



C. Cesaroni on behalf of the WP 2 team

Innovation Day

4 December 2024

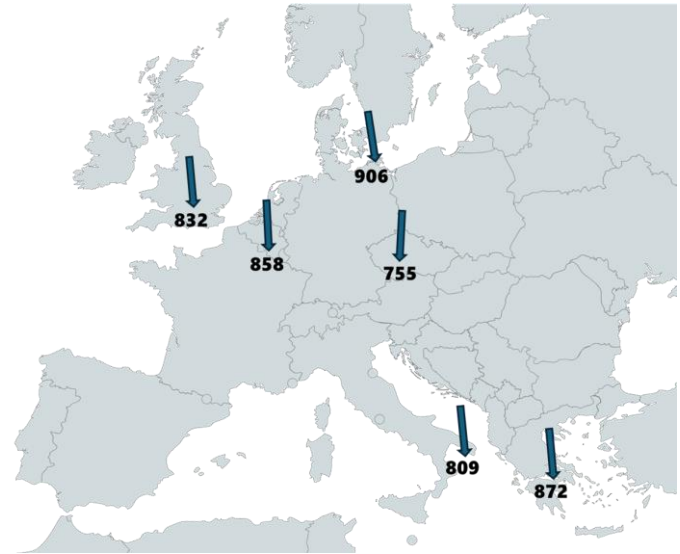
What do we have at hand?

Real-time monitoring and prediction of TIDs are very complex and a model that could form the basis for an alert system is of considerable scientific and technological interest for mitigating TID effects

Catalogue-based model

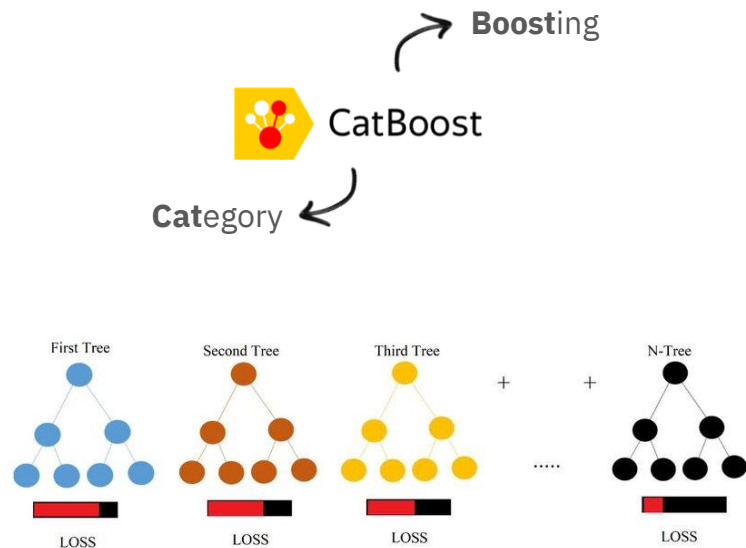
Forecasting horizon: 3hrs

European-wide forecasting



Estimated velocities (m/s) during an intense geomagnetic storm
(derived from **TechTIDE** catalogue)

Machine learning methodology – CatBoost & Optuna



Gradient boosting algorithm on decision trees

Categorical variables and **missing values** (NaN) are natively and efficiently supported

The **symmetric-trees** architecture:

- ensures efficient implementation on CPU/GPU
- reduces inference times
- naturally prevents overfitting (regularisation)

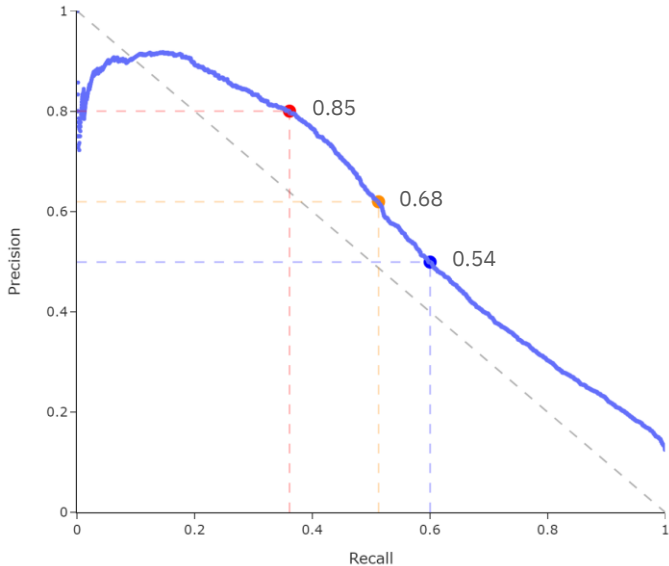
Integrates **SHAP** as a method of eXplainable Artificial Intelligence (**XAI**)

Easy to integrate with Optuna, a framework for **automatic hyperparameter optimisation**, which uses a Bayesian sampling method aborting unpromising trials



Machine learning methodology – operating modes

The **trade-off between precision and recall** is generally a **function of the user**, since the cost of a false positive is generally (very) different from that of a false negative



High precision
(high cost for FP)

TID doesn't occur	98.7%	1.3%
TID occurs	63.9%	36.1%
	TID not predicted	TID predicted



Balanced

TID doesn't occur	95.5%	4.5%
TID occurs	48.7%	51.3%
	TID not predicted	TID predicted



High sensitivity
(high cost for FN)

TID doesn't occur	91.4%	8.6%
TID occurs	40.0%	60.0%
	TID not predicted	TID predicted



Web application

T-FORS | Forecasting LSTIDs with AI


v0.2.3

Large Scale Travelling Ionospheric Disturbances (LSTIDs) are a type of space weather disturbance that could compromise the performance of critical space and ground infrastructure. The EU-funded [T-FORS project](#) is developing models that could aid in issuing forecasts and warnings for such events several hours ahead. Machine learning algorithms are used to forecast the occurrence and propagation of LSTIDs.



What can I find here?

The analysed data range from geomagnetic indices to sensor data from ionosondes scattered across the European continent. To get an intuition of the complexity behind our task, it may be informative to consider a low-dimensionality representation of the dataset. On the first page, you can find a representation of the data according to the *Uniform Manifold Approximation and Projection (UMAP)* algorithm, performing non-linear dimensionality reduction.

 [Show me the data](#)

The developed model comes from an efficient, fast and scalable gradient-boosting on decision trees framework ([CatBoost](#)). Our problem can be framed as a **multivariate time-series binary classification**, with:

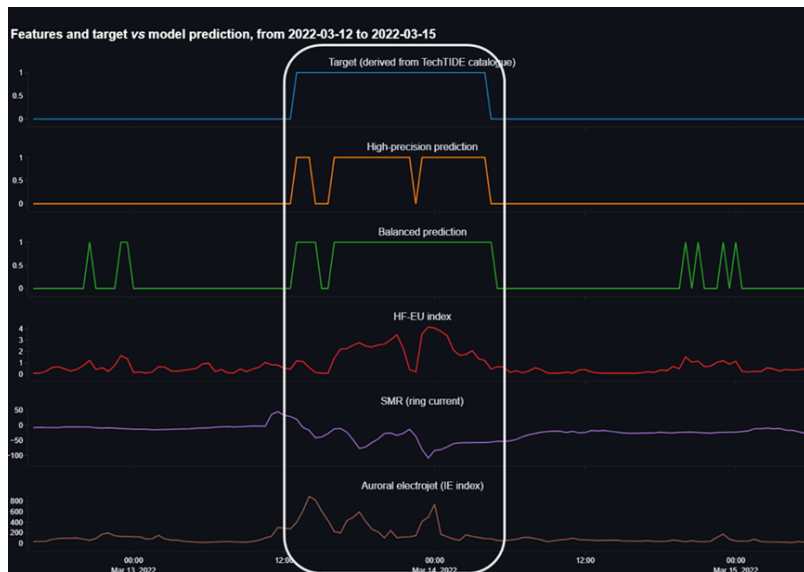
<https://t-fors-ai.streamlit.app/>

Selected events from validation phase

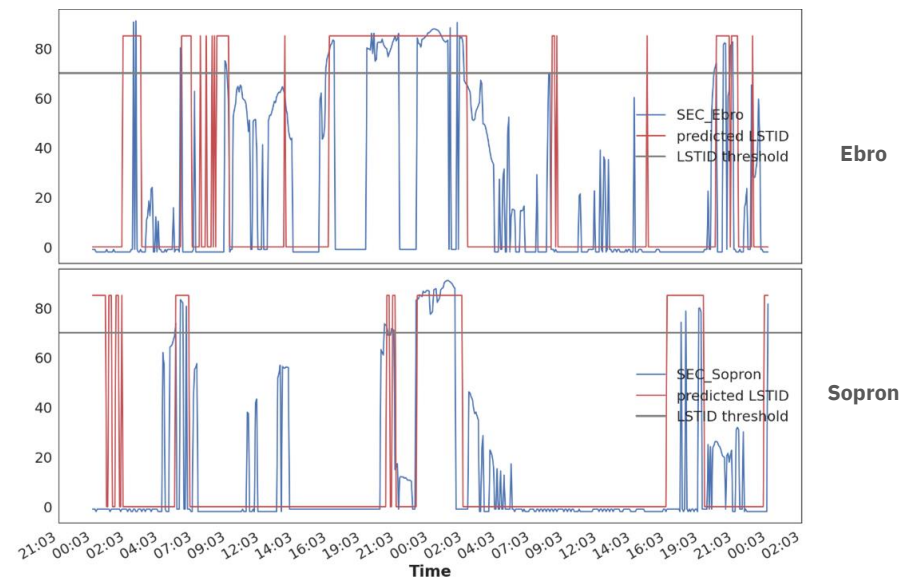
Event	Catalogue	Start (approx.)	End (approx.)	Drivers	Forecast (CB SB)
1	Yes	2022/03/13 16:05 2022/03/13 23:00	2022/03/13 21:40 2022/03/14 04:30	Yes Yes	Yes Yes Yes Yes
2	Yes	2022/07/01 23:10	2022/07/02 02:30	Yes	Yes Yes
3	No	2022/02/03 16:30	2022/02/04 00:00	Yes	Yes Yes
4	Yes	2022/07/27 03:00	2022/07/27 04:35	No	No Yes

Event 1 – catalogue- and SEC-based models

Event	Catalogue	Start (approx.)	End (approx.)	Drivers	Forecast (CB SB)
1	Yes	2022/03/13 16:05 2022/03/13 23:00	2022/03/13 21:40 2022/03/14 04:30	Yes Yes	Yes Yes Yes Yes



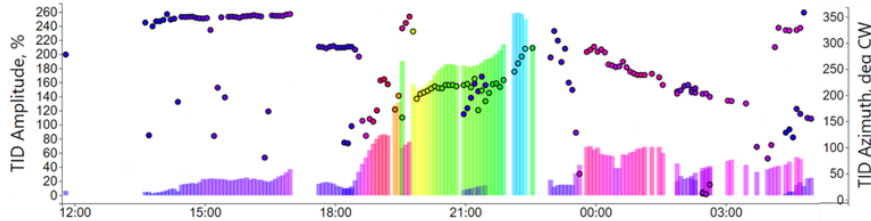
Catalogue-based



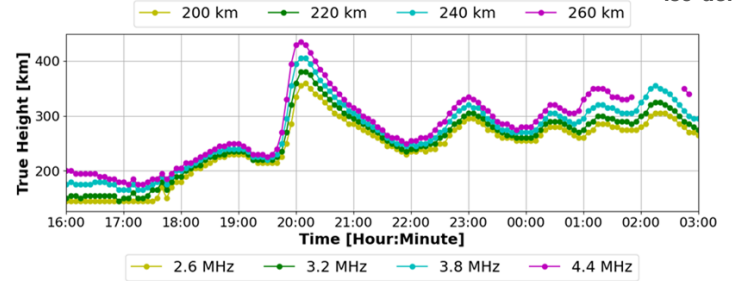
SEC-based

Event 1 – HF-INT, HF-TID, iso-density and detrended TEC

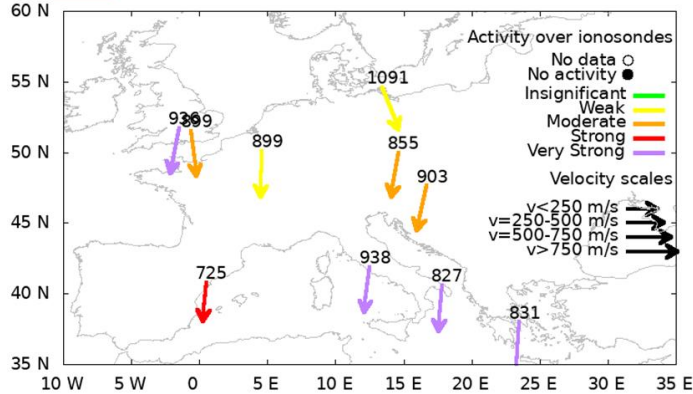
HF-TID Pruhonice-Juliusruh



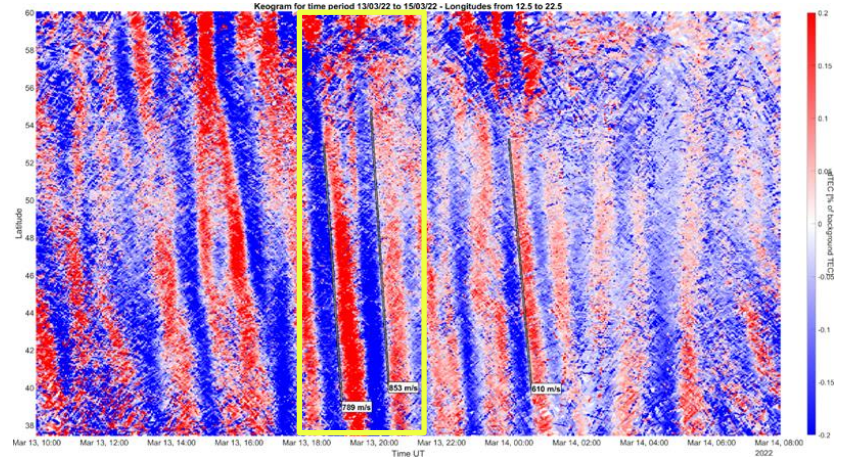
iso-density



Global Index: TID Vector velocities on 2022-03-13 at 21:25 UT



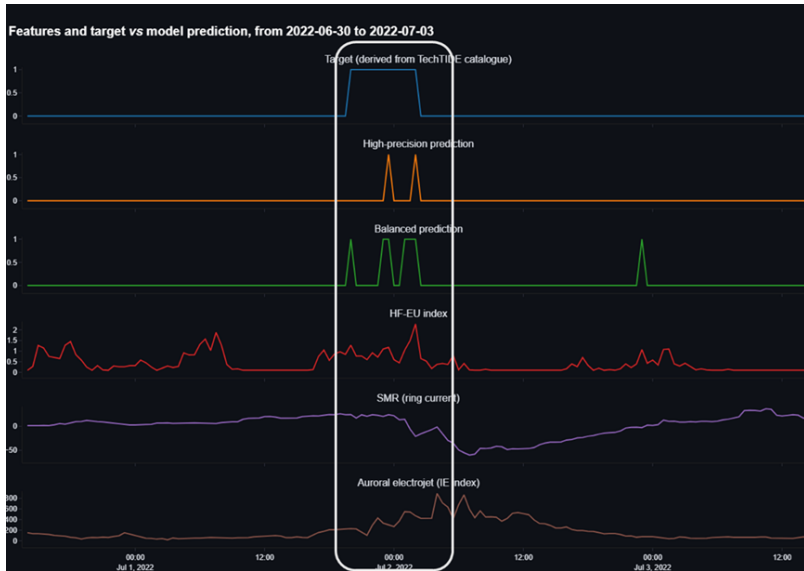
HF-INT



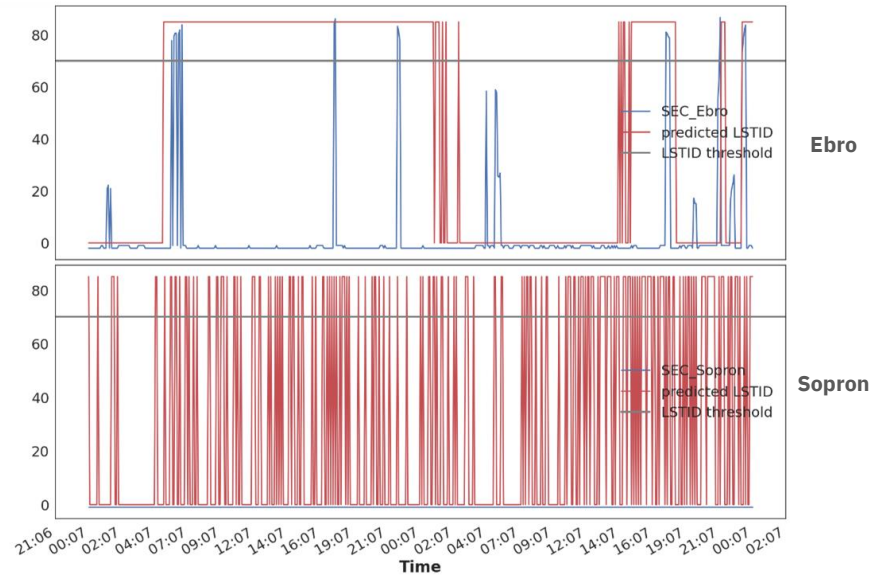
dTEC

Event 2 – catalogue- and SEC-based models

Event	Catalogue	Start (approx.)	End (approx.)	Drivers	Forecast (CB SB)
2	Yes	2022/07/01 23:10	2022/07/02 02:30	Yes	Yes Yes



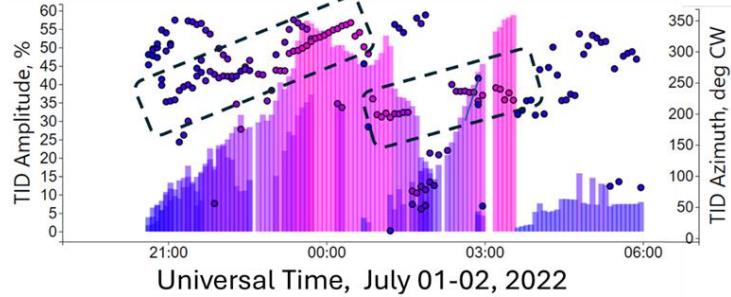
Catalogue-based



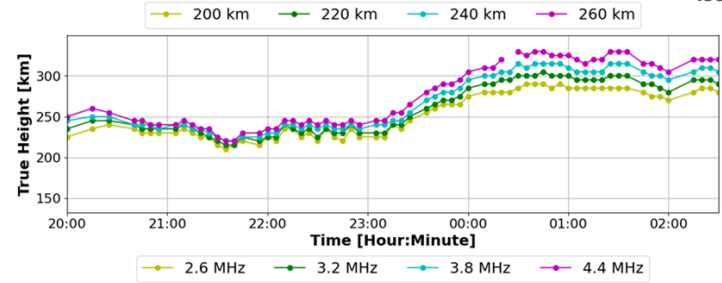
SEC-based

Event 2 – HF-INT, HF-TID, iso-density and detrended TEC

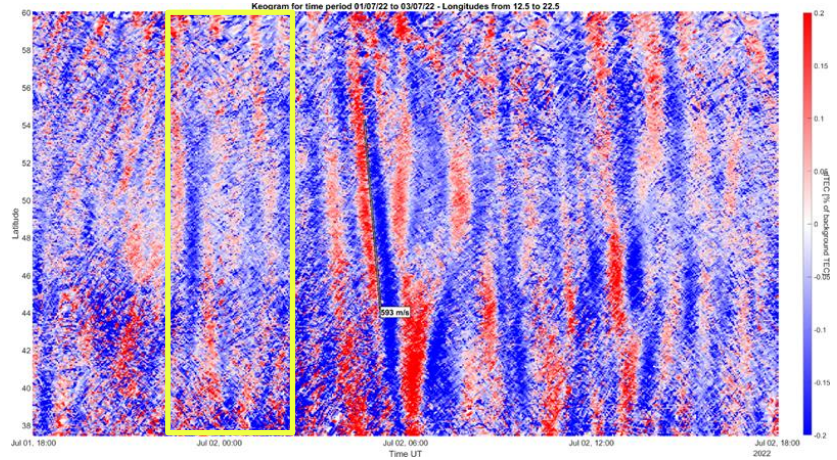
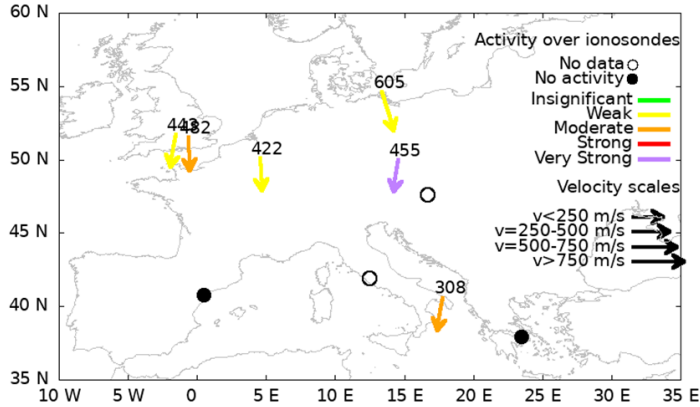
HF-TID Pruhonice-Juliusruh



iso-density



Global Index: UNCERTAIN Vector velocities on 2022-07-01 at 23:45 UT

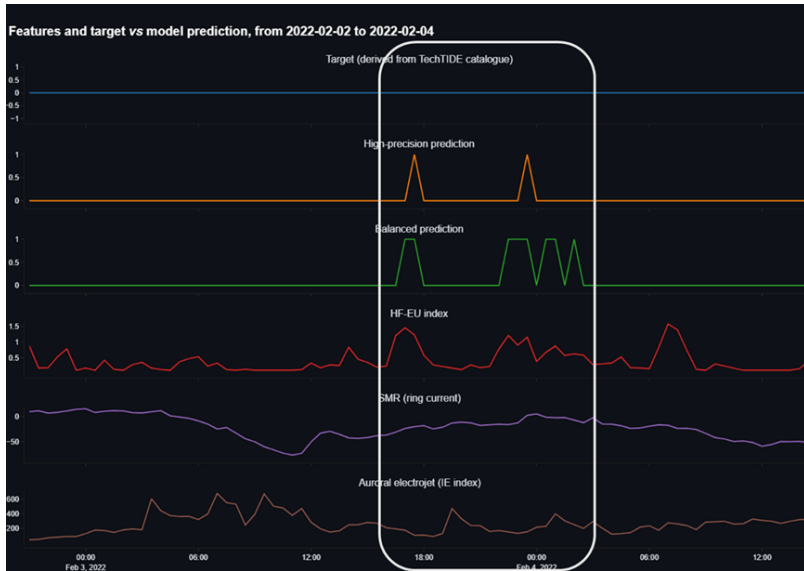


HF-INT

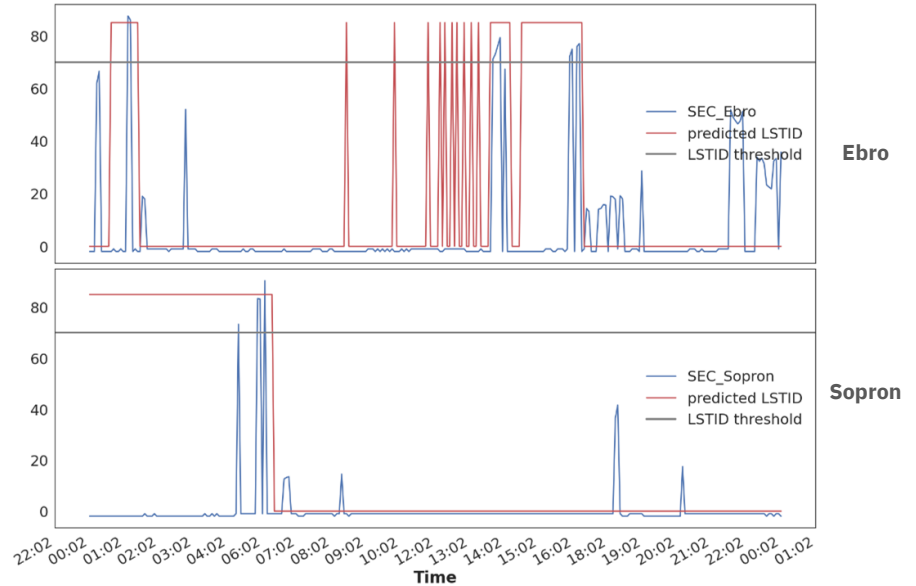
dTEC

Event 3 – catalogue- and SEC-based models

Event	Catalogue	Start (approx.)	End (approx.)	Drivers	Forecast (CB SB)
3	No	2022/02/03 16:30	2022/02/04 00:00	Yes	Yes Yes



Catalogue-based

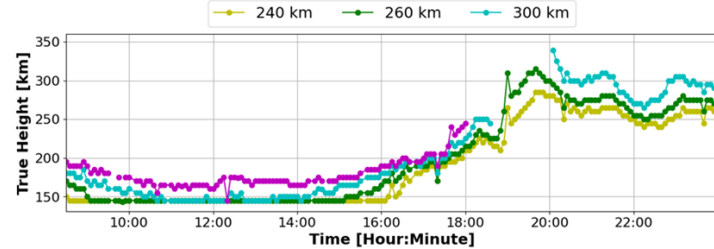


SEC-based

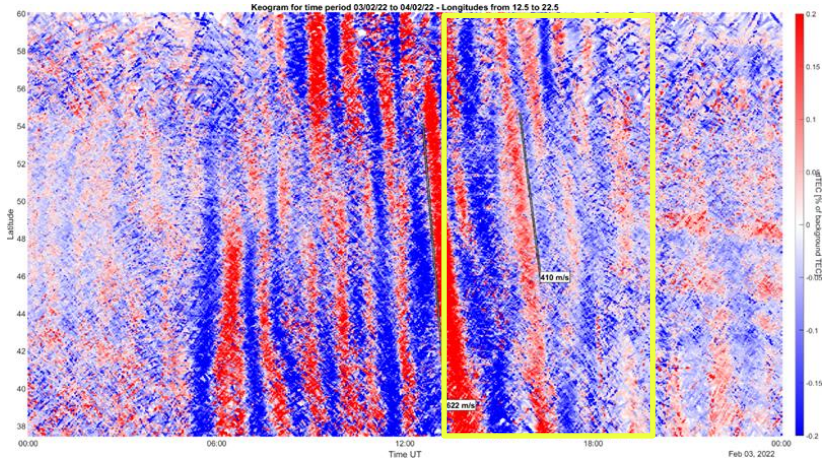
Event 3 – HF-INT, HF-TID, iso-density and detrended TEC

The HF-TID method reports unknown LSTID activity because no relevant data was found due to the unavailability of the Pruhonice-Juliusruh radio link

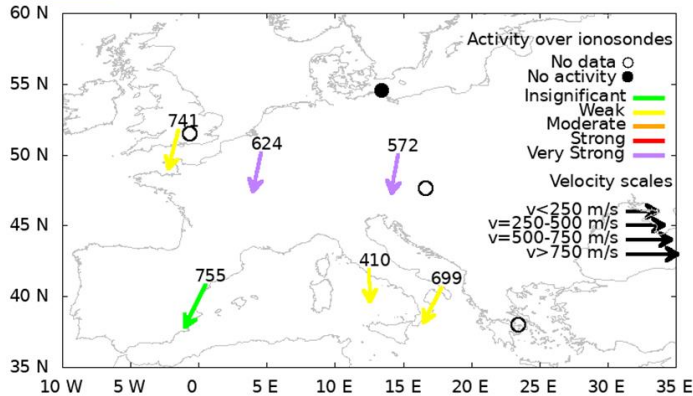
iso-density



2.6 MHz 3.2 MHz 3.8 MHz 4.4 MHz



Global Index: UNCERTAIN Vector velocities on 2022-02-03 at 17:20 UT

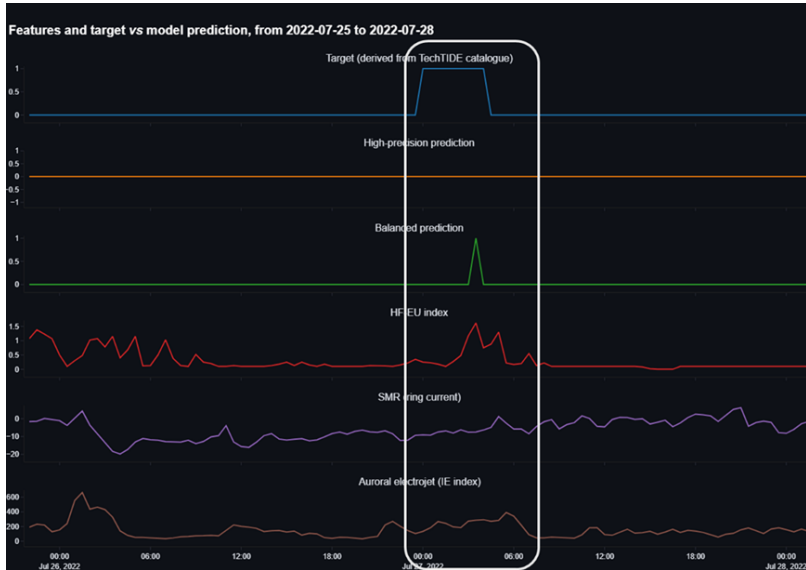


HF-INT

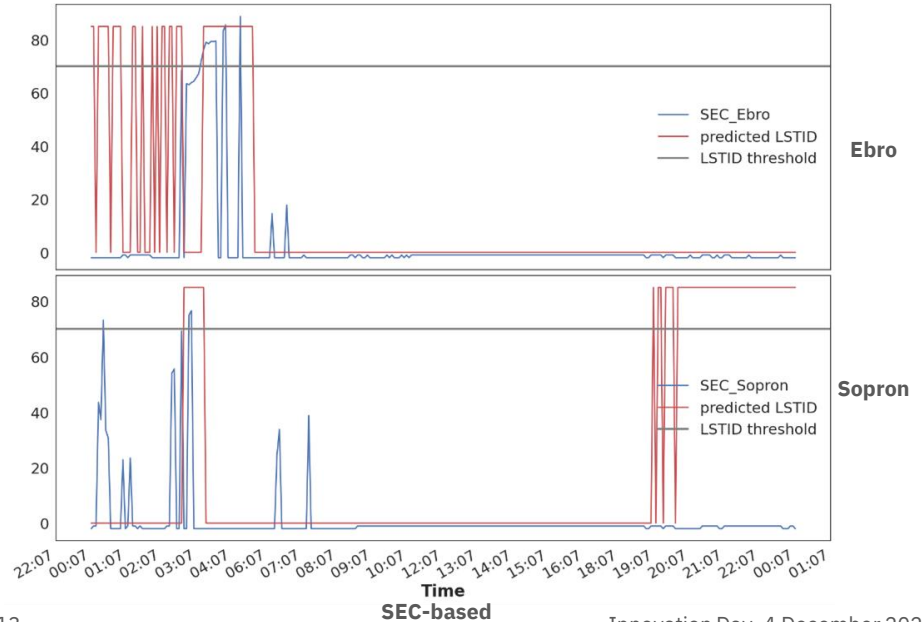
dTEC

Event 4 – catalogue- and SEC-based models

Event	Catalogue	Start (approx.)	End (approx.)	Drivers	Forecast (CB SB)
4	Yes	2022/07/27 03:00	2022/07/27 04:35	No	No Yes

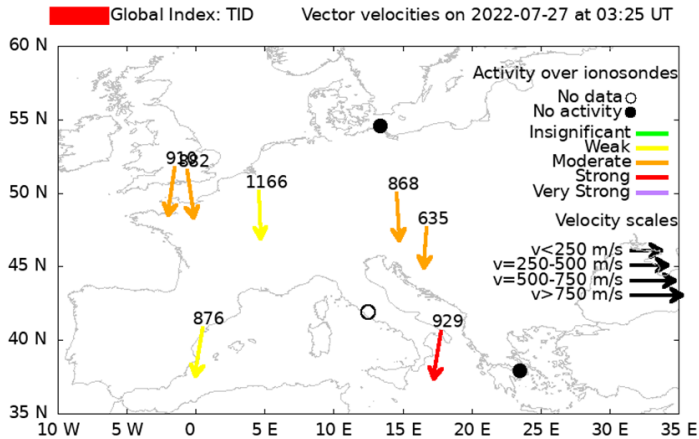


Catalogue-based

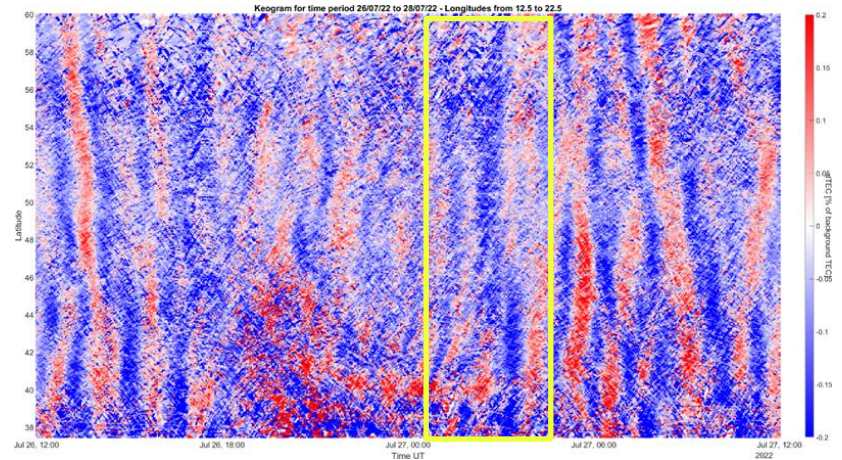
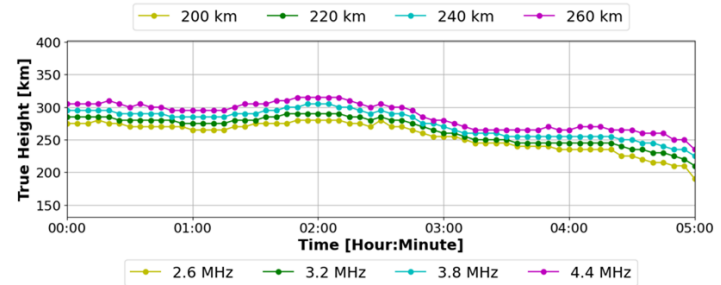


Event 4 – HF-INT, HF-TID, iso-density and detrended TEC

The HF-TID method reports unclear LSTID activity because no relevant data was found due to the unavailability of the Pruhonice-Juliusruh radio link due to weak/missing reception of the D2D, with a complicated mode structure due to the presence of a blanketing Es layer



HF-INT



dTEC

Summary and future steps (beyond T-FORS lifetime)

- 2 different AI-based models are available in T-FORS for LSTID forecasting ([CB model](#), [SB model](#))
- Catalogue based model gives a forecasting at European level 3 hours in advance
- SEC based model gives a forecasting at ionosonde level up to 2 hours in advance
- A first version of the demonstrator is available
- Validation of the models have been performed on a statistical and case event level (report under evaluation)