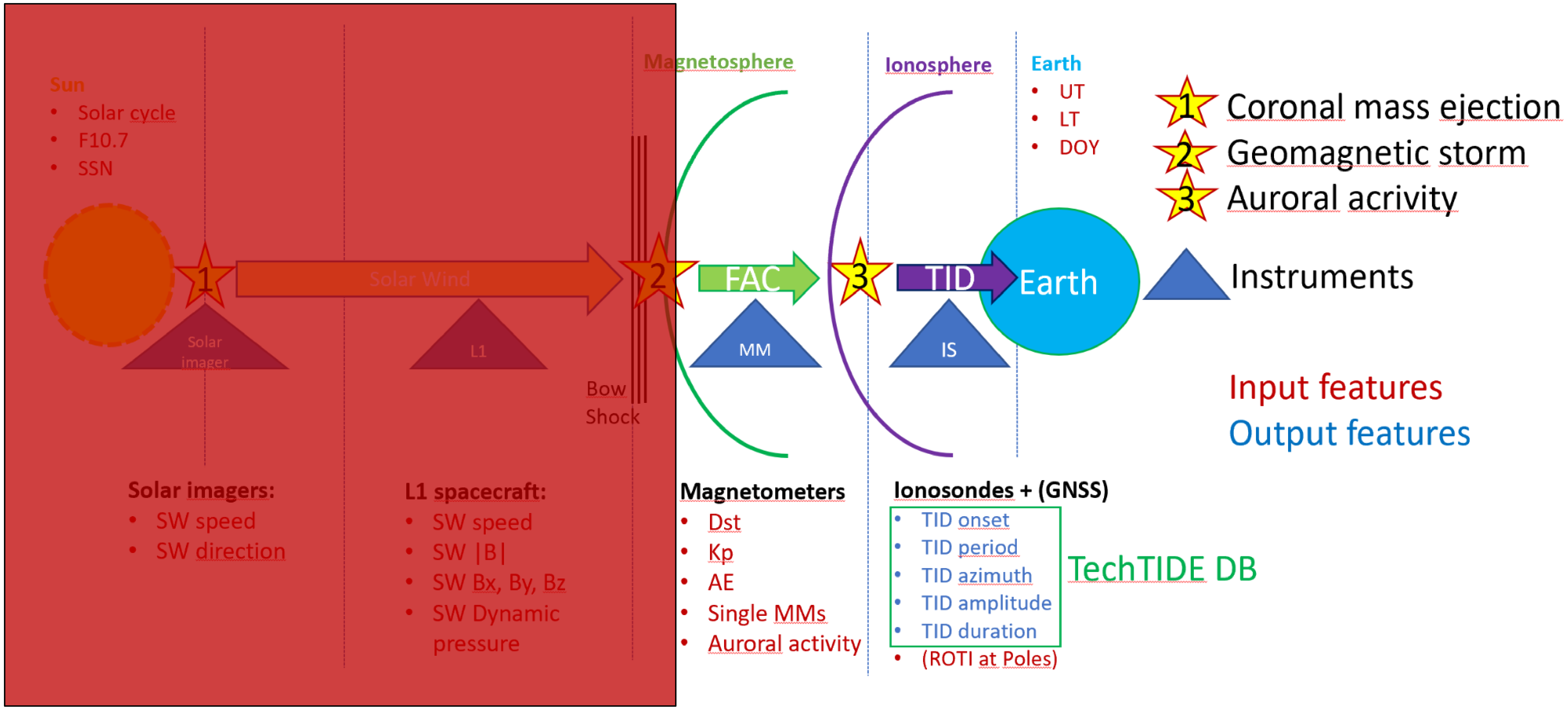
LSTIDs forecasting

LSTID forecasting: general strategy

- From solar imagers data, one can obtain the solar wind speed and following the estimated time of arrival at the Earth together with the expected magnetic field vector. This is the first possible driver of LSTIDs. Highest time horizon, smallest reliability/accuracy.
- From L1 spacecraft, we can calculate the values forecasted by solar imagers-based models. Medium time horizon, medium reliability/accuracy.
- From magnetometers, we can understand the current state of the Magnetosphere-Ionosphere system. Smallest time horizon, highest reliability/acuracy.

LSTIDs forecasting

LSTID short-term forecasting: general strategy

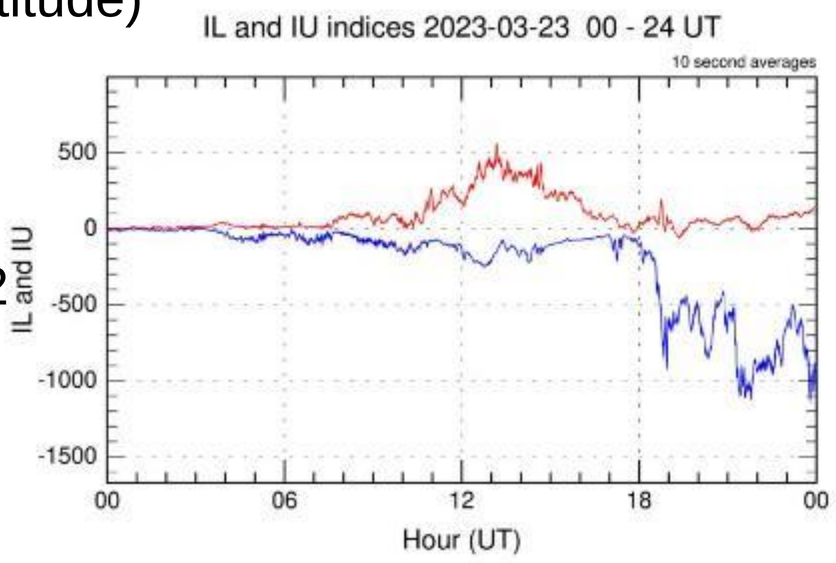


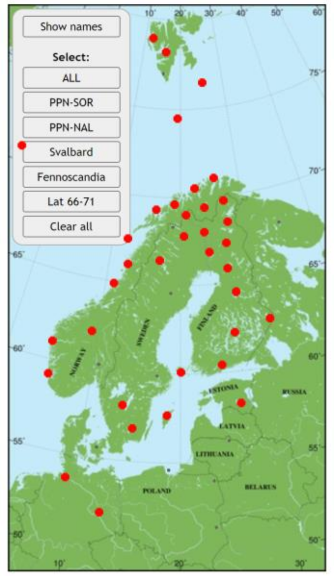
LSTID ST forecasting model

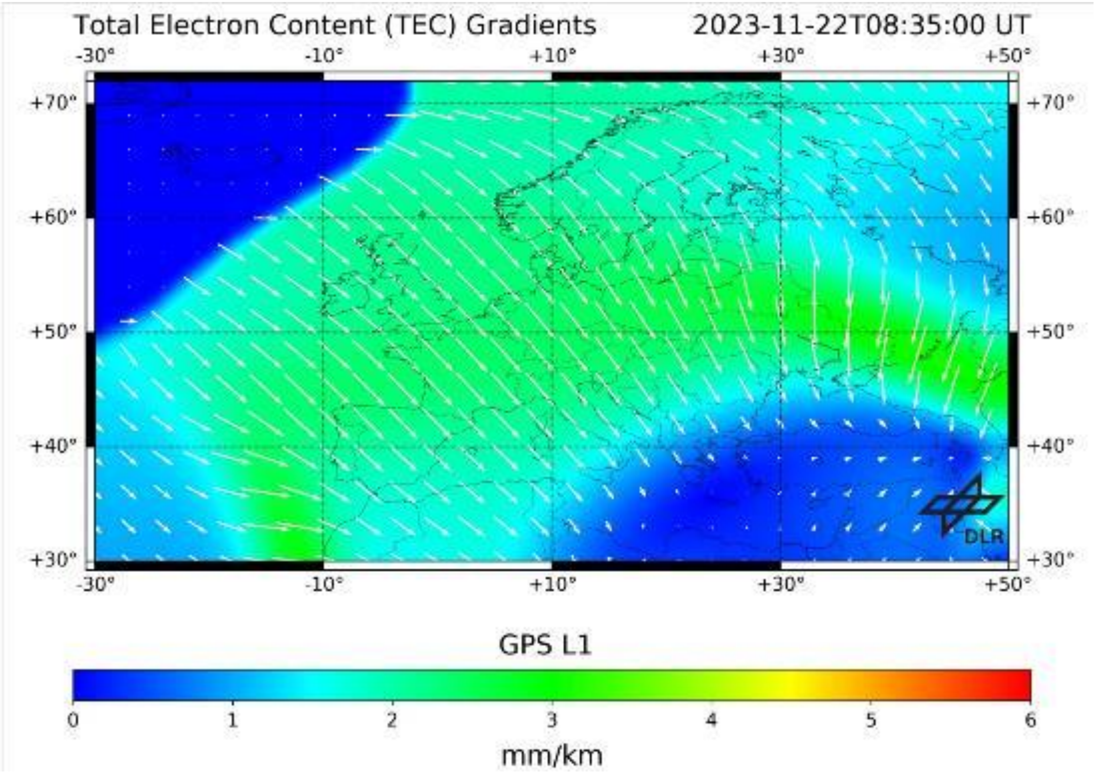
Developing the ML models: features & labels

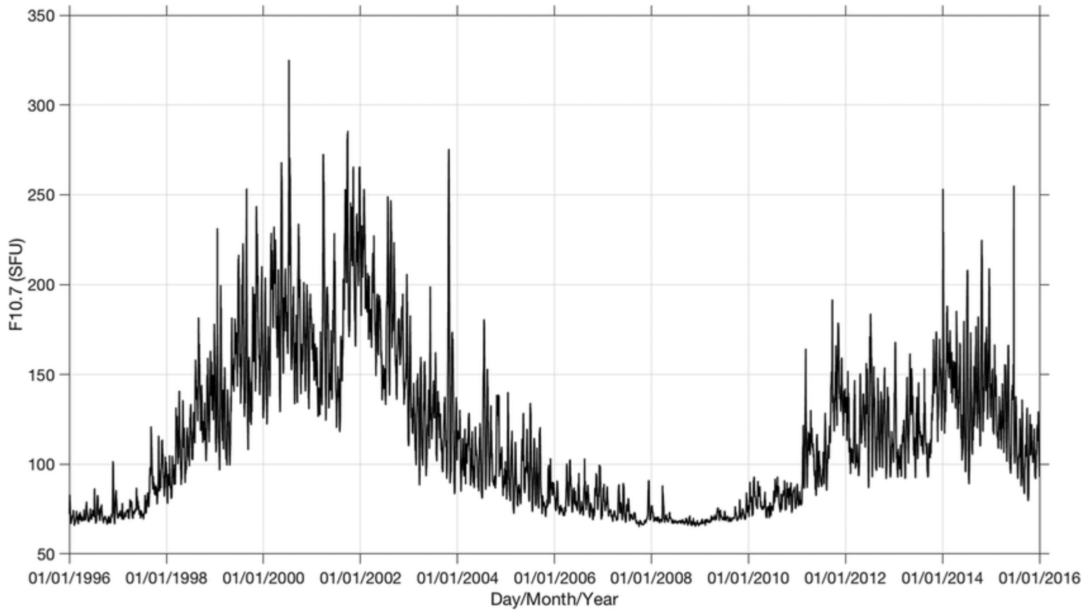
Features

- IE (IL, IU) values
- GNSS TEC Gradients over Europe (high latitude)
- 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
- SPCONT (Spectral Contribution
  from single ionosonde, FFT on 6-h)









Labels

- HF-INT refined LSTID catalogue provided by Ebro Observatory
- SPCONT (Spectral Contribution, FFT on 6-h)
- HF-EU index (Activity index based on HF interferometry)

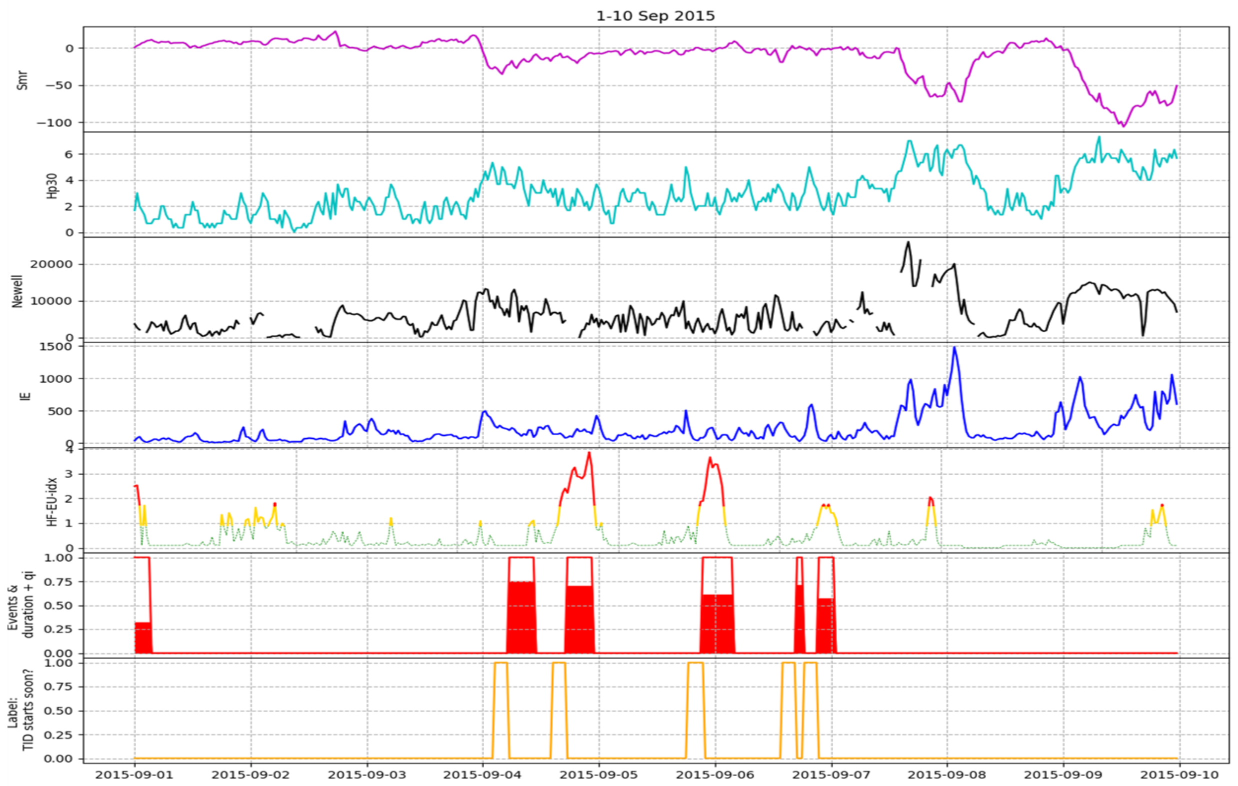
LSTID ST forecasting model

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
- SPCONT (Spectral Contribution

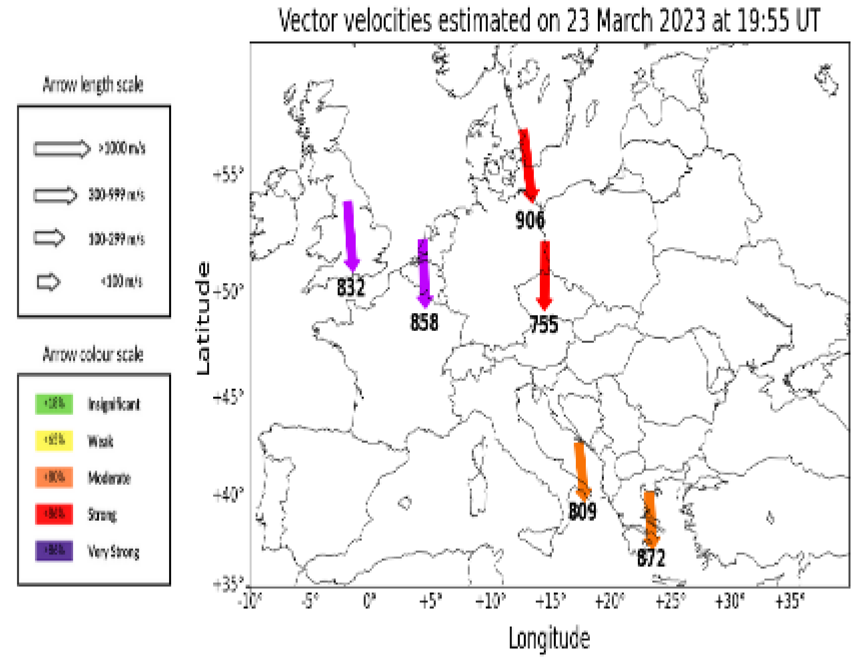
  from single ionosonde, FFT on 6-h)



Labels

- HF-INT refined LSTID catalogue provided by Ebro Observatory
  - SPCONT (Spectral Contribution

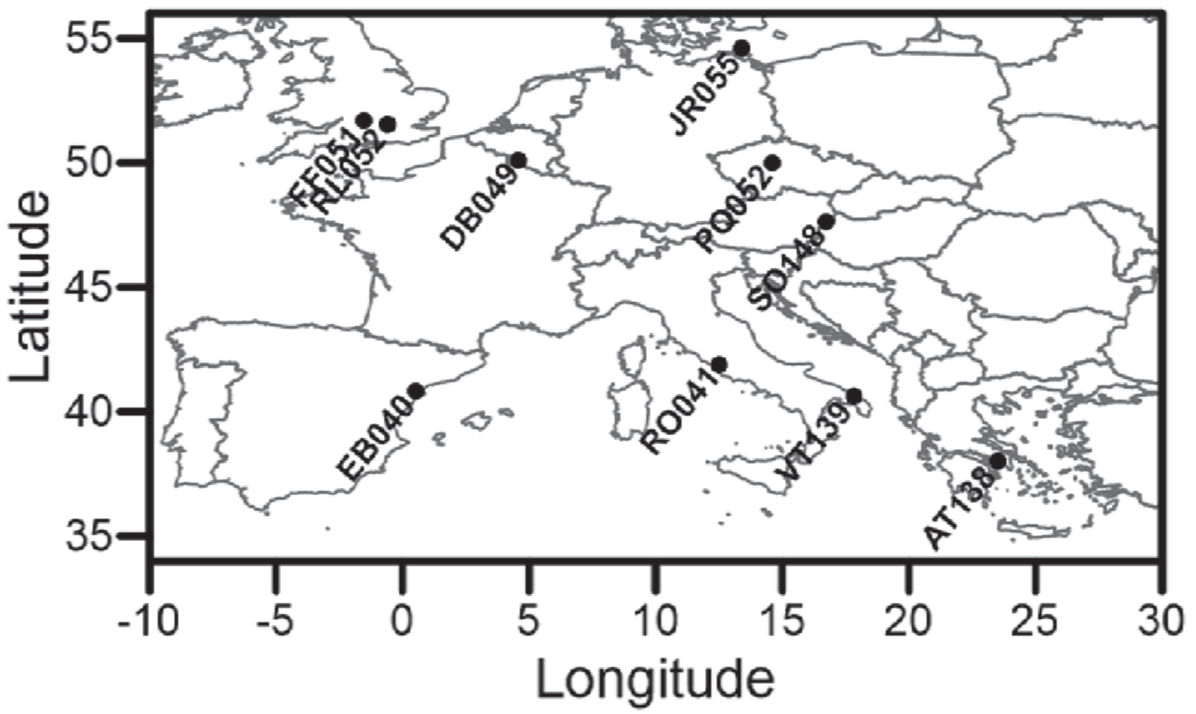
    from single ionosonde, FFT on 6-h)
- HF-EU index (European Activity index based on HF interferometry)



Detecting LSTIDs

The Detection method: HF-Interferometry

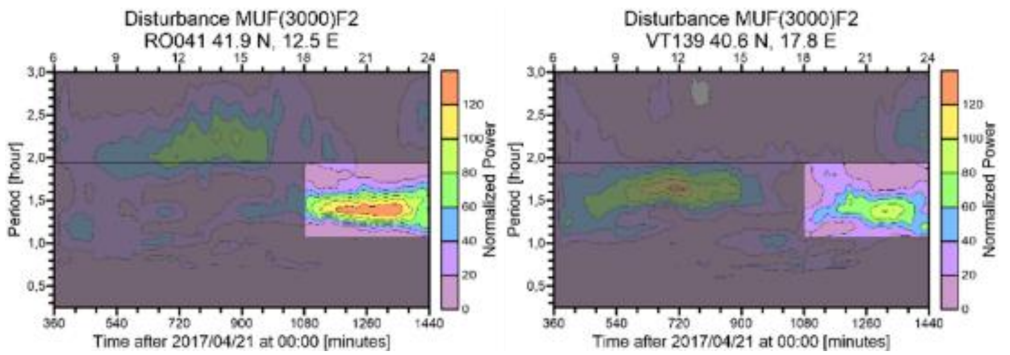
INPUT



- Characteristics from VI Ionospheric sounding (**MUF(3000)F2**).
- Network of DPS4D with stations working **synchronized**.
- GIRO DIDBase Fast Chars database http://giro.uml.edu/didbase/scaled.php

- **Detection of TID-like variation**

  Detect coherent TID-like variations by spectral analysis.



- **TIDs contribution to data variability.**

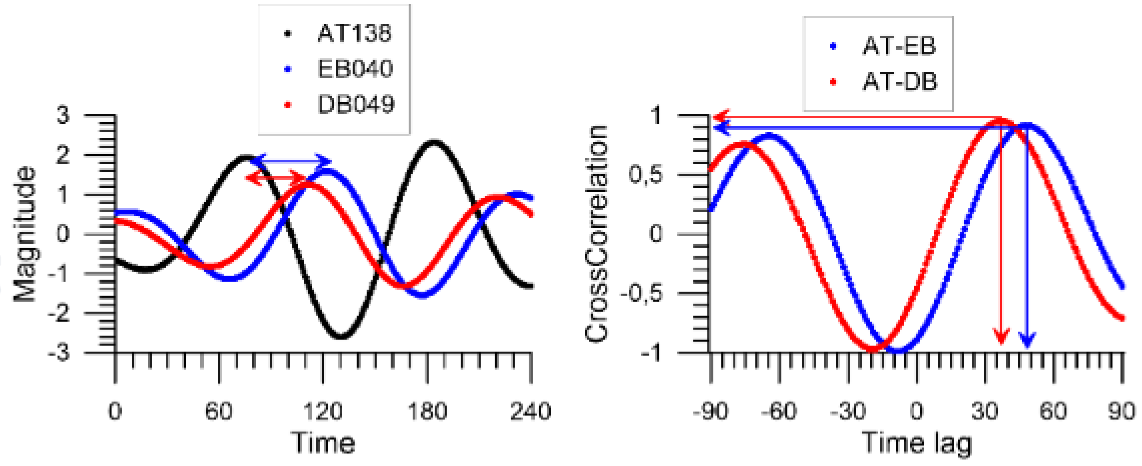
  Application of the Parseval's relation [A(ω) vs A(T)]

$$\sum_{n=-\infty}^{\infty}|x[n]|^2=\frac{1}{2\pi}\int_{-\pi}^{\pi}|X(\omega)|^2 d\omega \sim \sum_{T=T_S}^{T=T_E}A(\omega)^2 \qquad \mathrm{SPCont}(\%)=\frac{\sum_{T=T_{TID_S}}^{T=T_{TID_E}}A(T)^2}{\sum_{T=T_S}^{T=T_E}A(T)^2}$$

- **Estimation of the velocity and azimuth of the TID**

  Estimate time delays for different sites by cross-correlation, ΔTM_i. Estimate velocity of disturbance $\vec{v}$ assuming planar propagation.

$$\Delta TM_i - \vec{s}\cdot\Delta\vec{r}_i = 0 \;;$$



Detecting LSTIDs

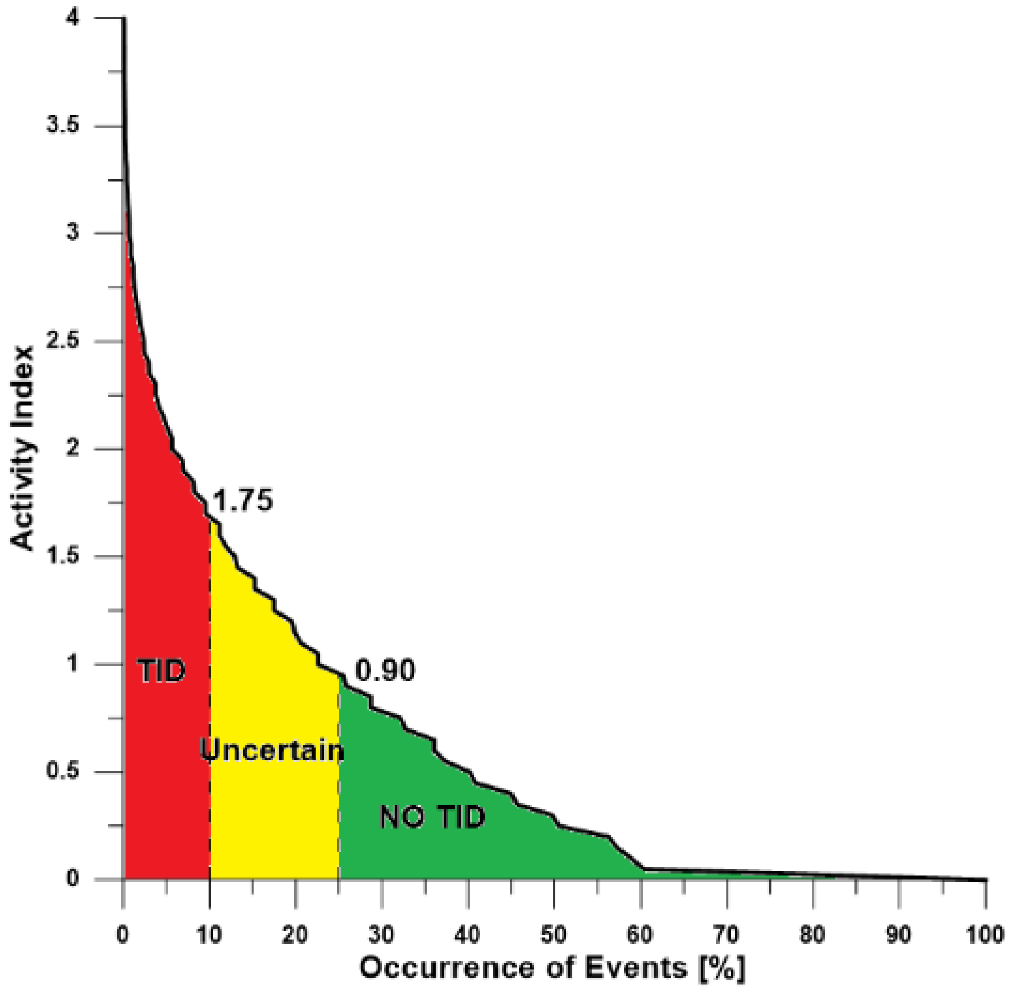
Products associated

- **HF-EU index**

```
EU 201701271330 TrL AV= 3.00 Area= 66.00% ActivityIndex= 1.98 TID
SA 201701271330 TrL AV= 0.00 Area=  0.00% ActivityIndex= 0.00 NO TID
```

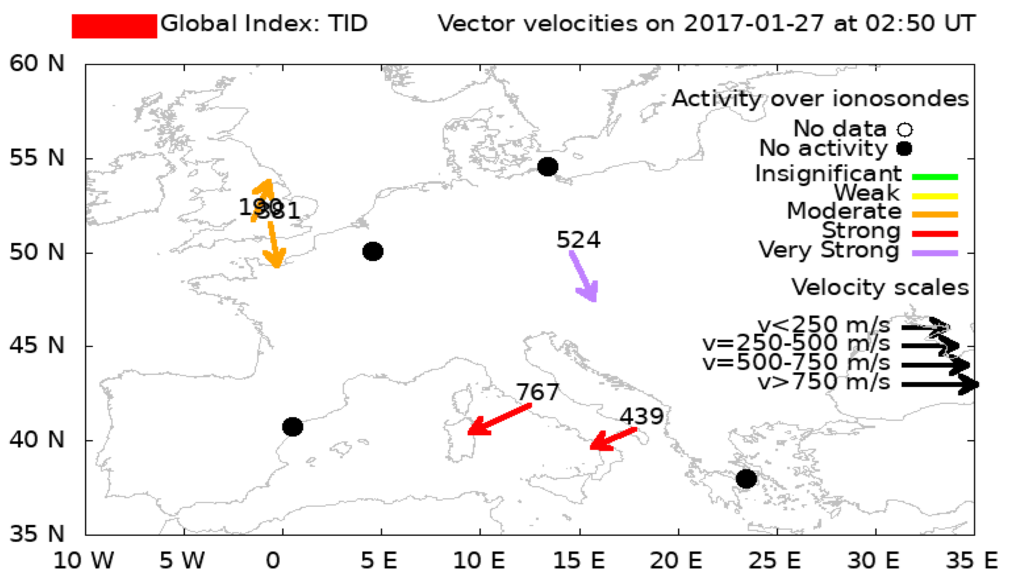
OE_HFI_YYYYMMDDHHmm_COND.log files every 5 minutes

- One index for the whole network.
  - It is the product of the average intensity of the TID (related to the spectral contribution) multiplied by the area affected (number of stations).
  - The thresholds have been established by statistics
    - 0 means no data
    - 0.1 means nothing detected



Detecting LSTIDs

Problems of the method



Sporadic E layer, Es

- We cannot see what is happening in F layer.
  - Affects specially on summertime at central hours of the day.

Lack of data

- Technical problems in some stations
- Connectivity problems with GIRO DIDBase
  - The TechTIDE portal storage the real-time data. To fix connectivity problems, time to time a reanalysis is carried out. But it is not storage in the TechTIDE portal

Uncertainty in the azimuth determination at the edge of the network

- The methodology to find the azimuth has an intrinsic uncertainty of 360º for stations located at the edge of the network (not usual but sometimes happens).

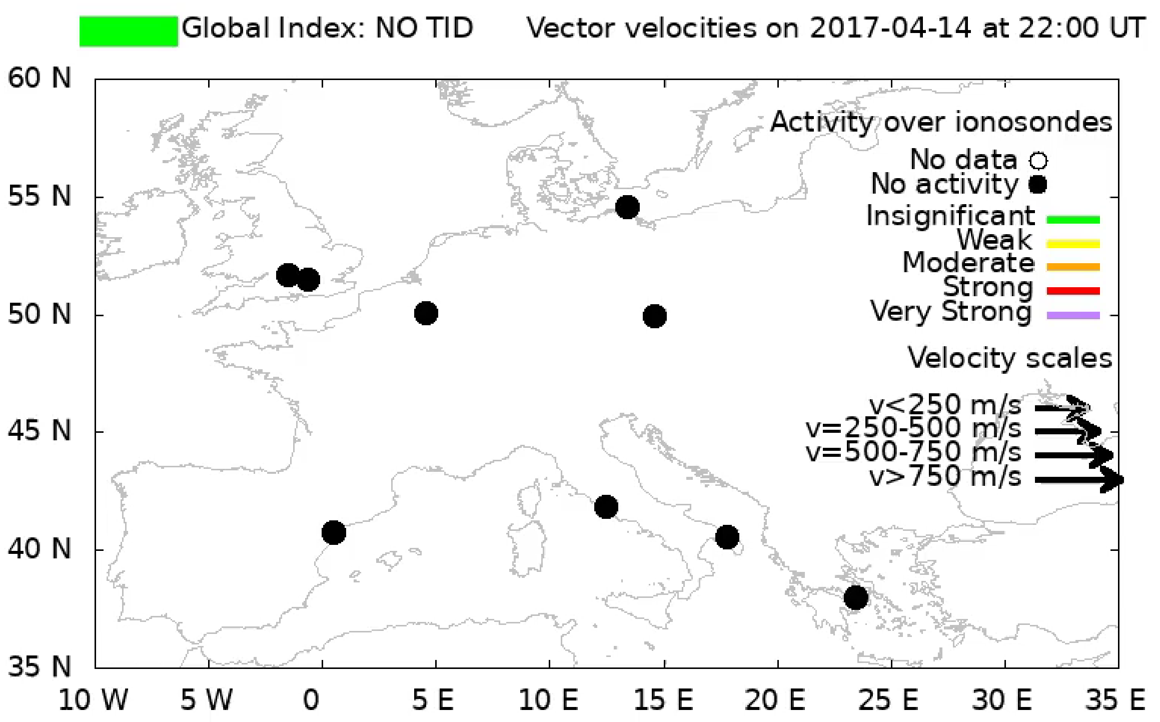
Intrinsic delay (Need to adopt a criteria for time detection)

- The detection time refers to the last download of the data. Then the method looks for periodicities in the previous 6 hours.
- As we look for periodicities in the input data, we need a full period to detect it.
- The method considers a detection if there is a **coherent periodicity in a minimum of 4 stations**. Then, a propagation time is needed to affect 4 station, it will depend on the azimuth of the perturbation and the velocity.
- Impact on the distribution of the time of detection

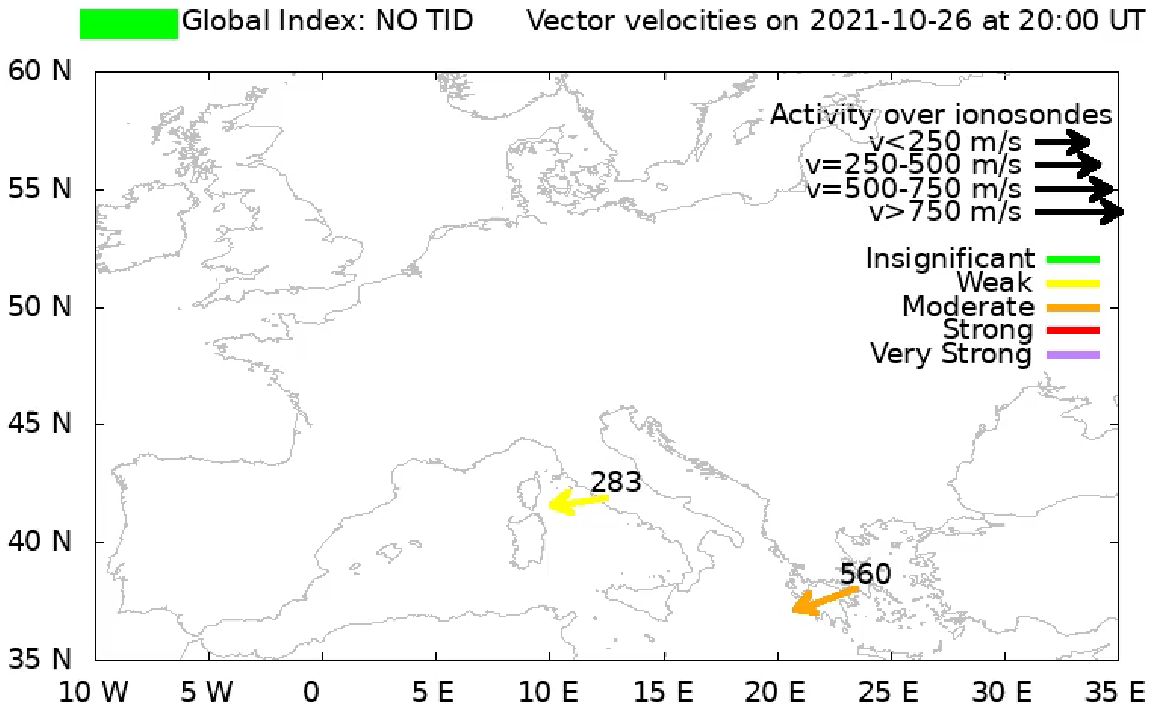
Detecting LSTIDs

HF-INT: Catalogue of events

- **Visual inspection to determine LSTIDs events**
  - ➢ Looking for coherent velocity propagation
  - ➢ 760 TIDs events detected and recorded above Europe between FEB 2014 to DEC 2022
- **Determination of onset time and duration**
  - ➢ Approximative
- **Average of the main characteristics of the TID for all stations and during the whole event.**



Included in the catalogue



Not included in the catalogue

HF-EU index vs HF-INT catalogue

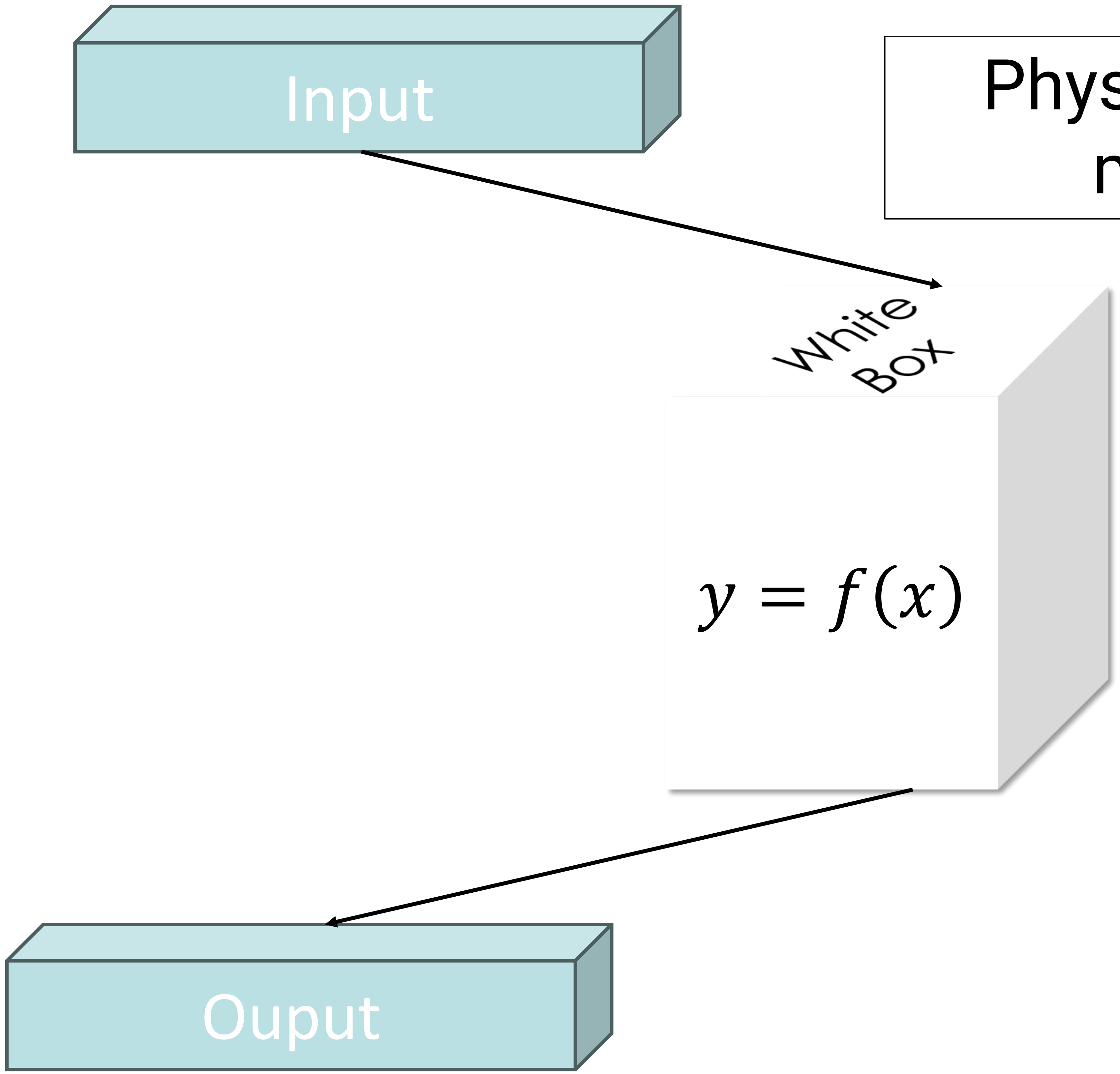
HF-EU index

- **Pros**
  - Automatic determination of the index clear criteria.
- **Cons**
  - Not all events with large index (above 1.75) are LSTIDs; presence of solar terminator effects and situations with a perturbation but with a non-coherent velocity.
  - No spatial information (you must go back to the raw data)
  - Although, you can determine an onset time automatically, you must keep in mind the delay problem of the method.

Catalogue

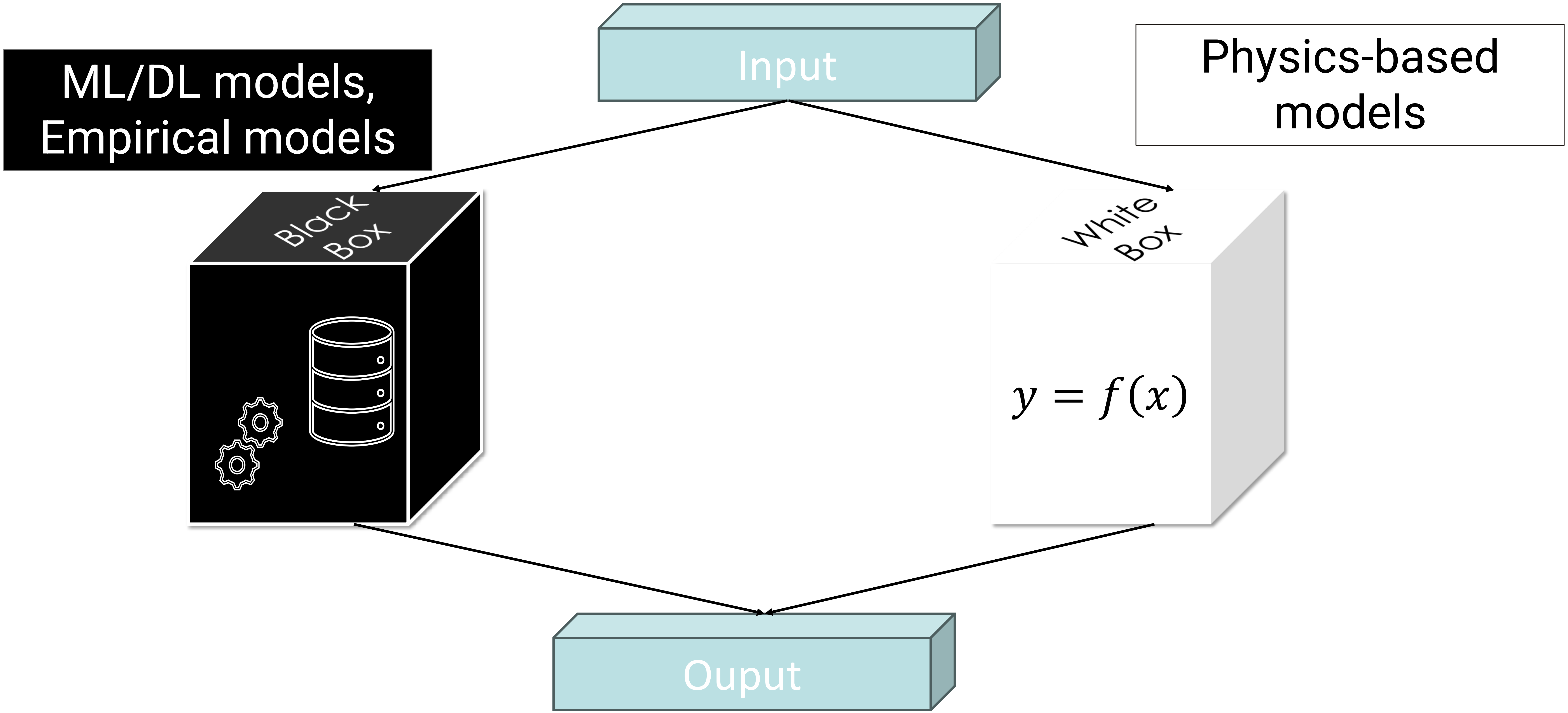
- **Pros**
  - We are sure that all events in the Catalogue are LSTIDs.
  - One file per year. Easy to work with.
- **Cons**
  - Not all TIDs are in the Catalogue, maybe not detected, no data, etc.
  - No spatial information (you must be back to the raw data)
  - Created by human inspection, probably biased.
  - Difficulties to determine the starting time and the duration of the event.

How to model and forecast an ionospheric phenomenon?



Physics-based models

How to model and forecast an ionospheric phenomenon?

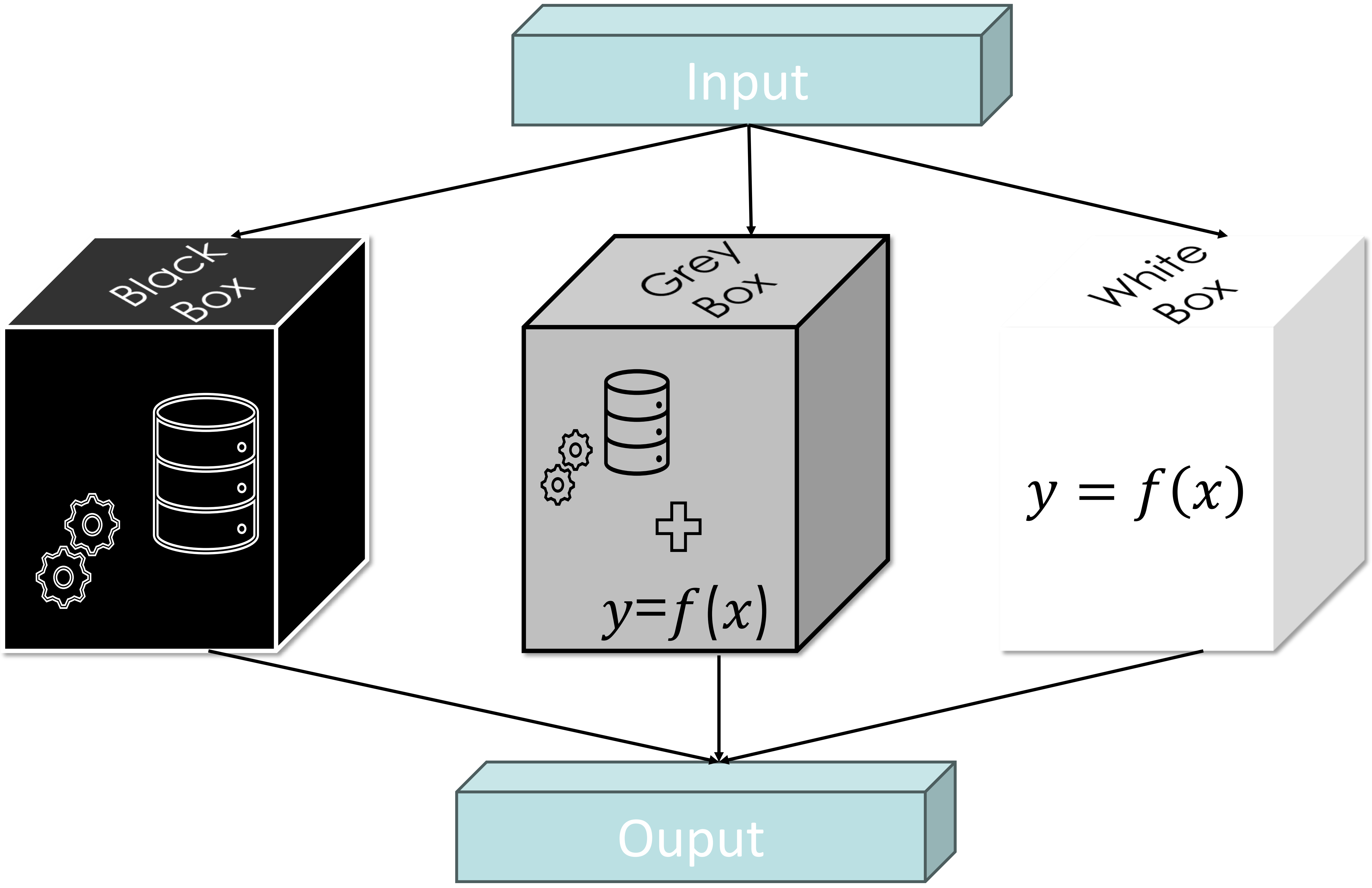


How to model and forecast an ionospheric phenomenon?

ML/DL models, Empirical models

Physics-based models

Physics Informed ML models, Data assimilation models, Explainable AI



ML based forecasting

1.

Exploratory data analysis

Impute missing values

Sync different sources

Data visualization & inspection

(garbage in, garbage out)

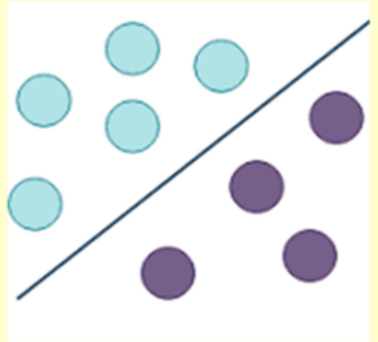
2.

Feature selection

Data normalization

Split training & validation sets

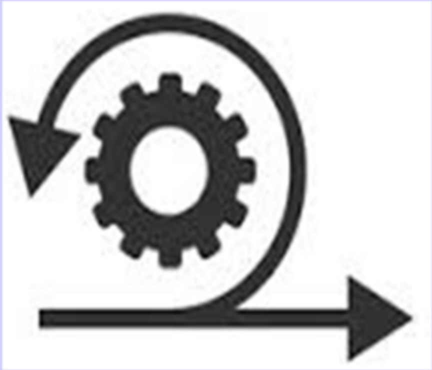
3.



Build and train Neural Network

Parameter tuning & cross-validation

4.



Deploy model

Performance evaluation

Re-train Neural Network

Interpretability vs Accuracy trade-off

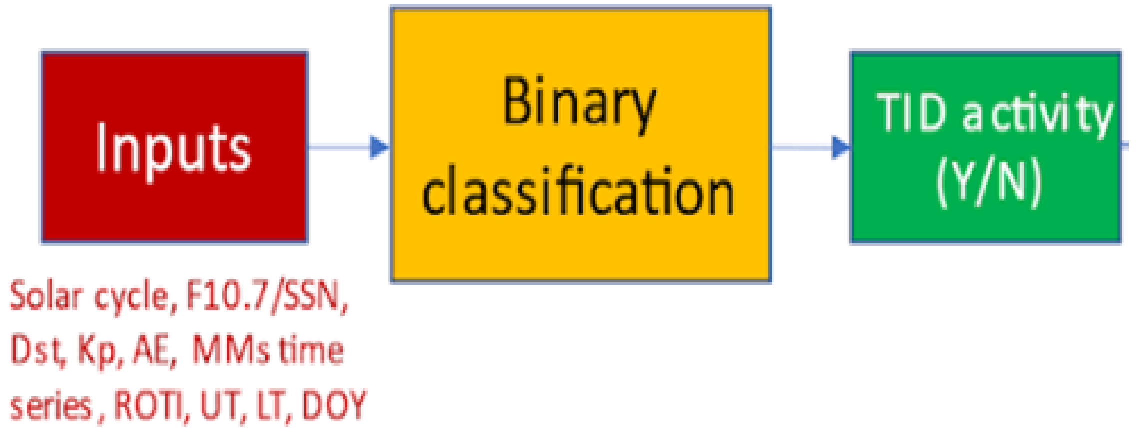
	Interpretable		Accurate
Complex model	✗	←	✓
Simple model	✓	→	✗

Interpretable or accurate: choose one

Catalogue based model

Developing the ST ML models: catalogue-based forecasting

- The problem is handled as a binary classification
- We are working with the **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°

Our features and labels so far…

Variables **input** to the model fall into the following categories:

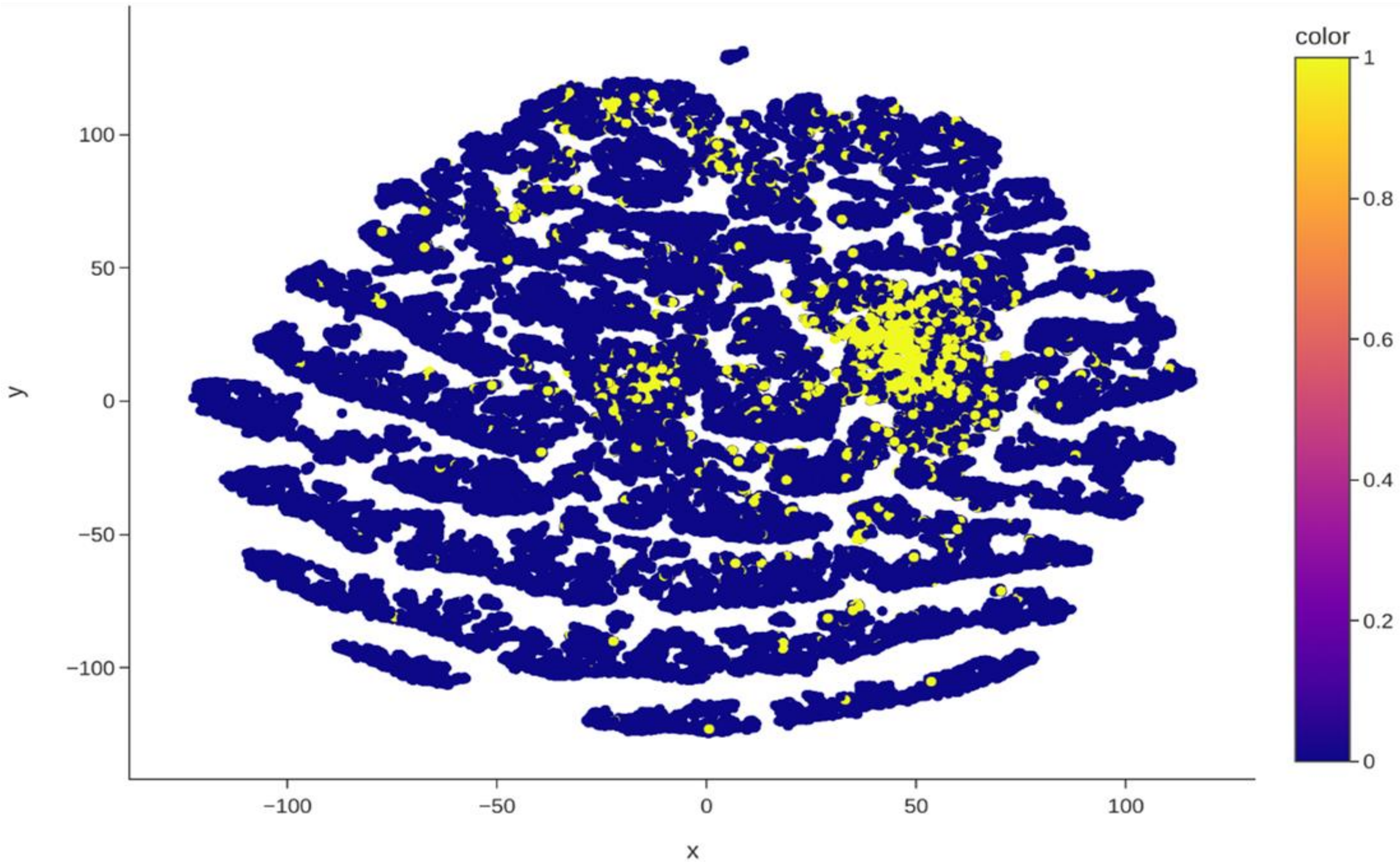
- **Geomagnetic indices**: IE, IL, IU, HP-30, SMR and moving averages for those variables;
- **HF related**: HF-EU index; single station spectral contribution, azimuth, velocity
- **Solar**: F10.7, solar zenith angle;
- **Solar wind and IMF:** B_z, v_x, ρ;

Given the time series (with a time resolution of 30 mins), we create the **target** as a feature taking on two values:

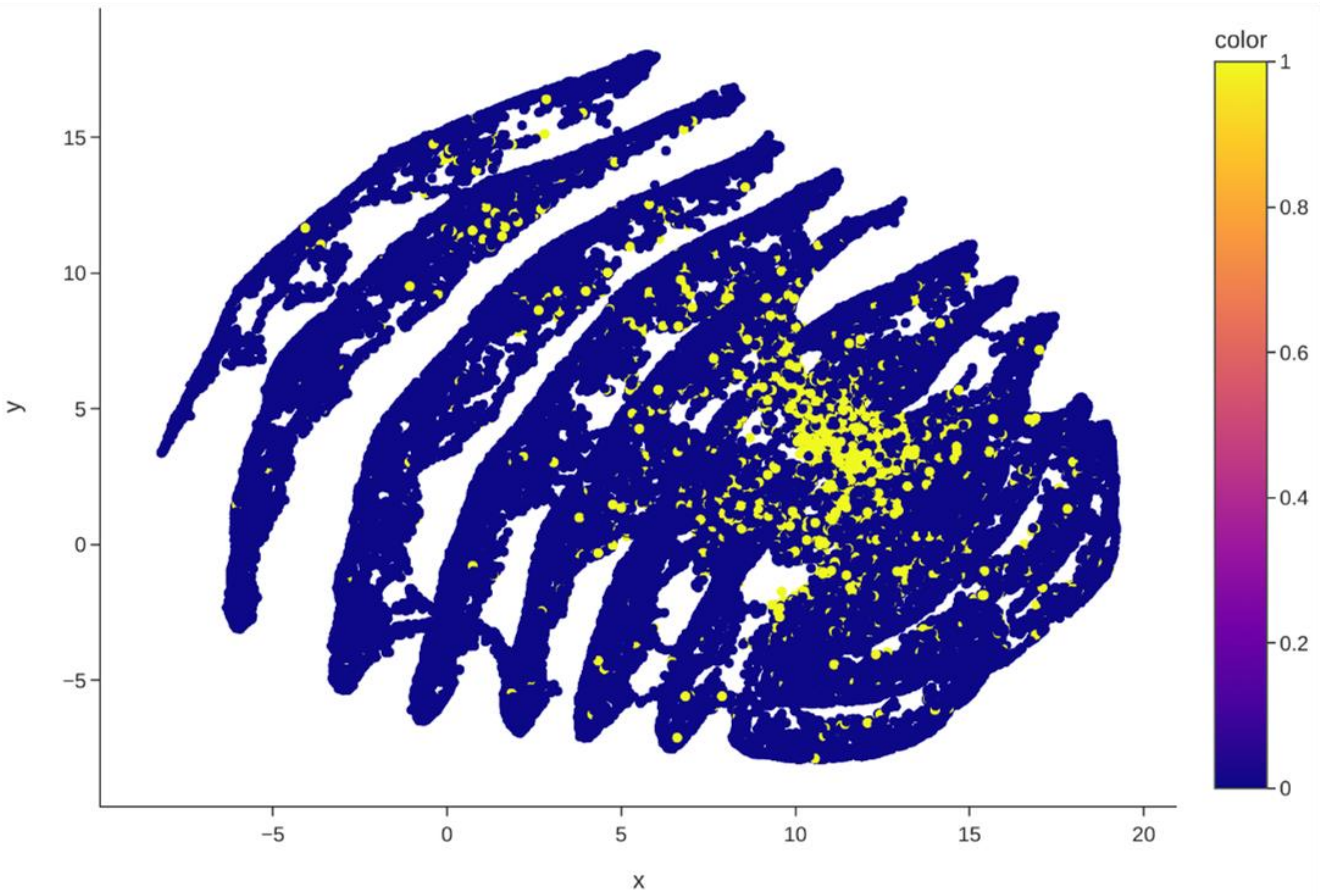
- **1** from 3 hours before the start of an LSTID until its end;
- **0** otherwise;

*We can frame our problem as a **multivariate time-series binary classification***

Features space



t-distributed Stochastic Neighbor Embedding (t-SNE)



Uniform Manifold Approximation and Projection (UMAP)

Our ML stack

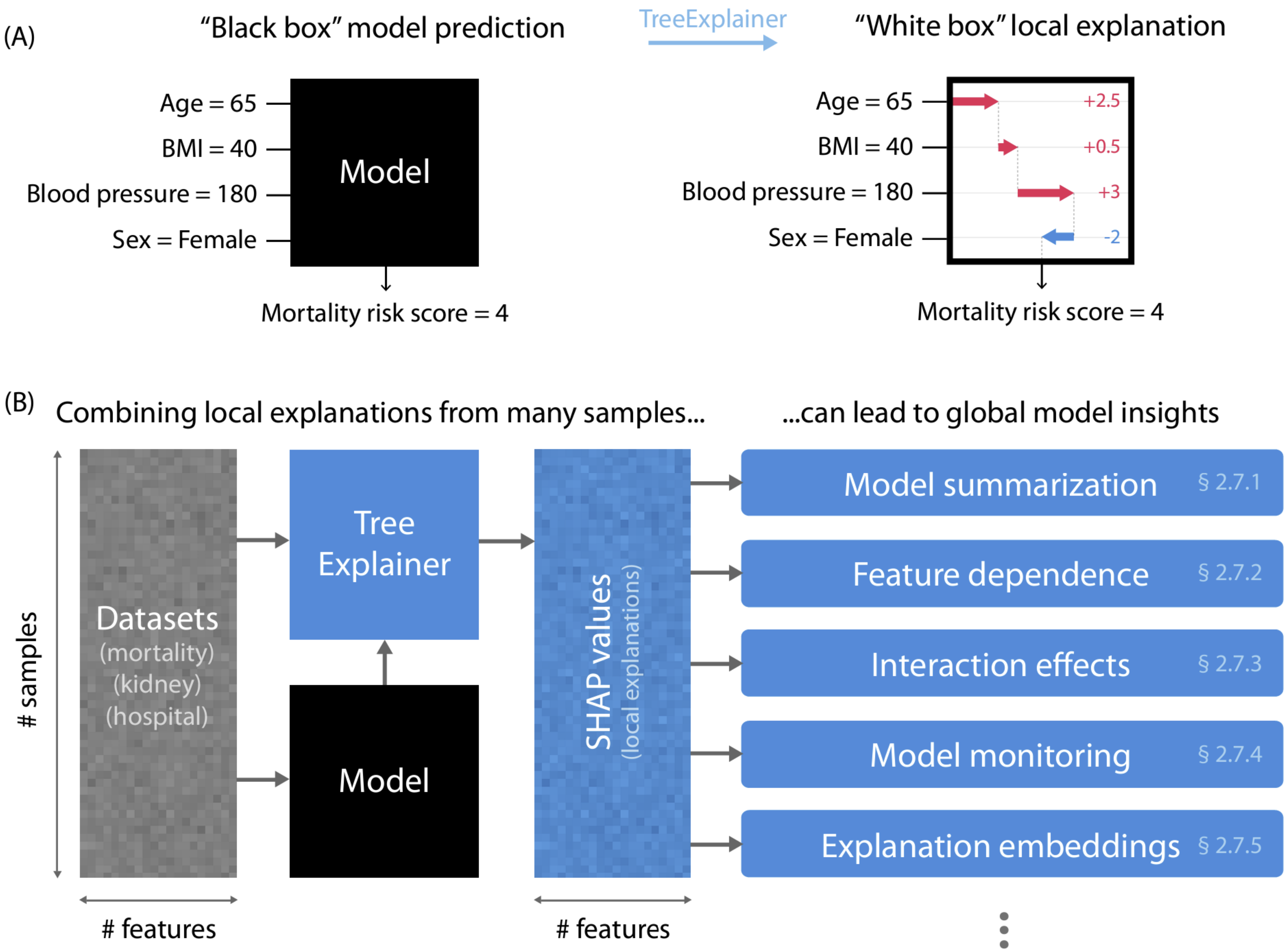
- Easily understandable and adaptable syntax
- One of the top languages for training ML models

- **Cat**egory & **Boost**ing (gradient boosting on decision trees)
- A symmetric balanced tree architecture leads to an efficient CPU implementation, decreases prediction time (great for real-time inference) and controls overfitting
- Categorical and missing values are handled natively
- Integrates SHAP to break predictions into features' contributions

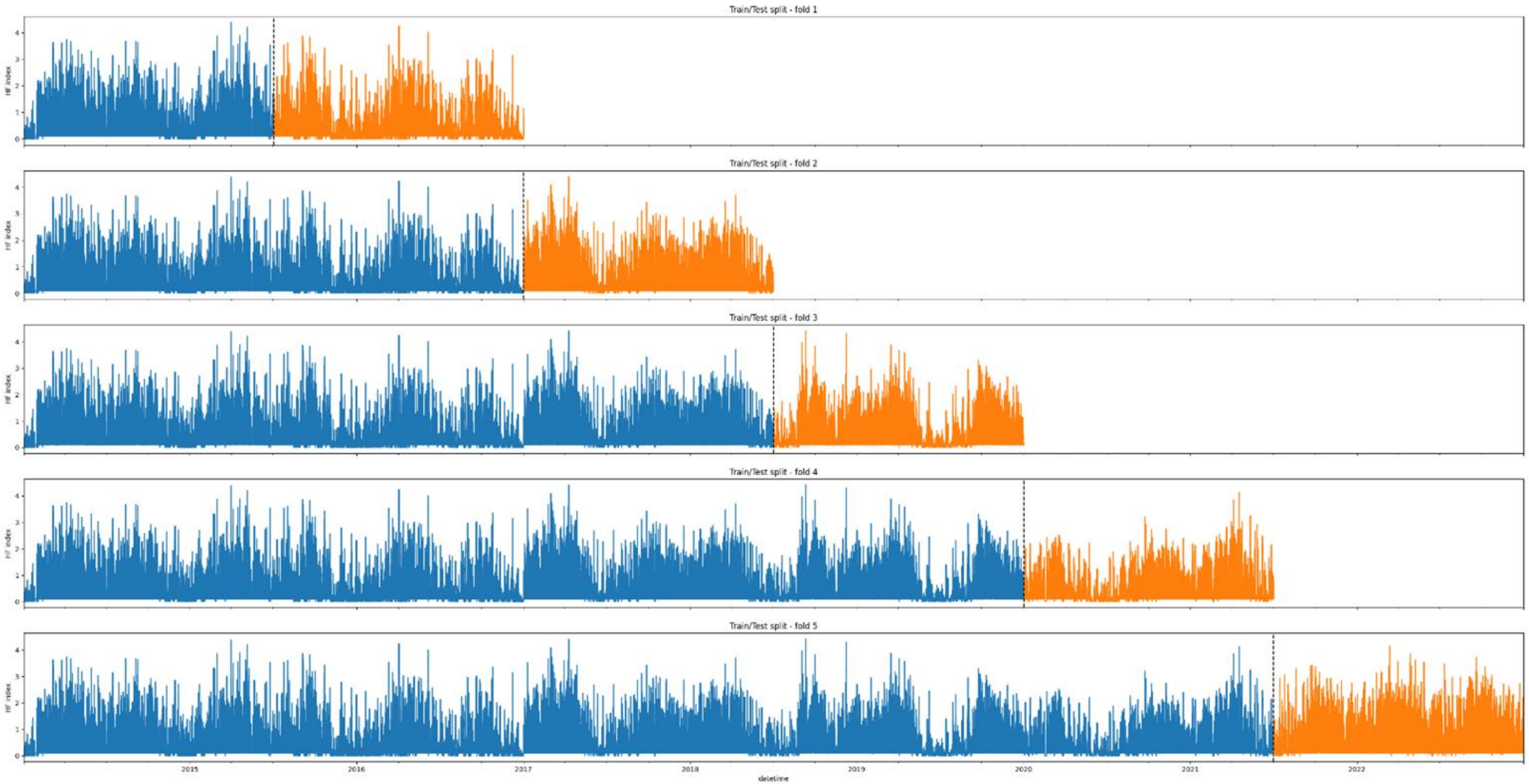
- Efficient optimisation framework for model hyper-parameters tuning
- **M**achine **L**earning **Op**eration**s** (MLOps) to organise and manage ML experiments

- The **SH**apley **A**dditive ex**P**lanation (SHAP) framework allows to test features influence on the model output from both global and local aspects
- Enhancement for interpretability and explainability of the model – very desirable features in potentially high-risk settings

How can we explain the model?

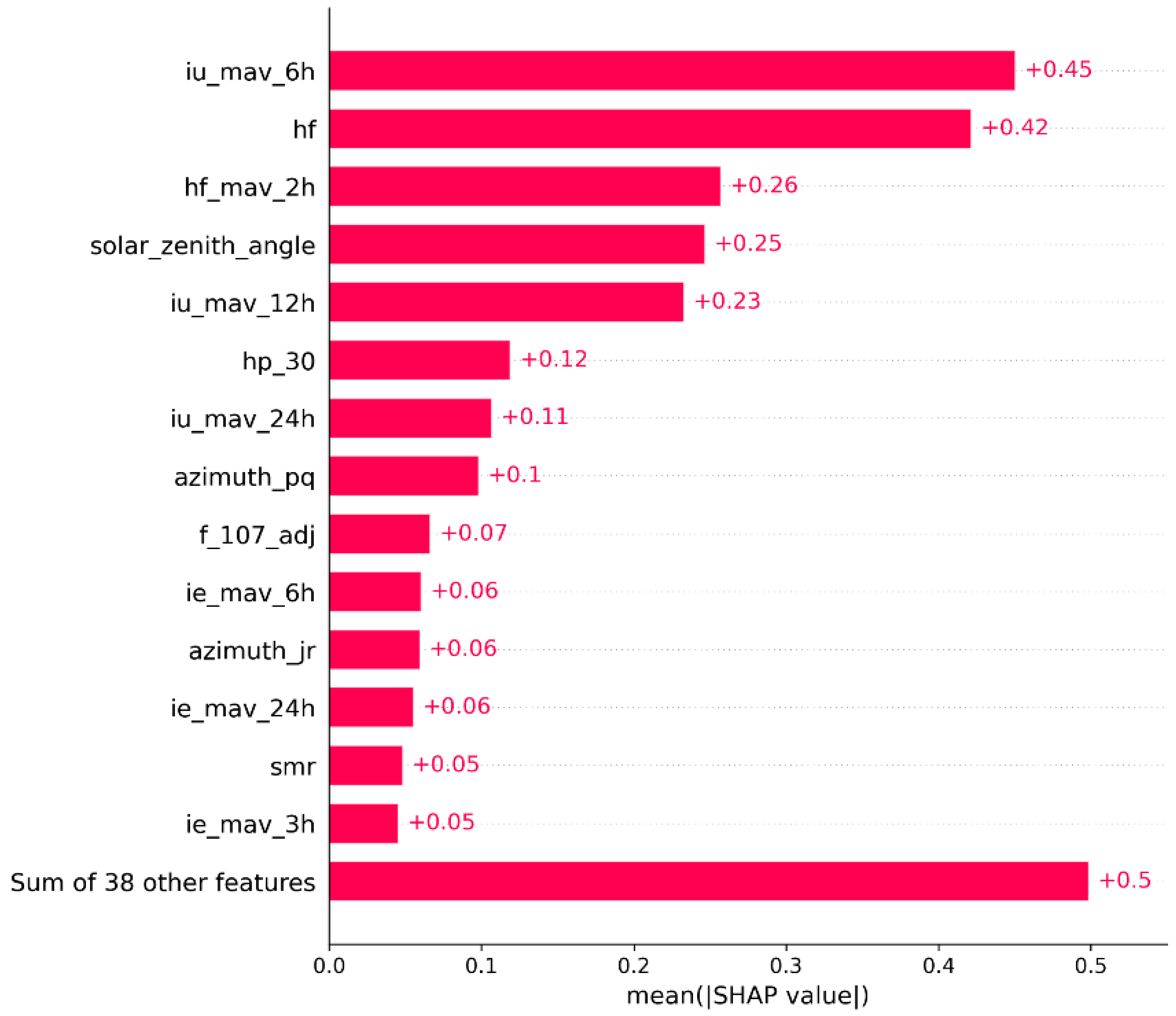


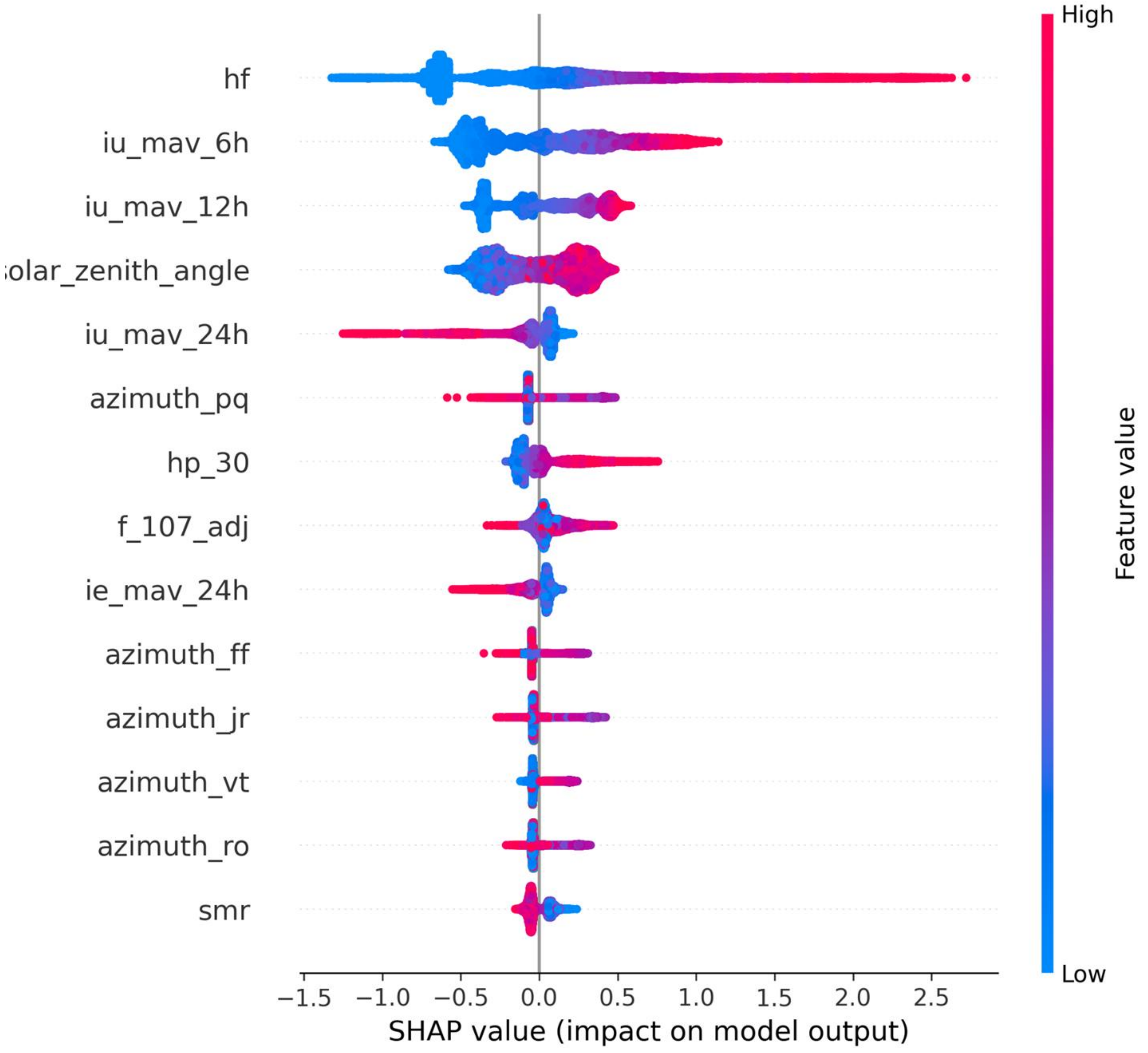
Training-Validation dataset splitting



Why so?
Because time series samples, in general, are not independent and identically distributed

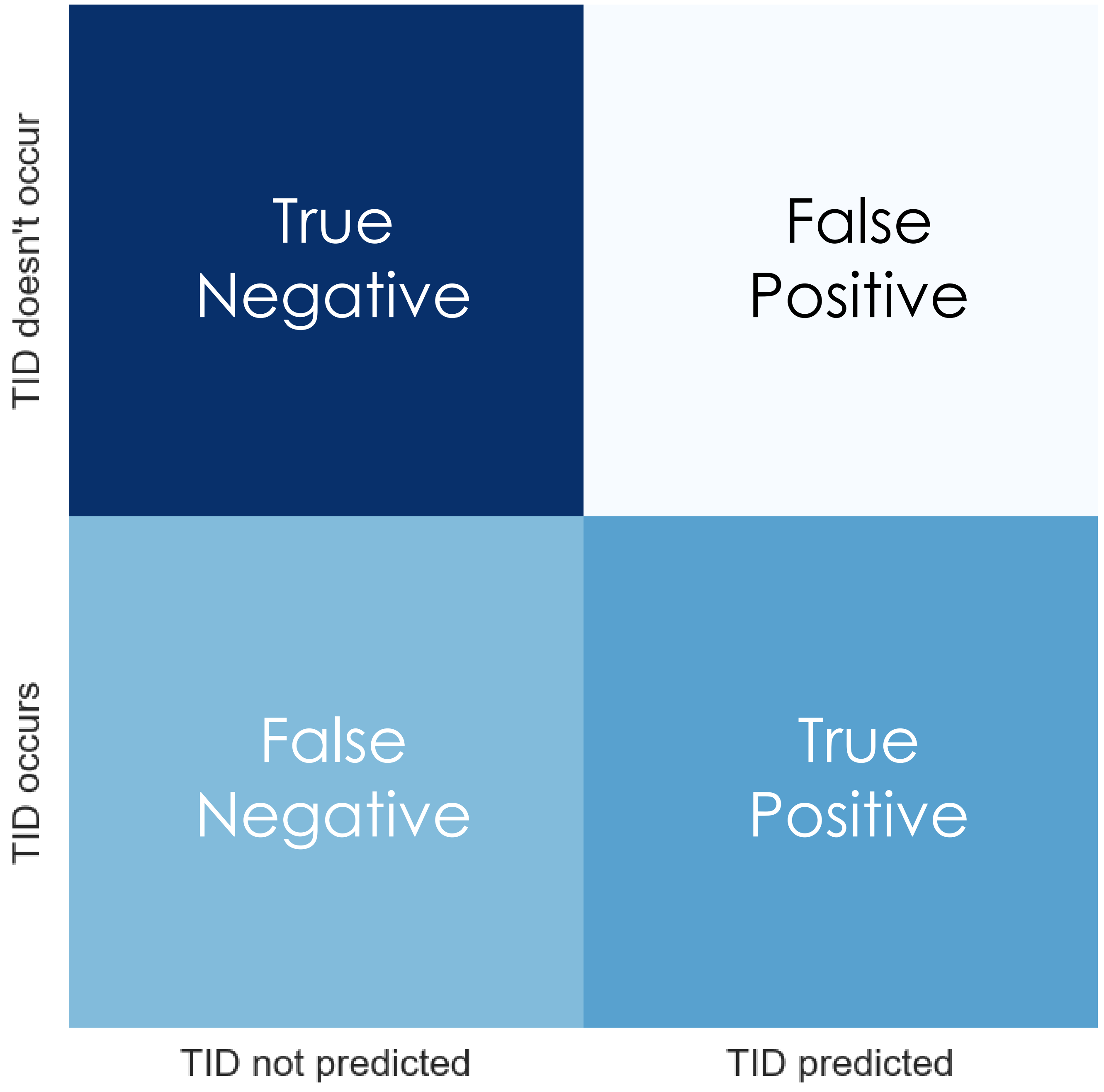
Local to global interpretation: feature importance





Local interpretation: 23/12/2022 22:20, LSTID lasted 1.5h

Global performance evaluation



$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

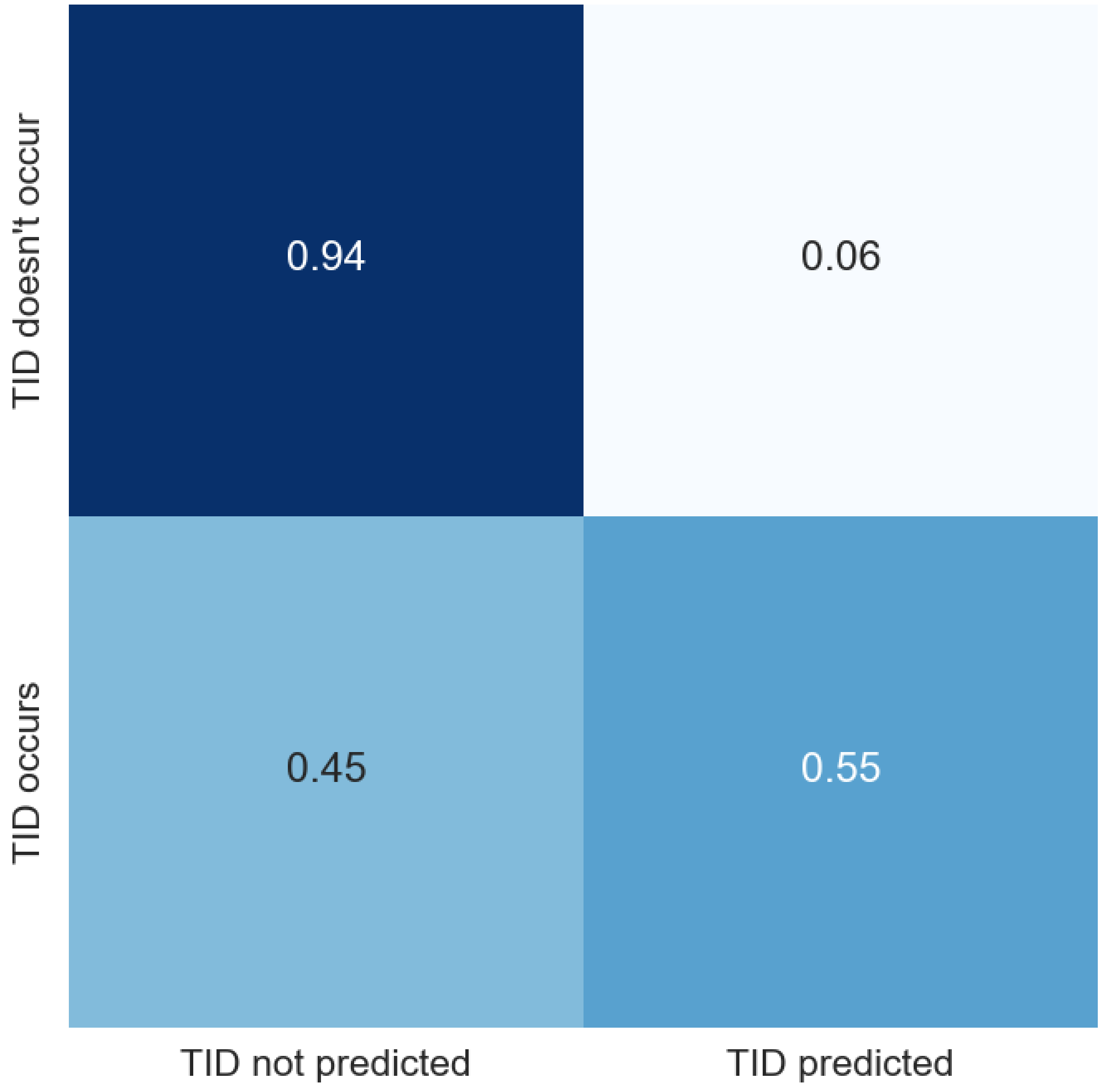
$$F_1 = \frac{P * R}{P + R}$$

Precision is a good measure to determinee, when the costs of False Positive is high

Recall actually calculates how many of the Actual Positives our model capture through labeling it as Positive (True Positive)

F1 Score might be a better measure to use if we need to seek a balance between Precision and Recall AND there is an uneven class distribution (large number of Actual Negatives).

Model performance so far…



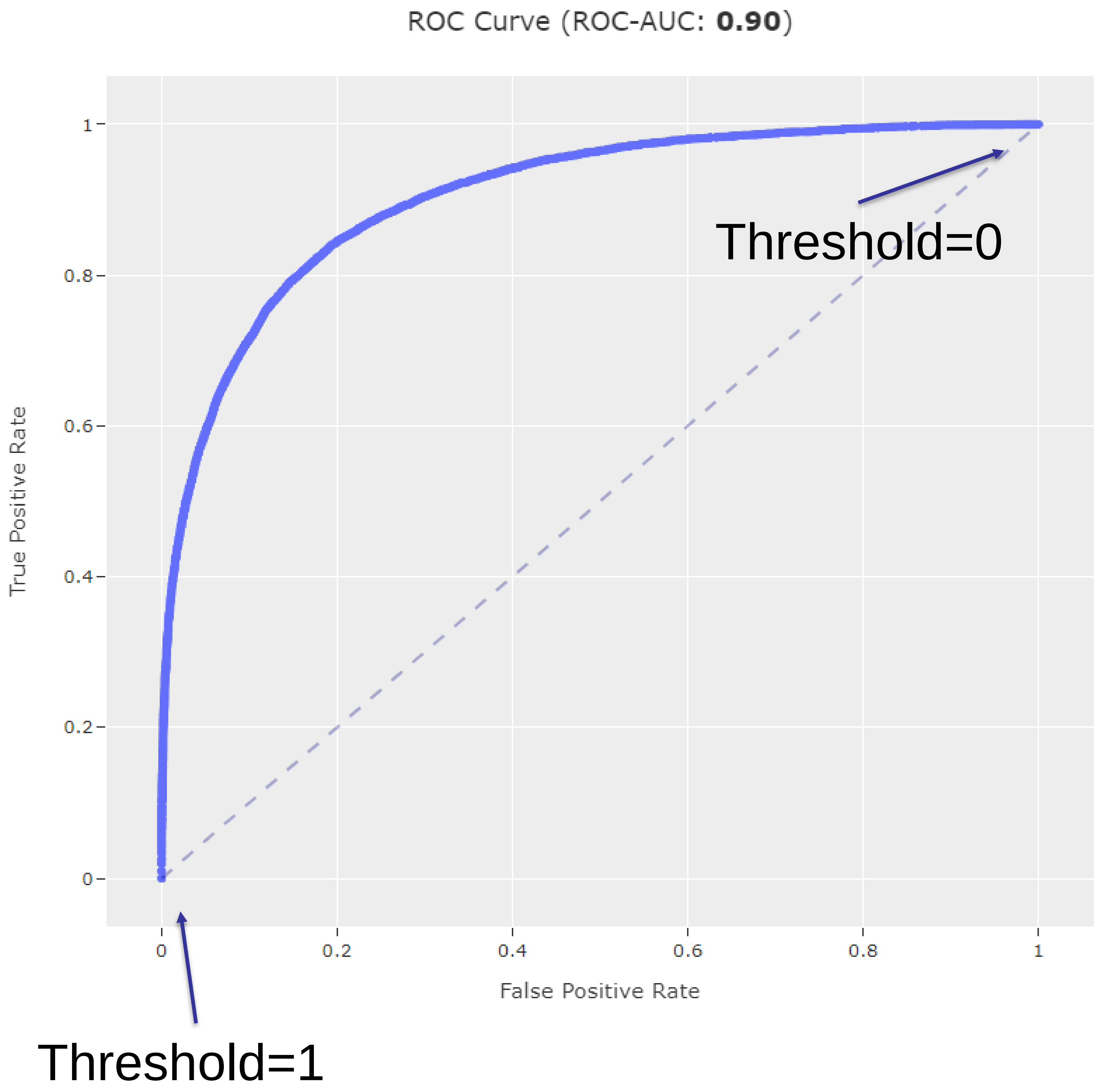
	No LSTIDs	LSTIDs
F1 score	0.97± 0.01	**0.49 ± 0.04**
Precision	0.97± 0.01	**0.49 ± 0.01**
Recall (sensitivity)	0.96± 0.02	**0.50 ± 0.07**

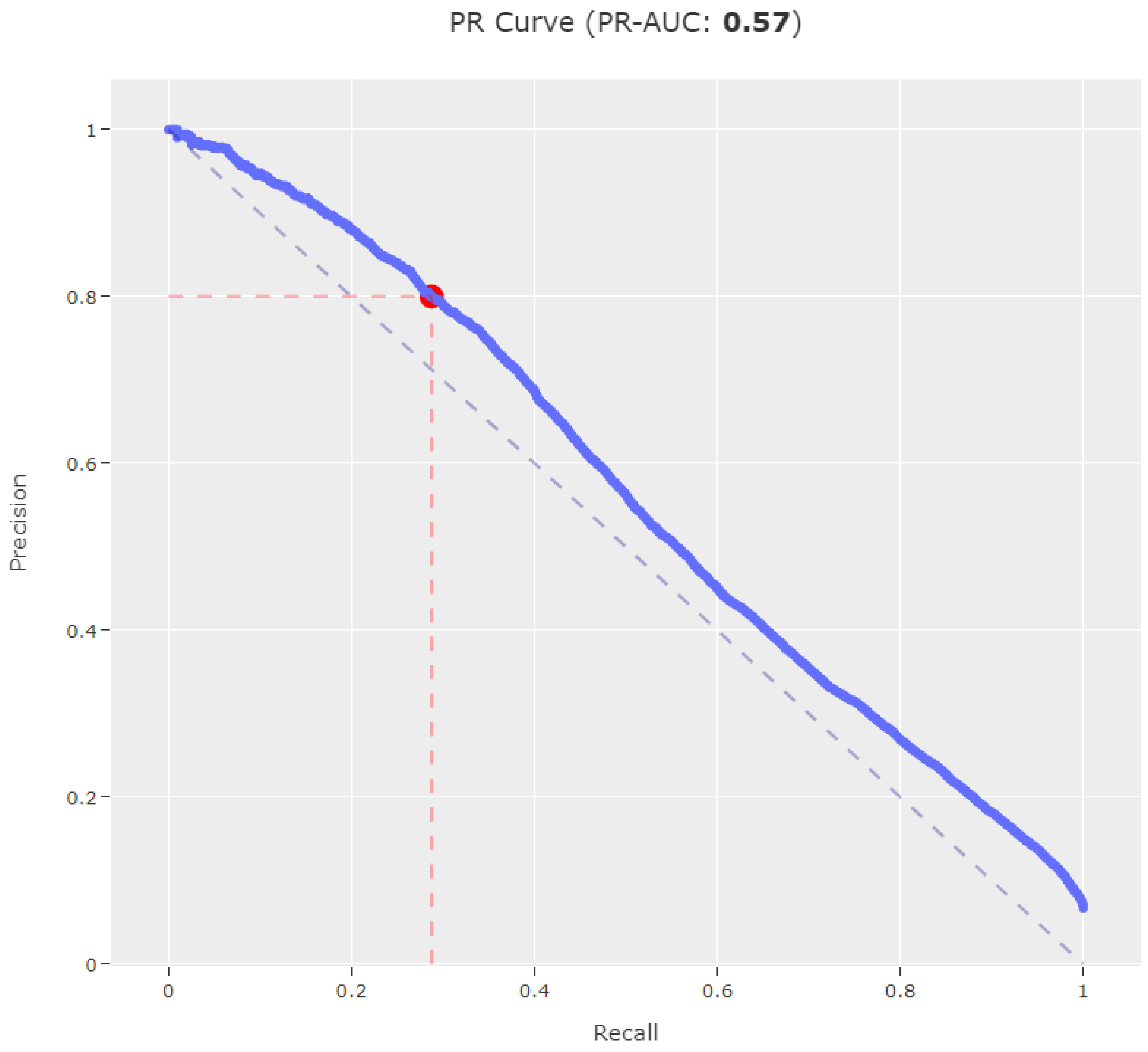
$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$F_1 = \frac{P * R}{P + R}$$

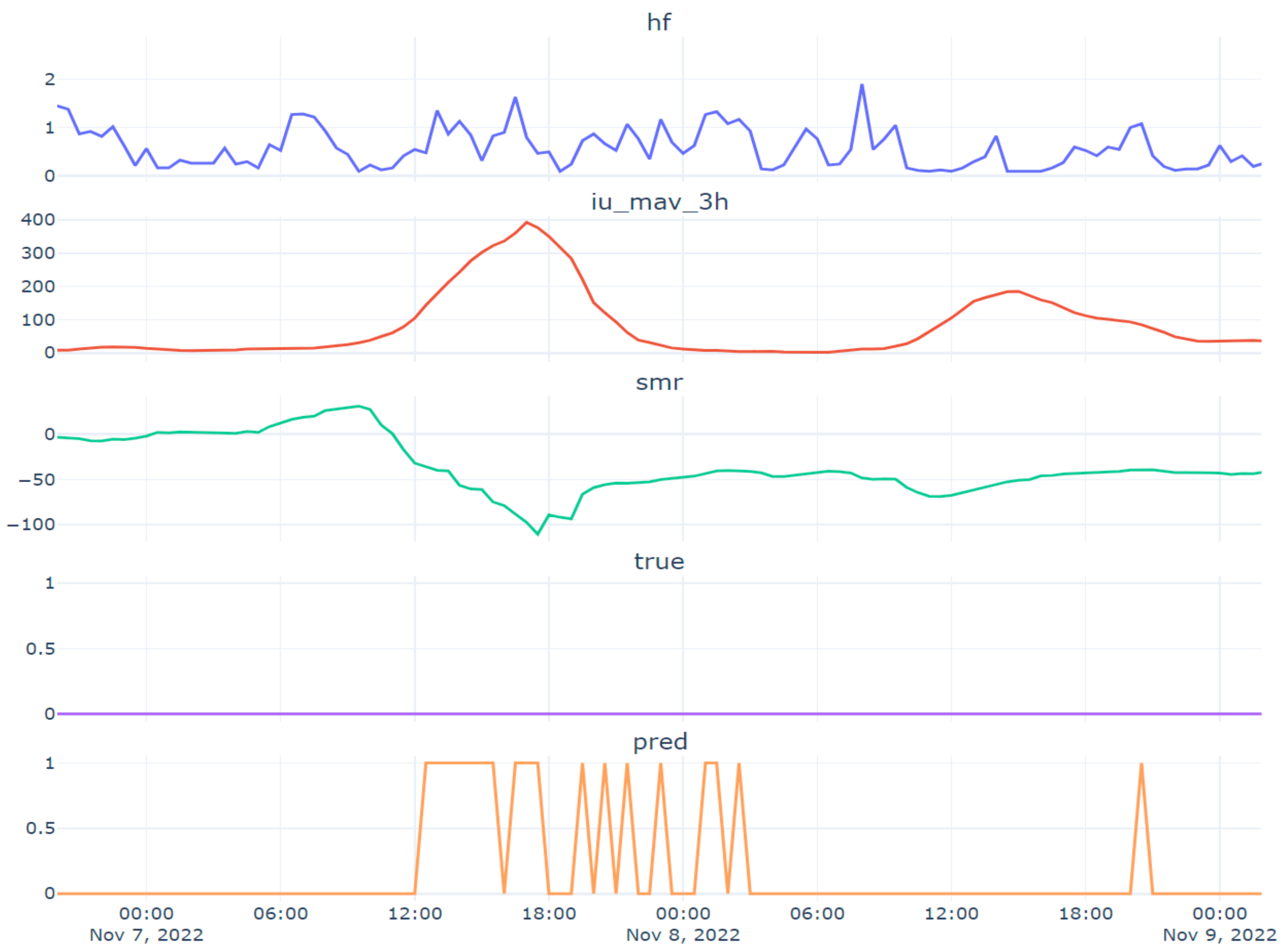
Model performance so far…

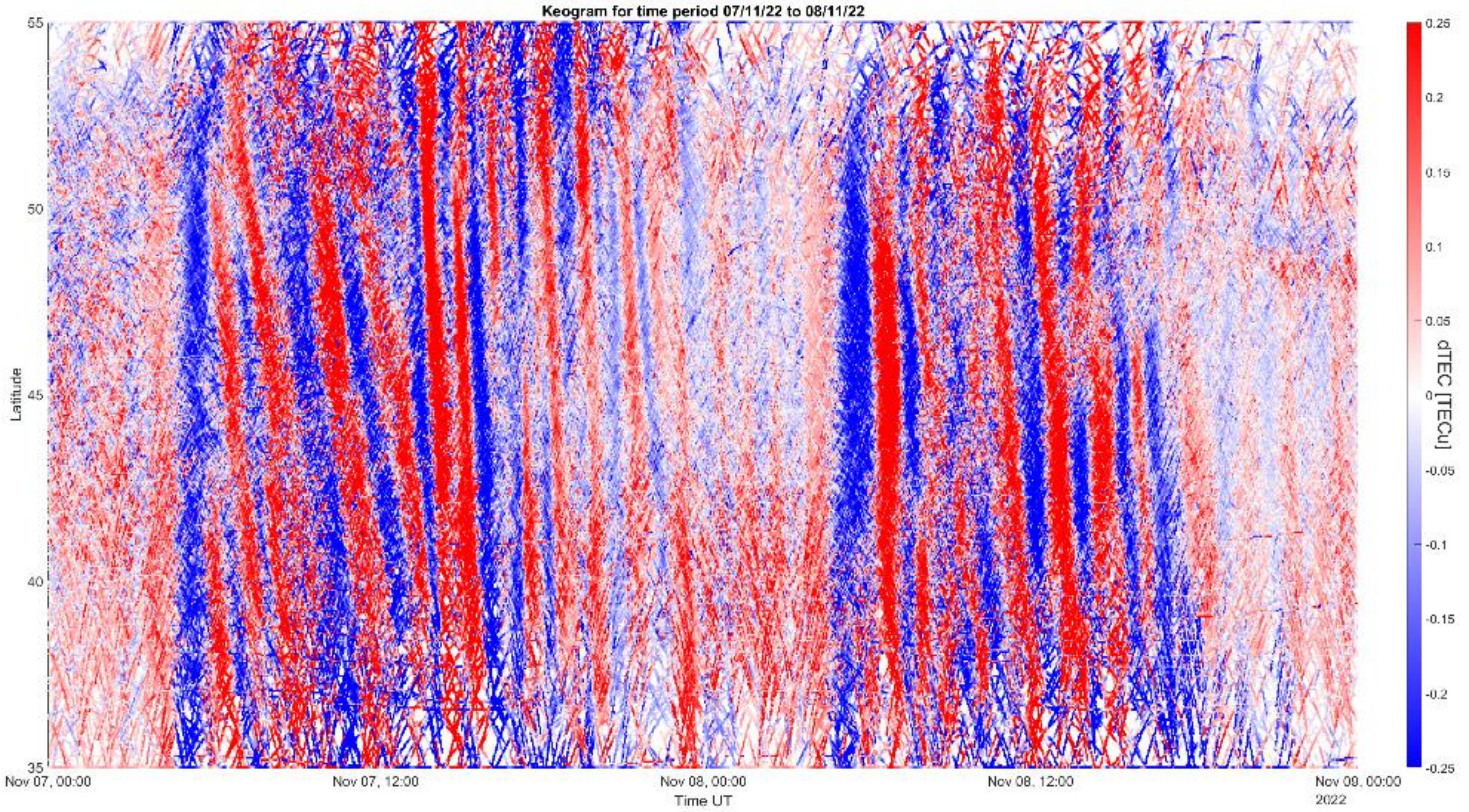




$$R = \frac{TP}{TP + FN} \qquad P = \frac{TP}{TP + FP} \qquad F_1 = \frac{P * R}{P + R}$$

How to improve the model performance…





- The model correctly predicts some LSTID occurrences (12:00 – 03:00), which were not in the HF catalogue (True) but apparent in GNSS-derived dTEC (Keogram)
- Despite that, the model still struggles to confidently predict TID occurrences (the prediction oscillates between true and false)
- Nevertheless, the model does not predict the LSTIDs happening during daytime of the 8th of November

Take home messages and next steps

- LSTIDs are due a complex chain of physical phenomena hardly predictable
- There are no physical models capable of predicting LSTID so far
- The T-FORS project is trying to develop a prototype model based on ML/techniques facing the problem as a binary classification and regression for HF-EU index prediction (not shown in this talk)
- So far, a gradient boosting on decision tree model seems to be promising in predicting the occurrence of LSTID a few hours in advance
- SHAP values give the opportunity to locally interpret the model results and, in turn, to globally define the features importance.
- Ancillary data (e.g. detrended TEC) can be used to identify both failures of the prediction model and LSTIDs not included in the catalogue
- Identify additional important features to drive the model
- Include other data sources to create a more complete and reliable catalogue of the events (automatic detection algorithm based on detrended TEC is under development)

Thank you for your attention!

Funded by the European Union

The T-FORS project is funded by the European Union (GA-101081835). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

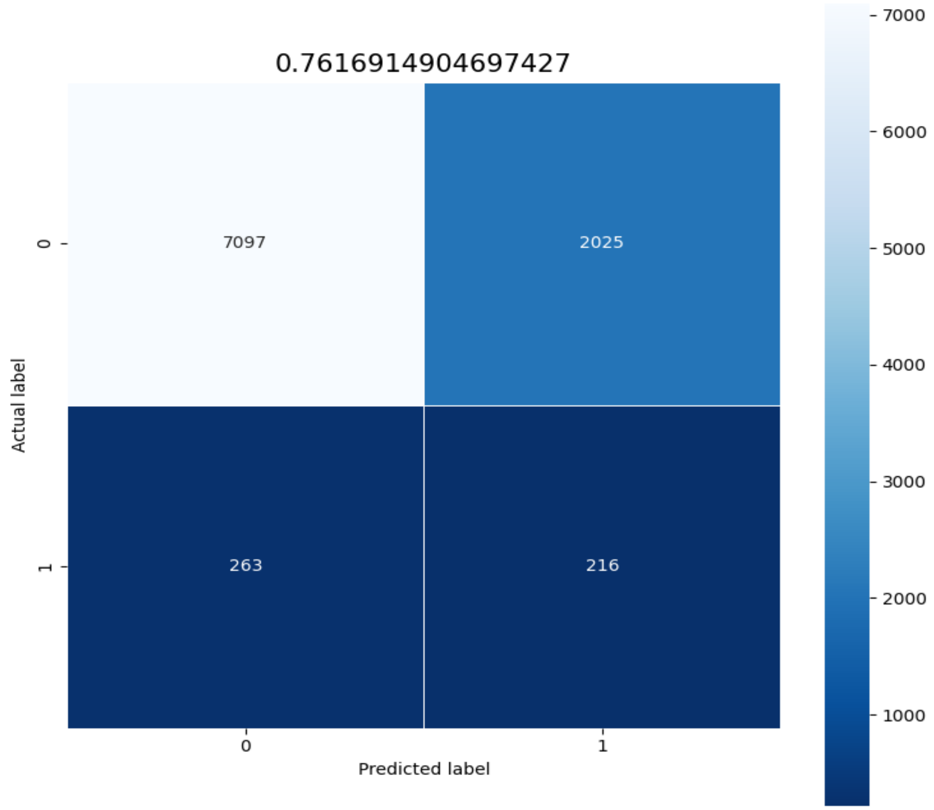
Backup slides

WP2 Results

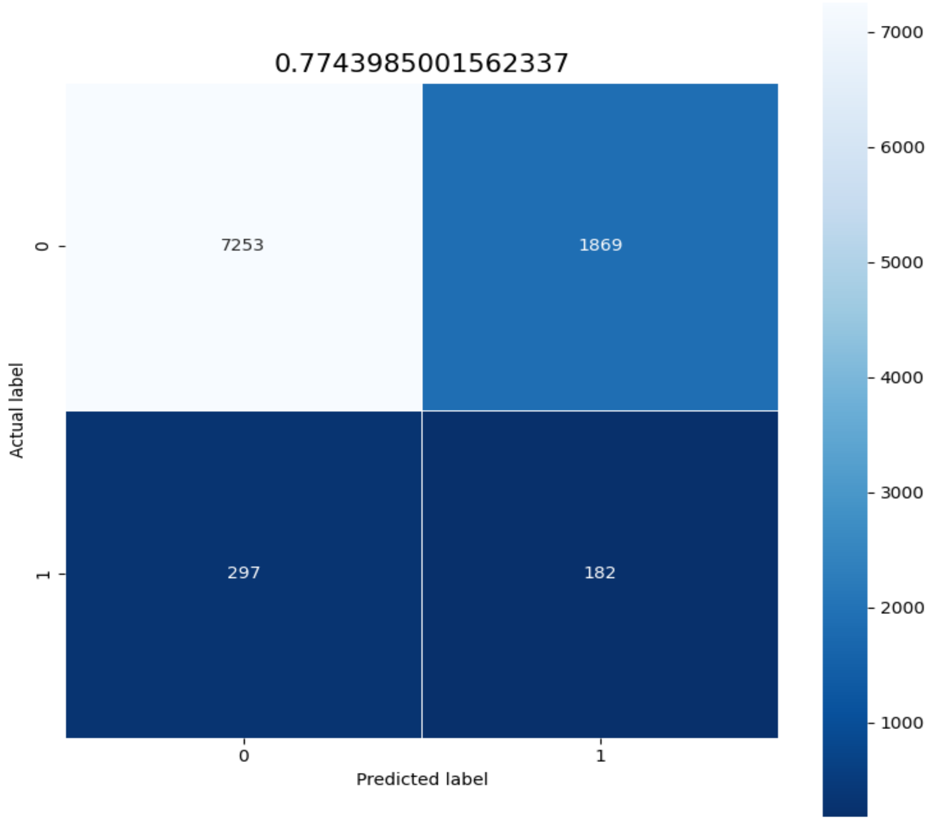
Developing the ML models: catalogue-based forecasting

- 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



WP2 Results

Developing the ST-HA ML models: LSTID indices-based forecasting

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

Research scenarios:

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

Scenario 2: **Prediction** of LSTID based **exclusively** on the **most recent** $IL, IU, GNSS, LT$ values.

Rationale: Is it possible to have an (even less reliable) decision on whether an $LSTID$ occurs, in the case where the most recent $SPcont$ values are missing?

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

Rationale: How reliable is a decision on whether an $LSTID$ occurs, in the case where (a) the **available** most recent $\boldsymbol{SPcont}$ values and (b) the most recent $IL, IU, GNSS_{HL}, LT$ values are considered?

Classifiers employed:

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

WP2 Results

Developing the ML models: LSTID indices-based forecasting

<u>Scenario 1:</u> **Prediction** of LSTID based **exclusively** on the **most recent** $SPcont$ values.

$$c(t) = \begin{cases} 0, & SPcont < T \\ 1, & SPcont > T \end{cases}$$

<u>**Data set creation**</u>

FNN classifier (P-dimensional input $\leftrightarrow$1-dim. output)

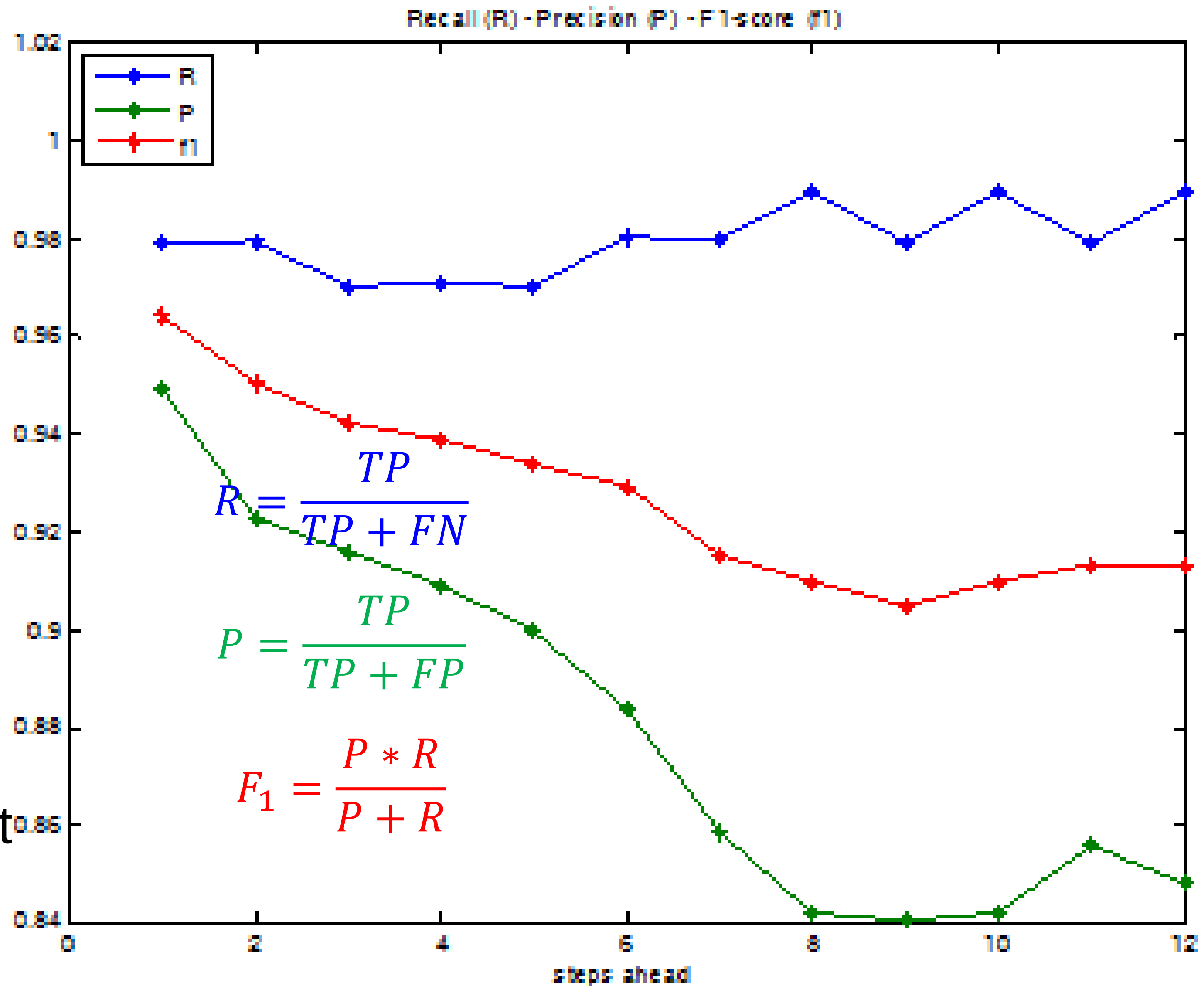
$$[SP(t-P-s+1), \ldots, SP(t-s)] \leftrightarrow c(t)$$

$\boldsymbol{P}$ = past-size window**,** $\boldsymbol{s}$ = no. of time-steps ahead **(5mins resolution)**

$$T = 0.5\,, \quad P = 24\,, \quad s = 1, \ldots, 12$$

Remarks:

- Trained **FNN** classifier has high classification performance as evidenced by precision and recall metrics
- $SPcont$ classes can be forecasted with high accuracy, given that past $SPcont$ observations are available (5% of cases)



WP2 Results

Developing the ST-HA ML models: LSTID indices-based forecasting

Scenario 2: Prediction of LSTID based exclusively on the most recent IL, IU, $GNSS$, LT values.

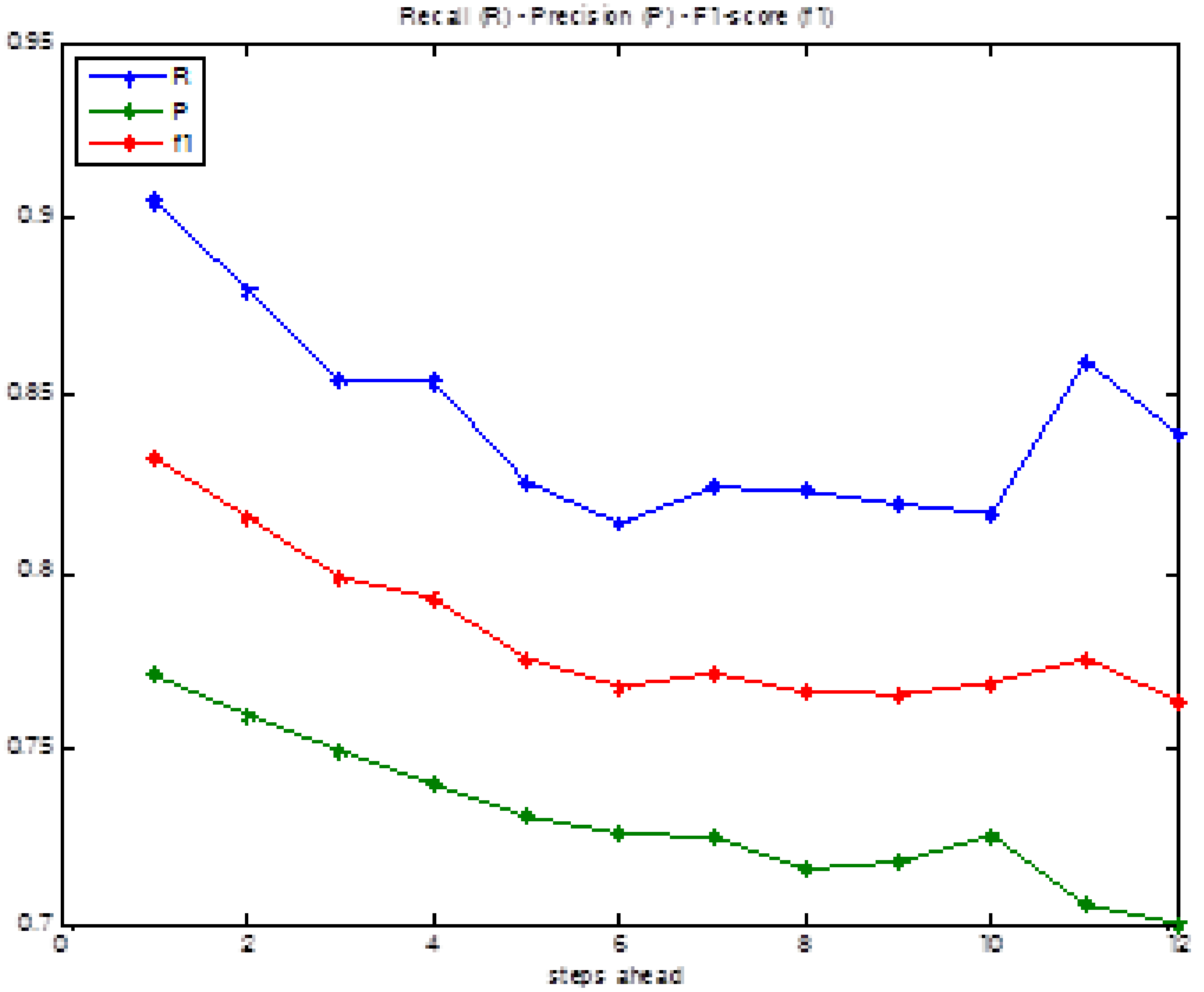
Data set creation

FNN classifier (4-dimensional input ⟷1-dim. output)

$$c(t) = \begin{cases} 0, & SPcont < T \\ 1, & SPcont > T \end{cases}$$

$$[min(IL(t-P-s+1:t-s)), max(IU(t-P-s+1:t-s)), max(GNSS(t-P-s+1:t-s)), LT(t)] \leftrightarrow c(t)$$

$\boldsymbol{P}$ = past-size window, $\boldsymbol{s}$ = no. of time-steps ahead **(5mins resolution)**



Remarks:

The results of **the FNN classifier** are less accurate than in the case where **only the $SPcont$** past values are considered (Scenario 1) (however, in the latter case, no missing values were considered).

WP2 Results

Developing the ST-HA ML models: LSTID indices-based forecasting

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

FNN classifier (5-dimensional input ⟷1-dim. output)

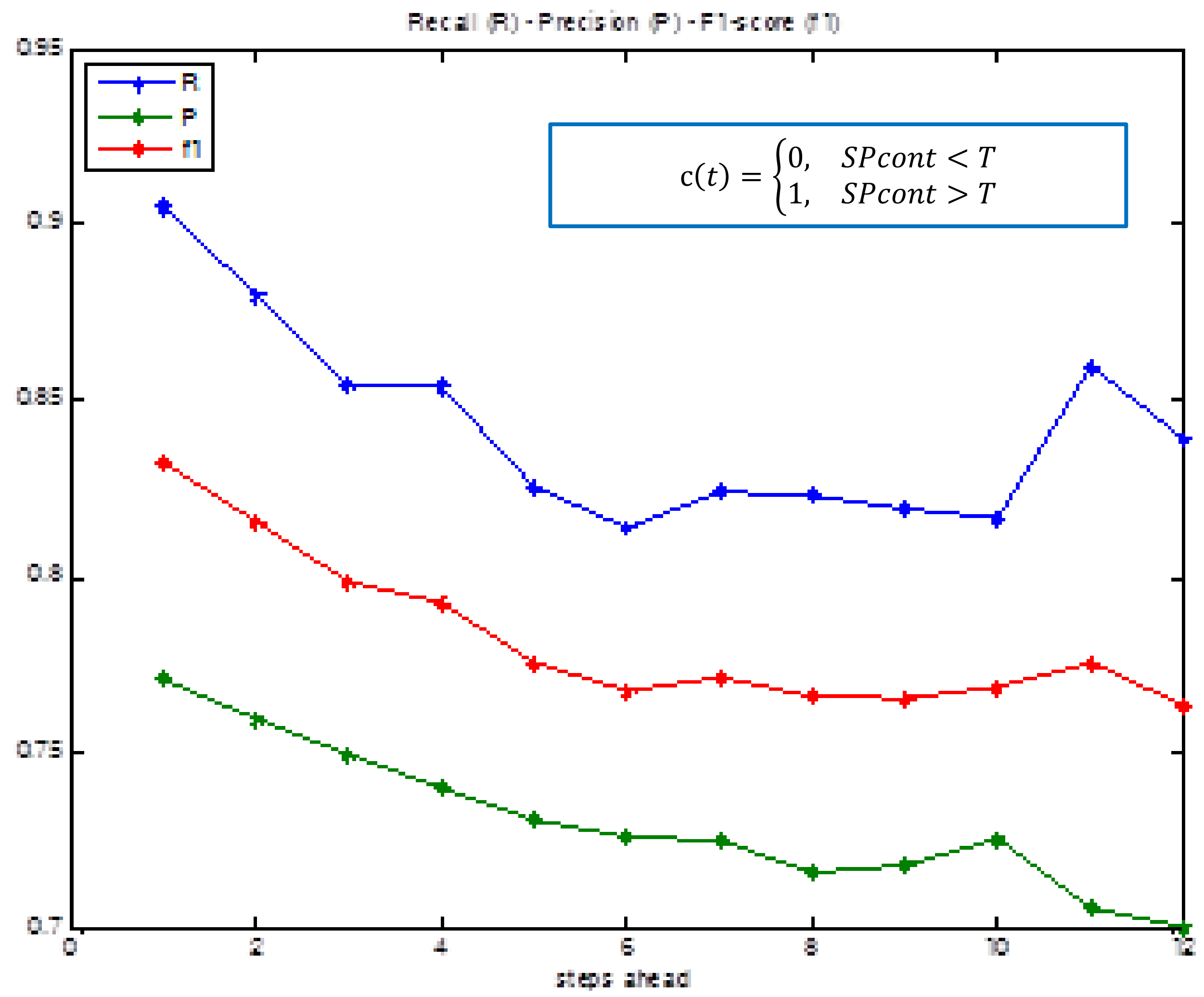
$$[max(SPcont(t-P-s+1:t-s)), min(IL(t-P-s+1:t-s)), max(IU(t-P-s+1:t-s)), max(GNSS(t-P-s+1:t-s)), LT(t)] \leftrightarrow c(t)$$

Remarks:

The results of **the FNN classifier** are more accurate than in the case where **only IL, IU, GNSS, LT** (Scenario 2) past values are exclusively considered as inputs.

The results of **the FNN classifier** are less accurate than in the case where **only the** $\boldsymbol{SPcont}$ past values are considered (Scenario 1) (however, in the latter case, no missing values were considered).

The **block-RNN** classifier exhibits higher performance compared to the **FNN** one (this may be due to the different way the input information is treated in the two cases).



WP2 Results

Developing the ST-HA ML models: LSTID indices-based forecasting

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

$$c(t) = \begin{cases} 0, & SPcont < T \\ 1, & SPcont > T \end{cases}$$

Data set creation

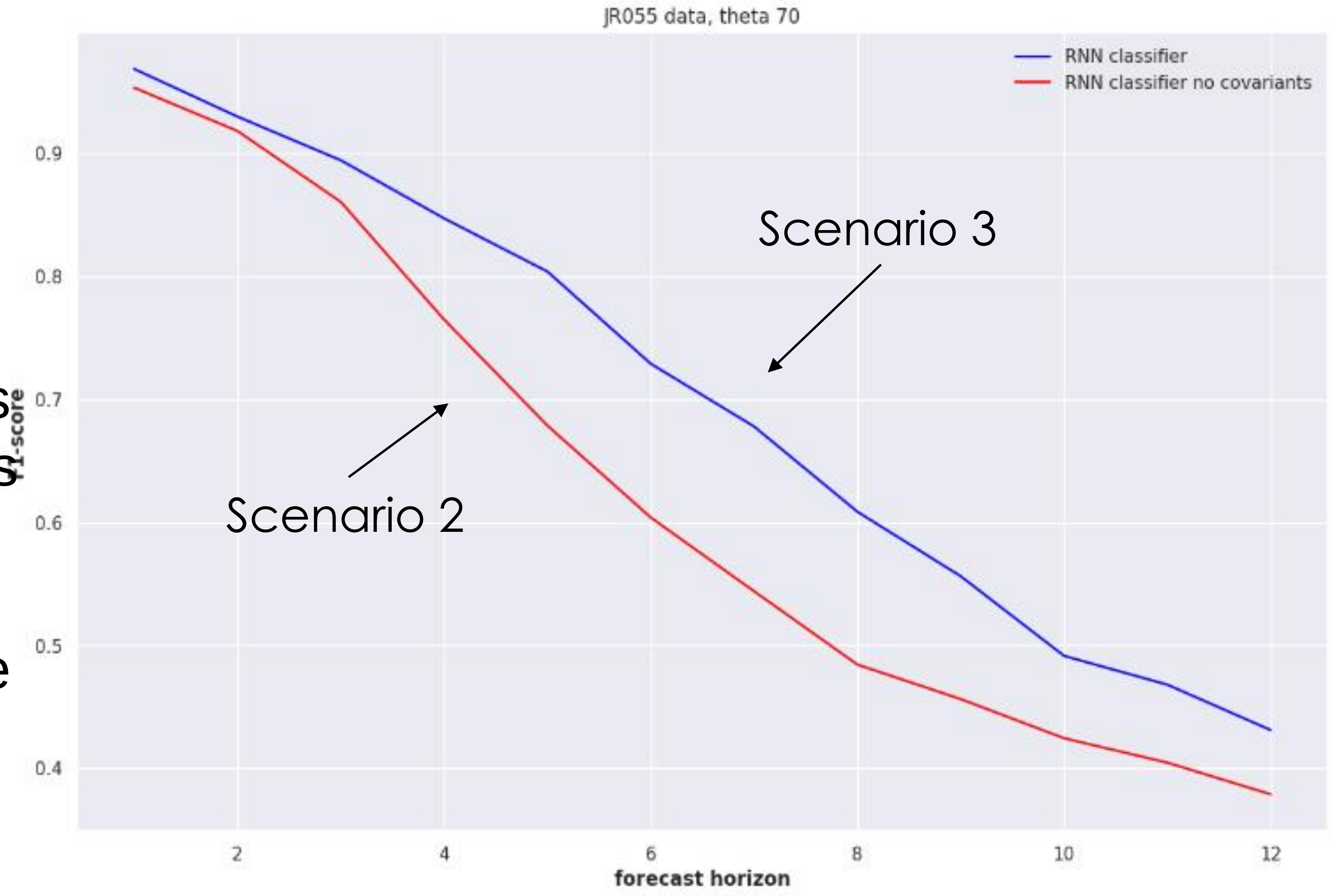
RNN classifier ($5 \cdot P$-dimensional input ⟷1-dim. output)

$$[SPcont(t-P-s+1), \ldots, SPcont(t-s), IL(t-P-s+1), \ldots, IL(t-s), IU(t-P-s+1), \ldots, IU(t-s), GNSS(t-P-s+1), \ldots, GNSS(t-s), LT(t)] \leftrightarrow c(t)$$

$SPcont$ and $GNSS$ missing values have been interpolated

Remarks:

- When interpolated $SPcont$ values are considered, $IL, IU, GNSS, LT$ drivers boost classification performance (>80% times)
- The results of the **RNN** classifier are more accurate than in the cases where **only** $IL, IU, GNSS, LT$ past values are exclusively considered as inputs (**Scenario 2**).
- The results of the **RNN classifiers** are less accurate than in the case where **only the** $SPcont$ past values are considered (**Scenario 1**) (however, in the latter case, no missing values were considered).



WP2 Results

Remarks and way forward

Based on the user needs collected by WP1, we focused on the ST-HA model development.

Catalogue-based forecasting:

- Add new input features
- Investigate other catalogue-based model time-delays\input time window
- Investigate the possibility to exploit HF EU index as a feature/label

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 higher latitude stations.

WP2 Results

Remarks and way forward

Medium term – medium accuracy:

To extend the forecasting horizon up to several hours:

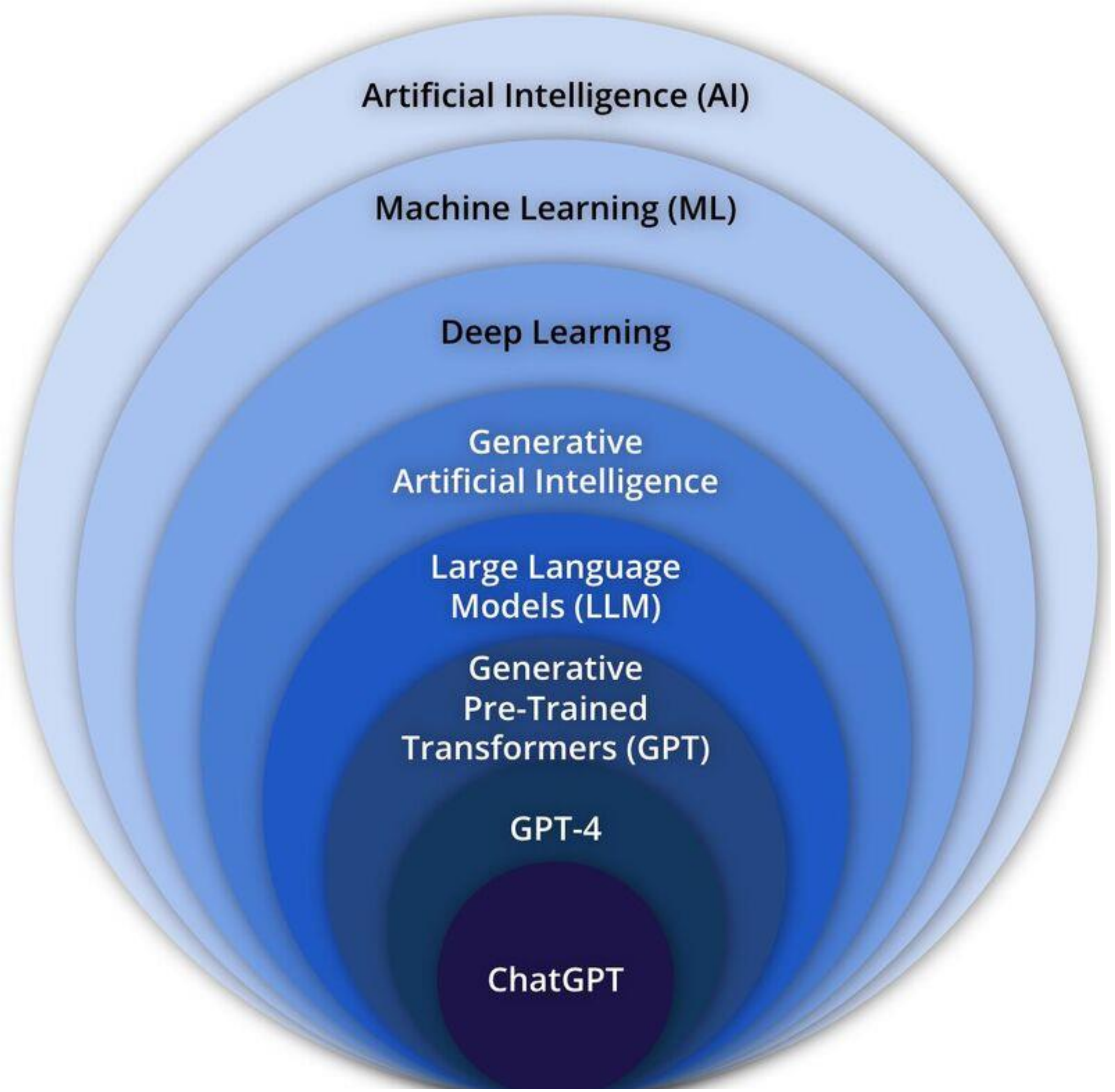
- Exploit the parameters measured at L1
- Investigate the possibility to exploit NOA model (SWIF) to relate IMF to ionospheric storm features

Long term – Low accuracy:

To extend the forecasting horizon up to one/two days:

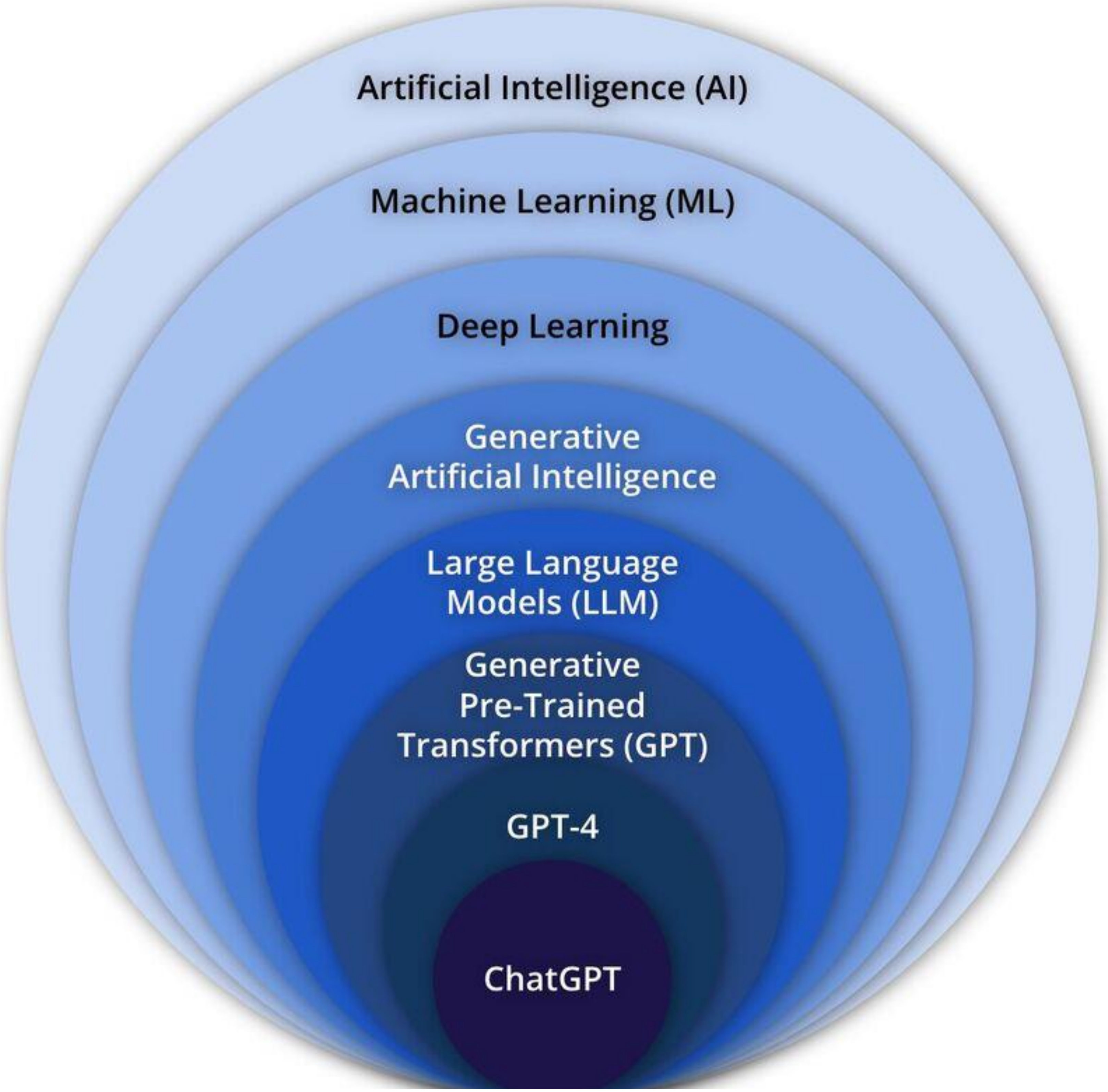
- Build a stacked model (including already existing models) starting from CME parameters provided by CACTUS to forecast SW parameters at L1 and, in turn, the geomagnetic indices at ground.

AI vs ML vs DL



Credits: E. Querci della Rovere

AI vs ML vs DL



Credits: E. Querci della Rovere