P3.5-643

Evaluation of automatic infrasound signal classification via Machine Learning deployed at the Central and Eastern European Infrasound Network

Marcell Pásztor[1,2], Tereza Sindelarova[3], Daniela Ghica[4], Ulrike Mitterbauer[5], Oleksandr Liashchuk[6], István Bondár[7]

1 Department of Geophysics and Space Science, Institute of Geography and Earth Sciences, ELTE Eötvös Loránd University 2 Kövesligethy Radó Seismological Observatory, HUN-REN Institute of Earth Physics and Space Science 3 The Czech Academy of Sciences, Institute of Atmospheric Physics 4 National Institute for Earth Physics 5 GeoSphere Austria 6 Main Centre of Special Monitoring, National Center for Control and Testing of Space Facilities, State Space Agency of Ukraine 7 Seismic Location Services

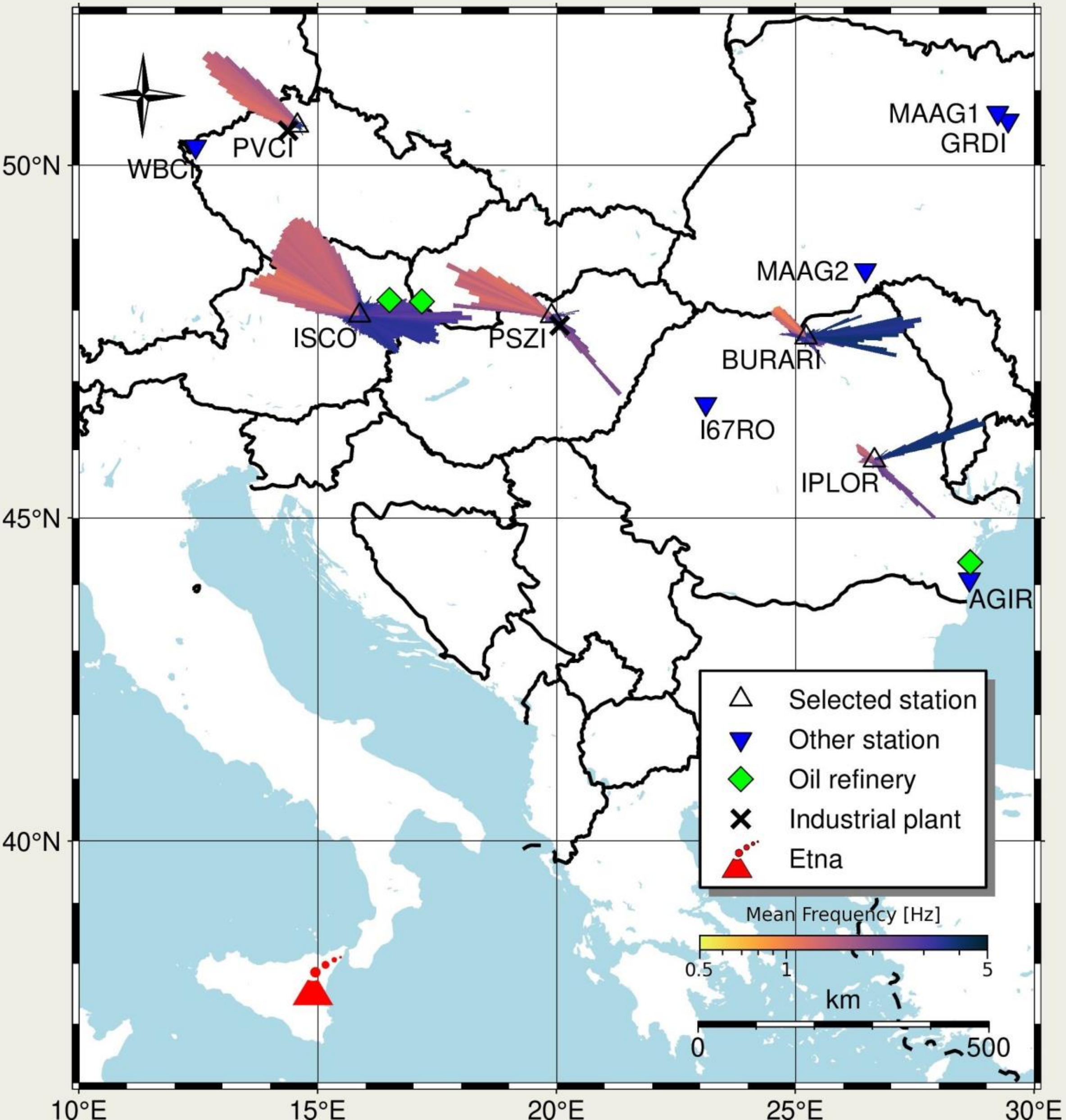
INTRODUCTION AND MAIN RESULTS

Using the Hungarian Infrasound array (PSZI) a Machine Learning based approach is presented to automatically classify infrasound detections. Detections are produced by the Progressive Multi Channel Correlation (PMCC) method. An ensemble model has been used for more than six months which consists of a Random Forest model trained on PMCC related features and a Convolutional Neural Network trained on spectrograms. Currently, the performance of the automatic monitoring scheme is under evaluation. So far, it has shown great results toward distinguishing between signals of interest (e.g., Etna eruptions, quarry blasts) and usage to remove detections that can be considered as noise, such as signals from oil refineries.

Evaluation of automatic infrasound signal classification via Machine Learning deployed at the Central and Eastern European Infrasound Network

Marcell Pásztor, Tereza Sindelarova, Daniela Ghica, Ulrike Mitterbauer, Oleksandr Liashchuk, István Bondár

P3.5-643



The Central and Eastern European Infrasound Network CEEIN

The Central and Eastern European Infrasound Network (CEEIN) was established in 2018 through a collaboration of research institutes from Hungary, Austria, the Czech Republic, and Romania. In 2019, Ukraine joined the network, further extending its reach. Within CEEIN, researchers work together to identify shared infrasound sources both within and beyond the network, carry out seismoacoustic investigations, and distinguish between natural and human-made events. The recorded events—such as quarry blasts, sonic booms, and bolides—are reported in the biannual CEEIN bulletin (Bondár et al., 2022, please see P2.4-886 at SnT 2025). Bondár et al. (2022) demonstrated that CEEIN significantly improves infrasound detection capabilities across Europe. CEEIN data can be accessed on the website below.

Motivation

The ever-growing data flow makes it not only possible but also desirable to automatically filter detection lists, particularly since PMCC can generate a large number of detections. Machine learning applications have shown great potential for classifying infrasound signals (e.g., Albert & Linville, 2020; Witsil et al., 2022; Bishop et al., 2022; Pásztor et al., 2023, 2025).

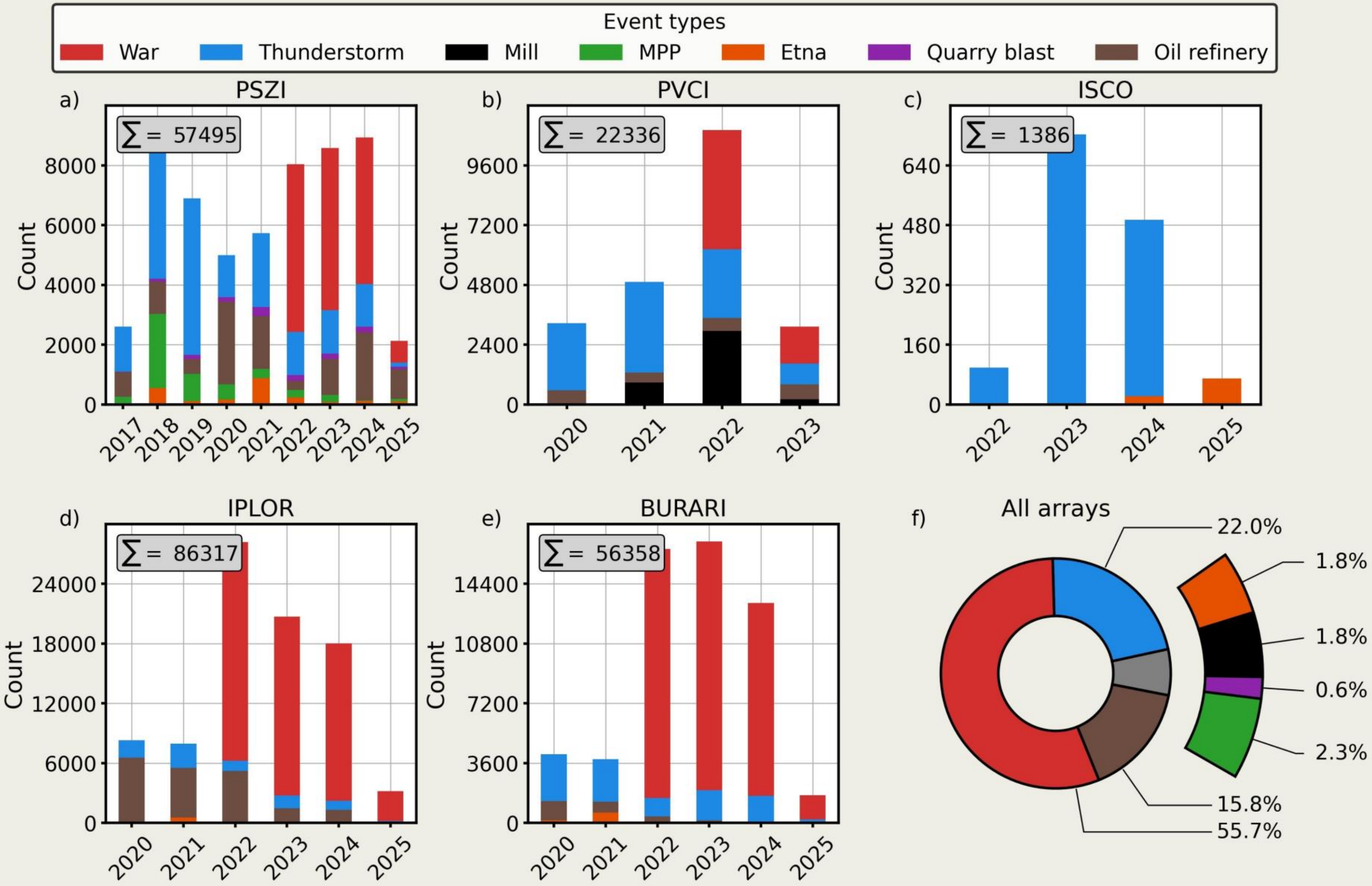
Here, we present the results from the first six months of deploying such an automated scheme, which is capable of distinguishing among quarry blasts, eruptions of Mount Etna, thunderstorms, oil refineries, a power plant, and signals related to the war in Ukraine.

CEEIN website

Bondár et al., 2022

Evaluation of automatic infrasound signal classification via Machine Learning deployed at the Central and Eastern European Infrasound Network

Marcell Pásztor, Tereza Sindelarova, Daniela Ghica, Ulrike Mitterbauer, Oleksandr Liashchuk, István Bondár



The Dataset

Over the past few years, we have been developing a large labeled infrasound dataset that incorporates five CEEIN arrays (PVCI, ISCO, PSZI, IPLOR, and BURARI). The dataset currently contains approximately 250,000 instances and is regularly updated.

For labeling, infrasound detections were matched with ground-truth data from various sources—for example, seismic data for quarry blasts and lightning distribution maps for thunderstorms. In addition, ray-tracing and cross-bearing methods were applied for confirmation. The full process is described in Pásztor et al. (2025). The current classification scheme includes the following categories:

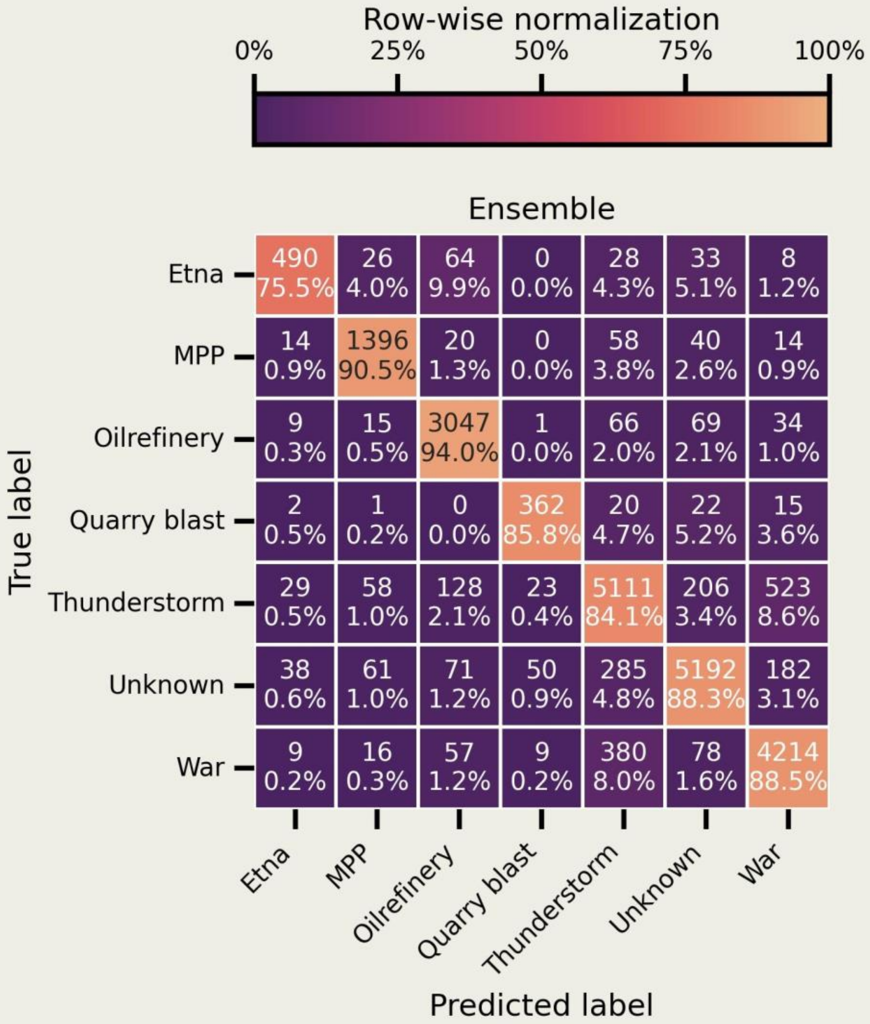
- Quarry blast
- Etna activity
- Thunderstorm
- Oil refinery
- Mátra Power Plant (MPP)
- Mill and wood processing plant
- War in Ukraine

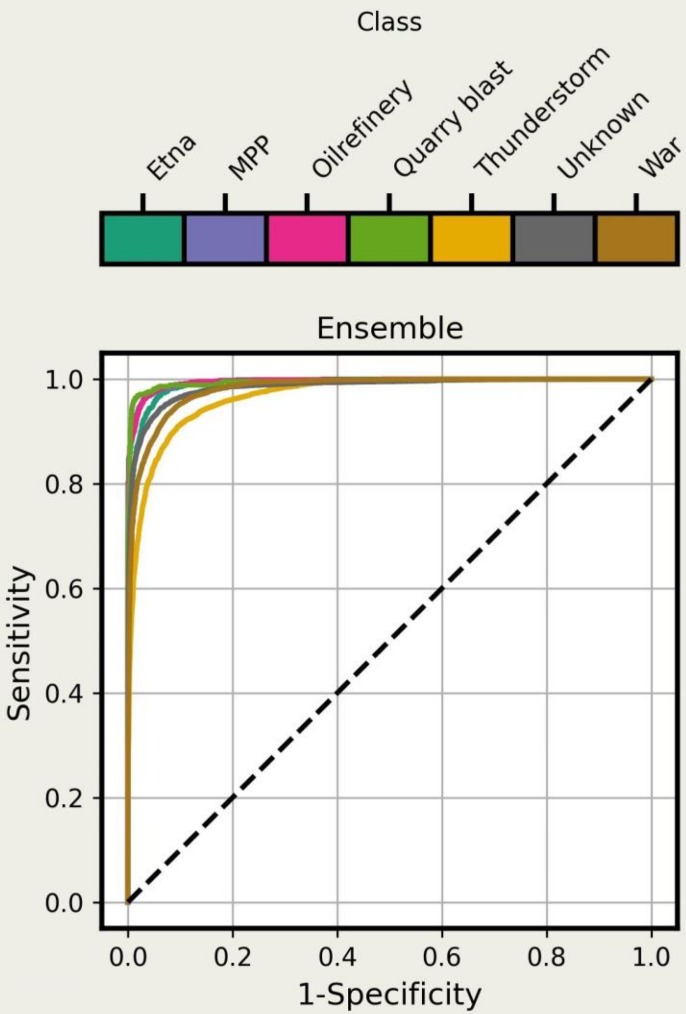
The dataset is considered consistent, as the PMCC settings were not changed and the labeling method was applied uniformly across all arrays. Nevertheless, some limitations remain, such as class imbalance and differences between arrays (aperture, instruments etc.).

Evaluation of automatic infrasound signal classification via Machine Learning deployed at the Central and Eastern European Infrasound Network

Marcell Pásztor, Tereza Sindelarova, Daniela Ghica, Ulrike Mitterbauer, Oleksandr Liashchuk, István Bondár

P3.5-643





Modelling

The ensemble model is composed of two components. A Random Forest model is trained on tabular data, i.e., features extracted from PMCC along with three features that quantify the similarity between neighboring detections (see Pásztor et al., 2023 for details).

In parallel, a Convolutional Neural Network is trained on 95-second spectrograms computed from sum-and-delay beams. The ensemble model combines the probability outputs of the two models by averaging them, and the predicted class is assigned based on the argument of maxima.

In a recent study (Pásztor et al., 2025), we trained and evaluated the model using five CEEIN stations and further investigated its transportability. In the present work, however, we focus exclusively on PSZI, a four-element array in Hungary.

Pásztor et al., 2023

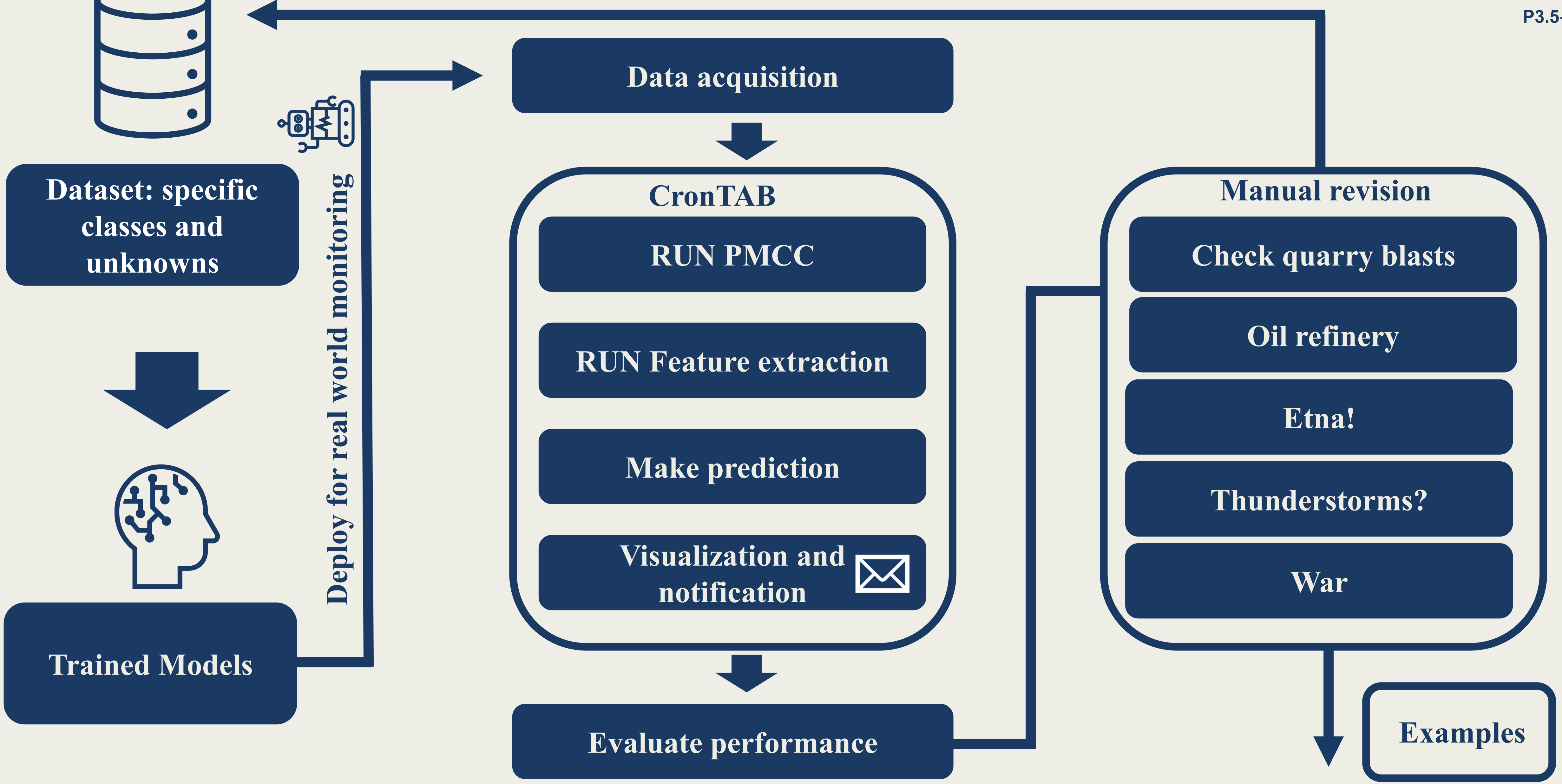
Pásztor et al., 2025

SnT 2025 CTBT: SCIENCE AND TECHNOLOGY CONFERENCE
8 SEPTEMBER ONLINE DAY
9 TO 12 SEPTEMBER AT HOFBURG PALACE, VIENNA & ONLINE

Evaluation of automatic infrasound signal classification via Machine Learning deployed at the Central and Eastern European Infrasound Network

Marcell Pásztor, Tereza Sindelarova, Daniela Ghica, Ulrike Mitterbauer, Oleksandr Liashchuk, István Bondár

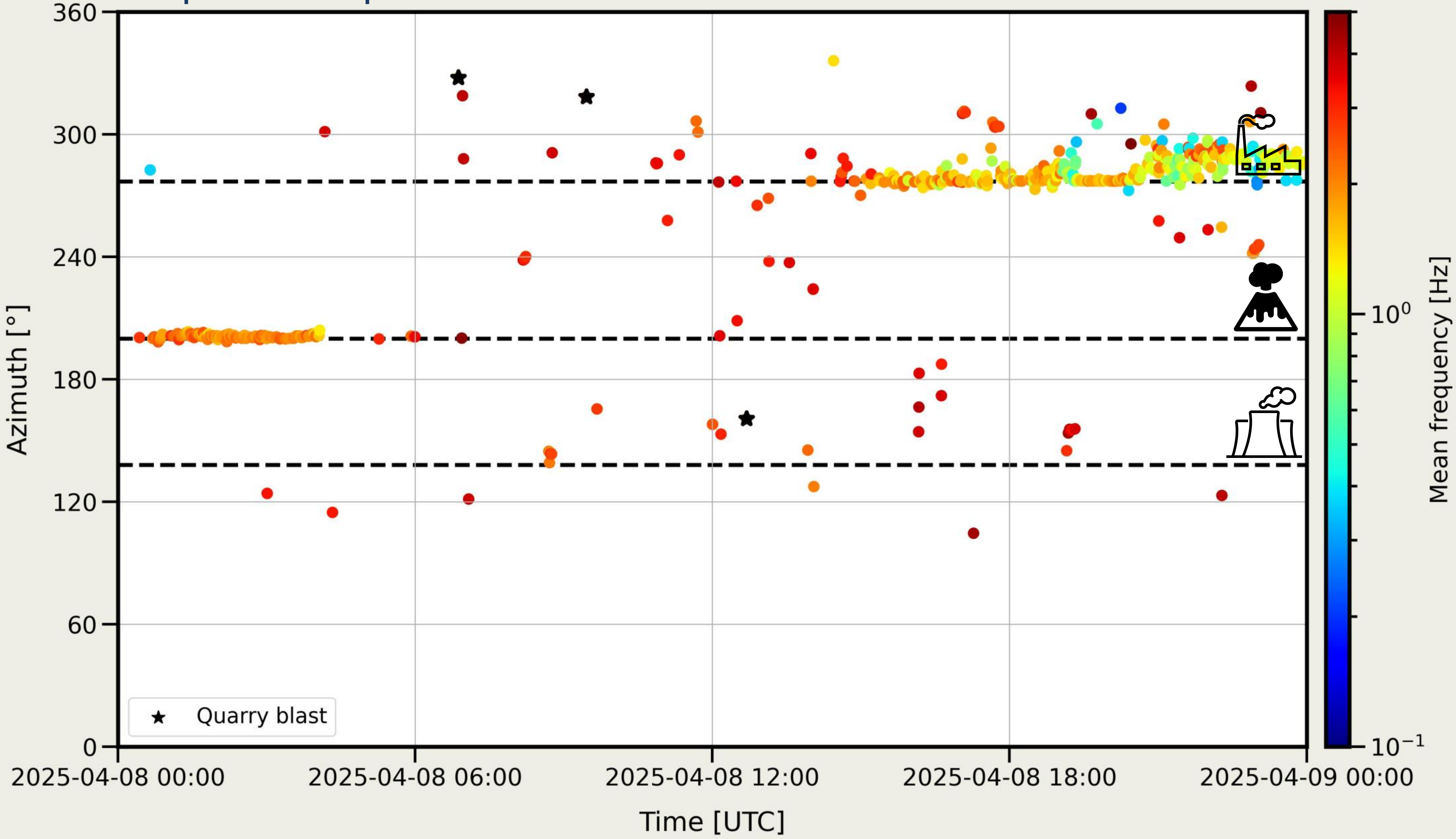
P3.5-643



Evaluation of automatic infrasound signal classification via Machine Learning deployed at the Central and Eastern European Infrasound Network

Marcell Pásztor, Tereza Sindelarova, Daniela Ghica, Ulrike Mitterbauer, Oleksandr Liashchuk, István Bondár

Example 1: Etna Eruption



Event Type	Count
Etna	50
MPP	24
Oil refinery	142
Quarry blast	3
Thunderstorm	45
Unknown	180
War	42

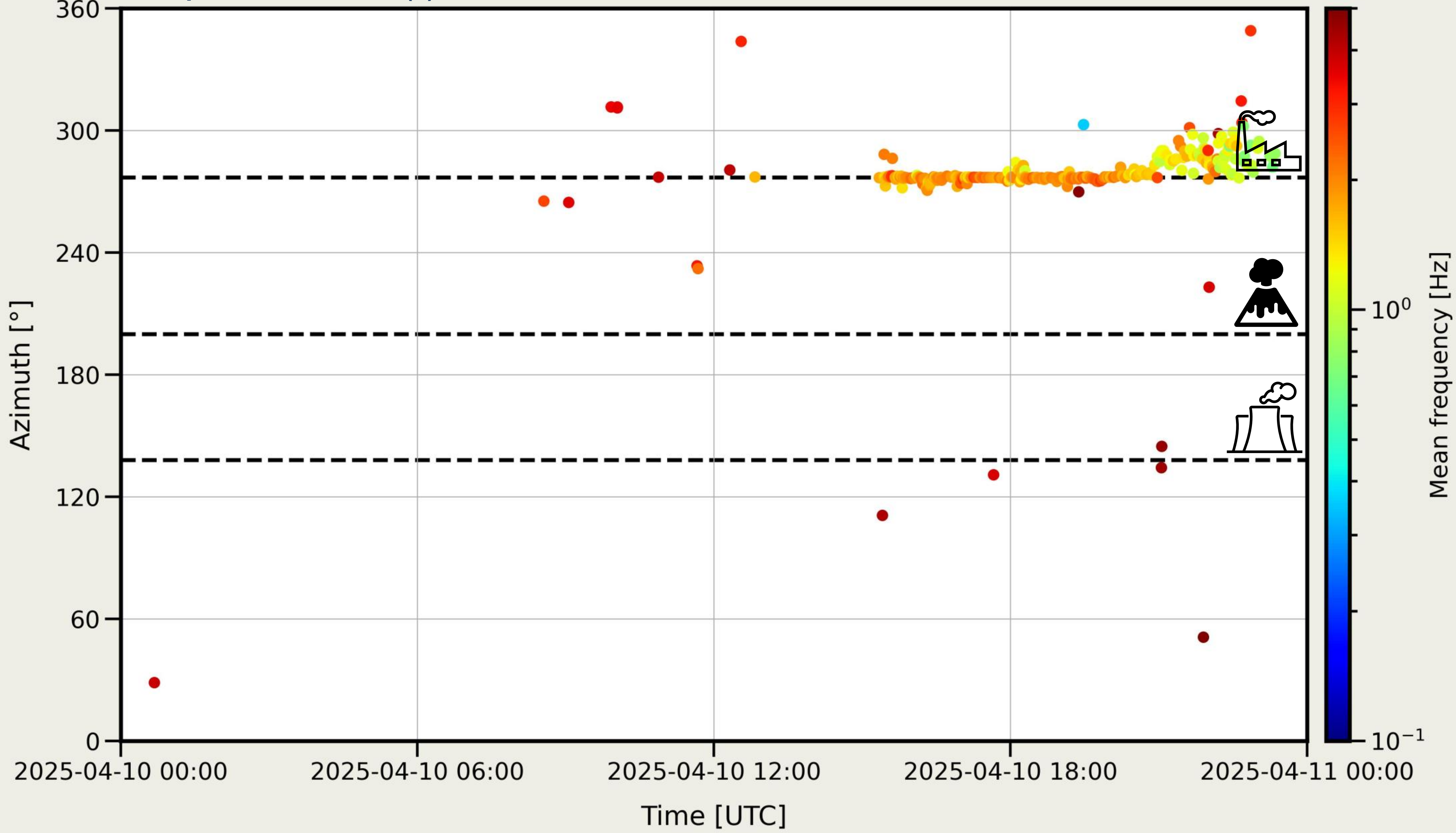
On April 8, 2025, an eruption was correctly identified by the algorithm, as shown in the figure at around 200° azimuth. A report was automatically generated based on the daily number of detections labeled as Etna.

Evaluation of automatic infrasound signal classification via Machine Learning deployed at the Central and Eastern European Infrasound Network

Marcell Pásztor, Tereza Sindelarova, Daniela Ghica, Ulrike Mitterbauer, Oleksandr Liashchuk, István Bondár

P3.5-643

Example 2: Oil refinerie(s)



Event Type	Count
Etna	15
MPP	5
Oil refinery	**134**
Quarry blast	0
Thunderstorm	33
Unknown	86
War	10

Signals from the oil refineries near Schwechat and Bratislava were correctly identified multiple times during the six-month period. An example is shown on the left. Please note that azimuth is not used as a feature, though it could be employed as an additional constraint.

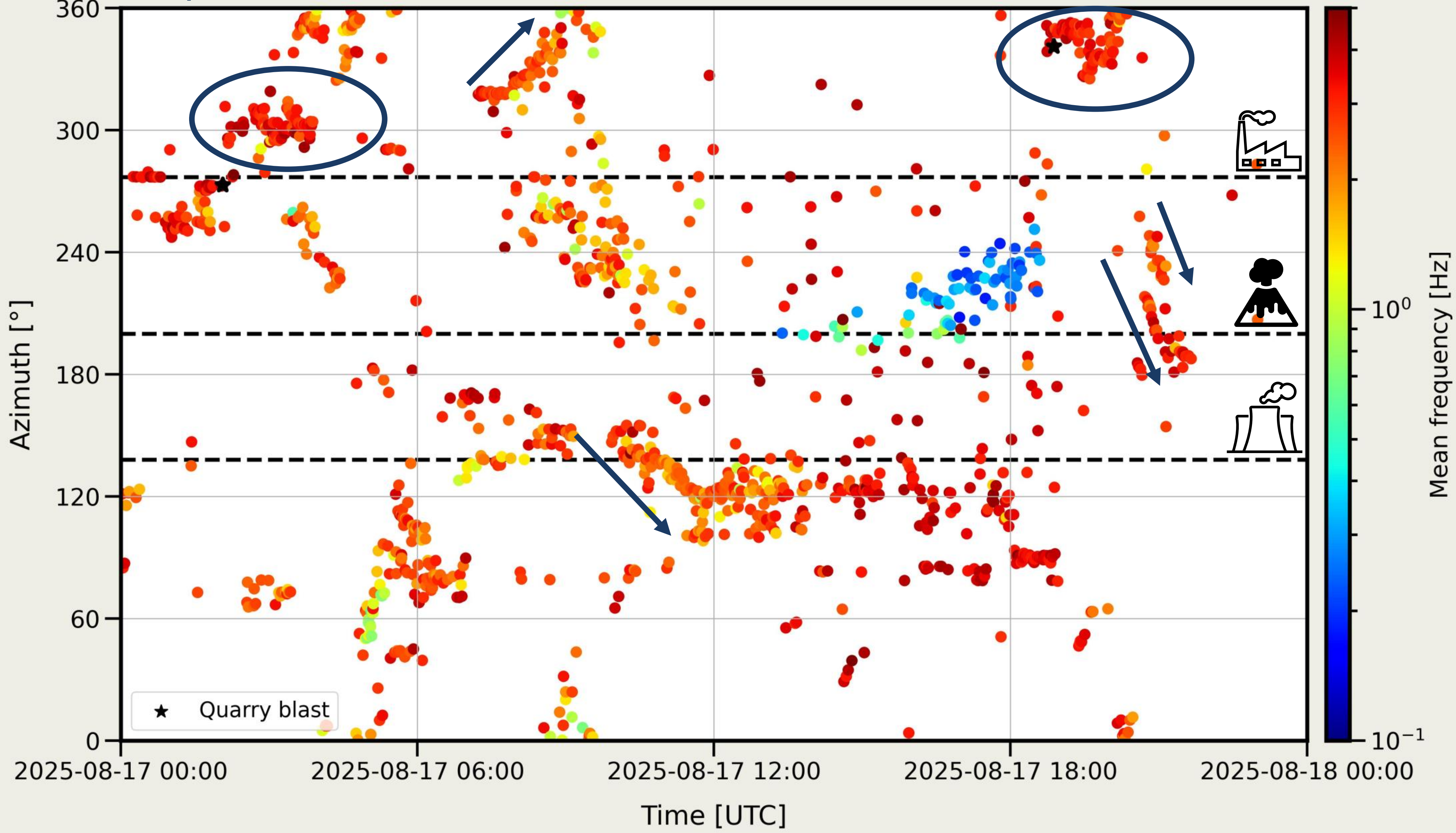
SnT 2025 8 SEPTEMBER ONLINE DAY 9 TO 12 SEPTEMBER AT HOFBURG PALACE, VIENNA & ONLINE
CTBT: SCIENCE AND TECHNOLOGY CONFERENCE

Evaluation of automatic infrasound signal classification via Machine Learning deployed at the Central and Eastern European Infrasound Network

Marcell Pásztor, Tereza Sindelarova, Daniela Ghica, Ulrike Mitterbauer, Oleksandr Liashchuk, István Bondár

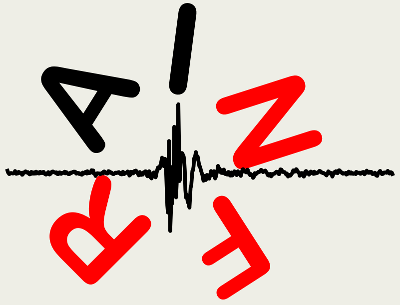
P3.5-643

Example 3: Thunderstorms



Event Type	Count
Etna	1
MPP	1
Oil refinery	7
Quarry blast	2
Thunderstorm	1037
Unknown	155
War	99

As shown in the confusion matrices and ROC curves, thunderstorms account for the highest number of misclassified detections. Here, we present a day when several thunderstorms struck Hungary.

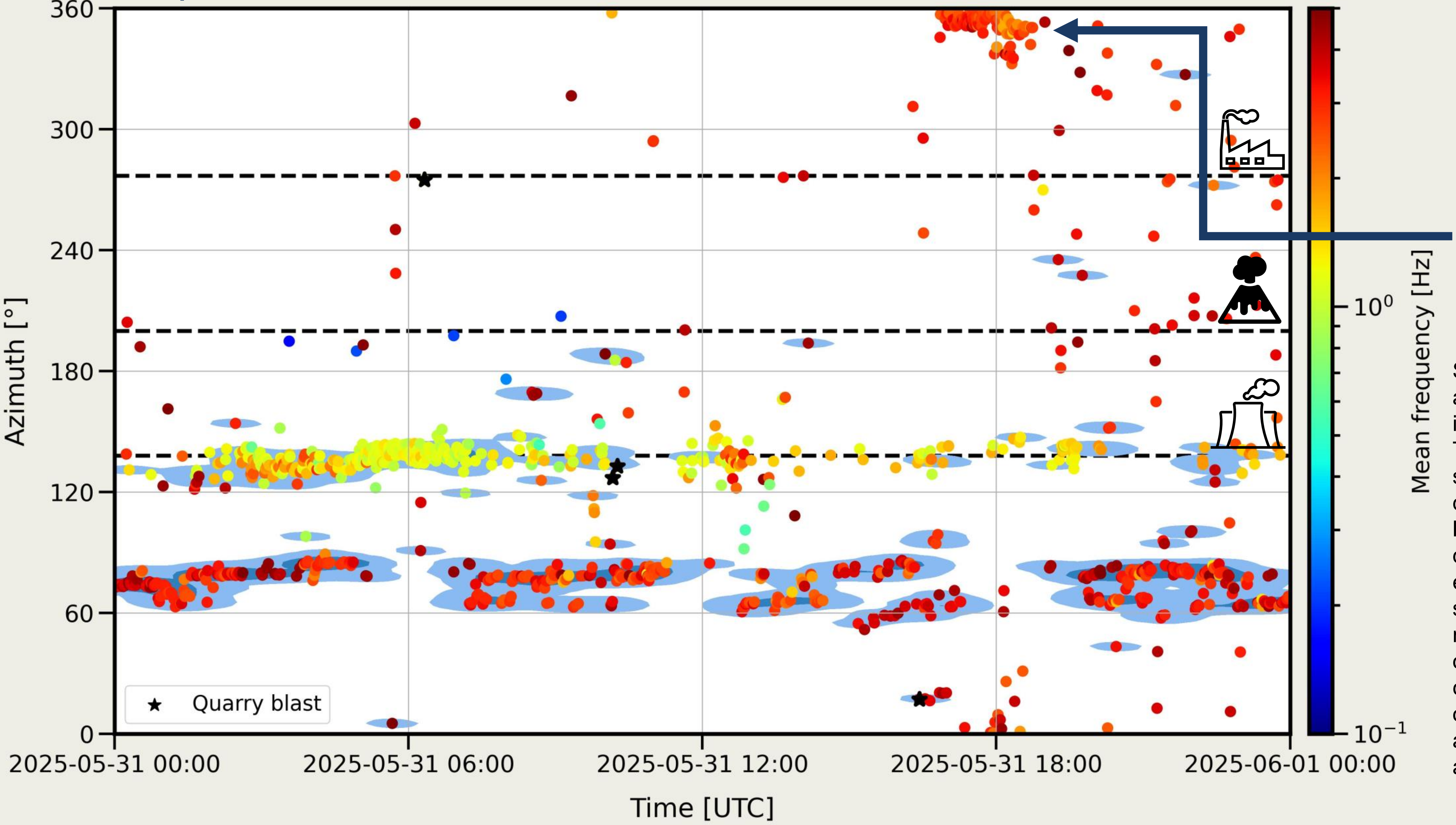


Evaluation of automatic infrasound signal classification via Machine Learning deployed at the Central and Eastern European Infrasound Network

Marcell Pásztor, Tereza Sindelarova, Daniela Ghica, Ulrike Mitterbauer, Oleksandr Liashchuk, István Bondár

P3.5-643

Example 4: War in Ukraine

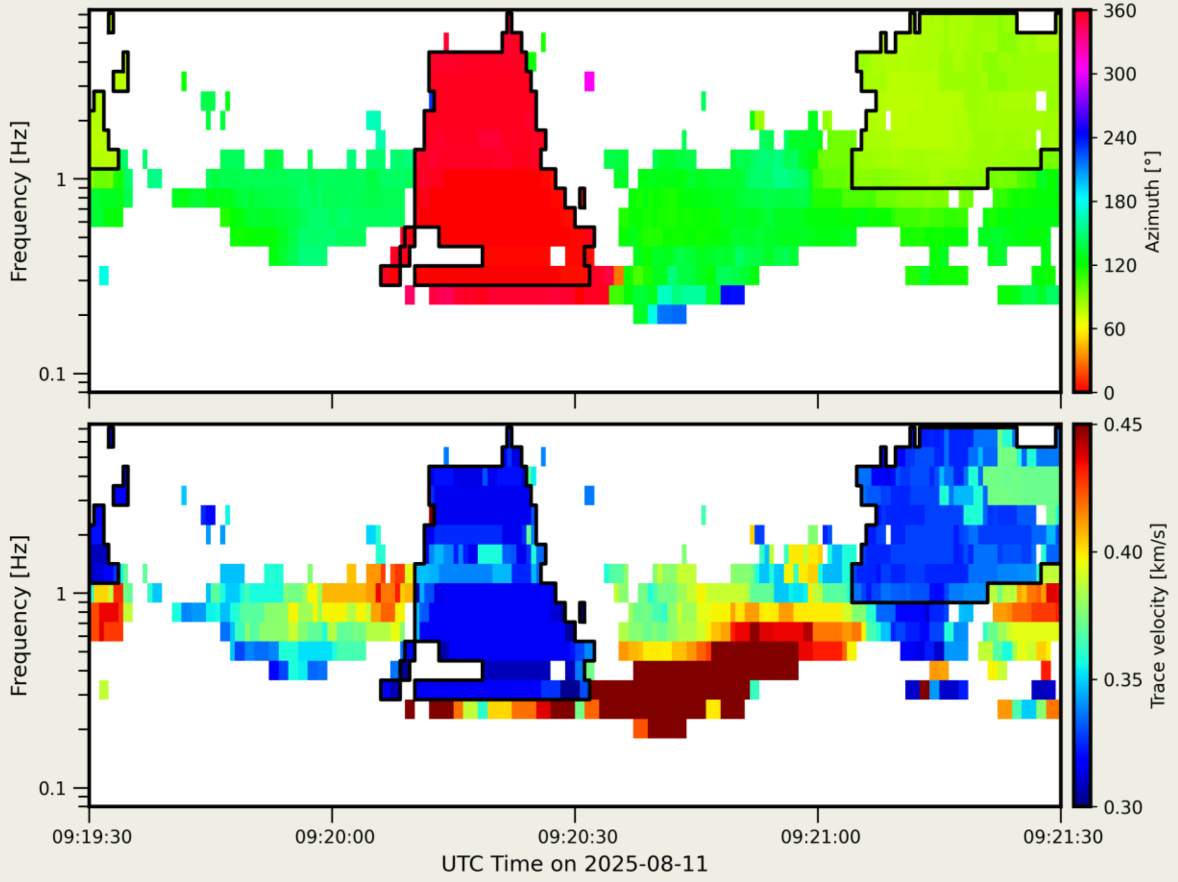


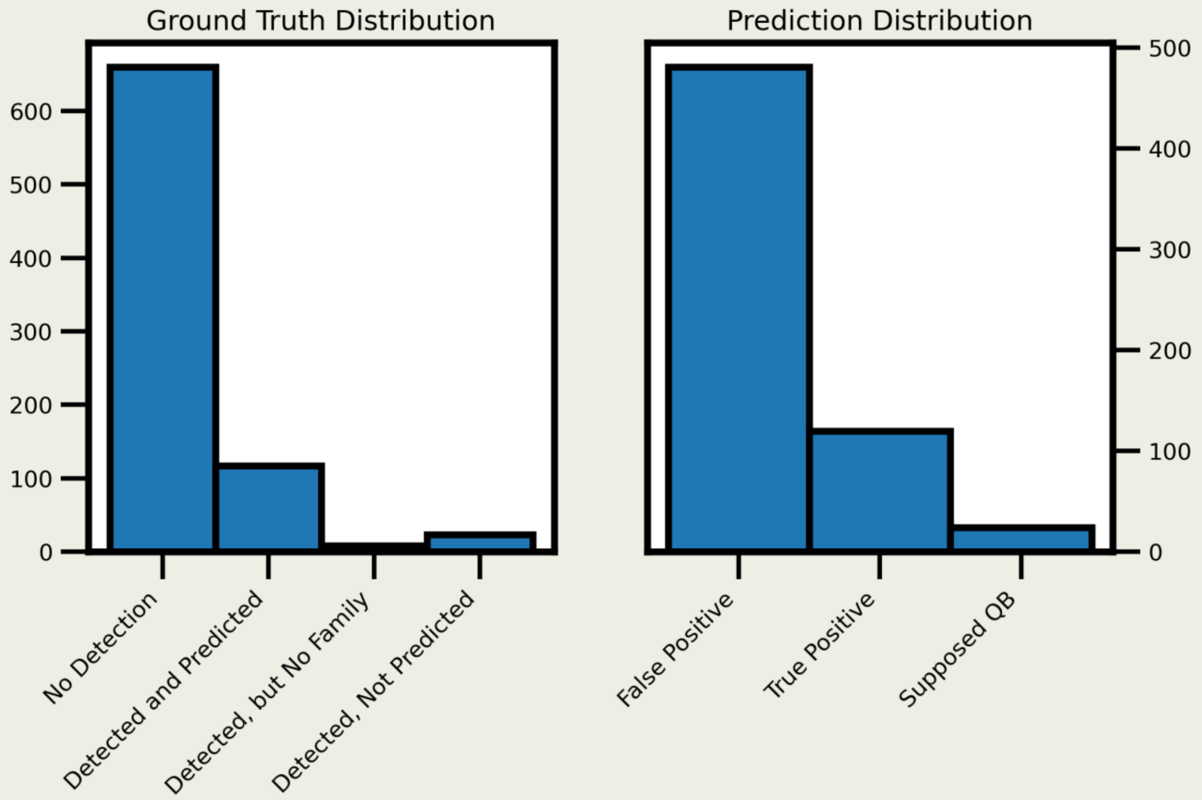
Event Type	Count
Etna	8
MPP	1
Oil refinery	31
Quarry blast	4
Thunderstorm	264
Unknown	215
War	514

Since April, war-related activity has been correctly identified on multiple days. The overlaid density plot shows the azimuthal distribution of detections labeled as 'war.' While most detections lie within the expected 60–90° range, a substantial number of misclassifications are observed, likely due to the concurrent presence of other signals. In such cases, azimuth could serve as an additional constraint.

Evaluation of automatic infrasound signal classification via Machine Learning deployed at the Central and Eastern European Infrasound Network

Marcell Pásztor, Tereza Sindelarova, Daniela Ghica, Ulrike Mitterbauer, Oleksandr Liashchuk, István Bondár





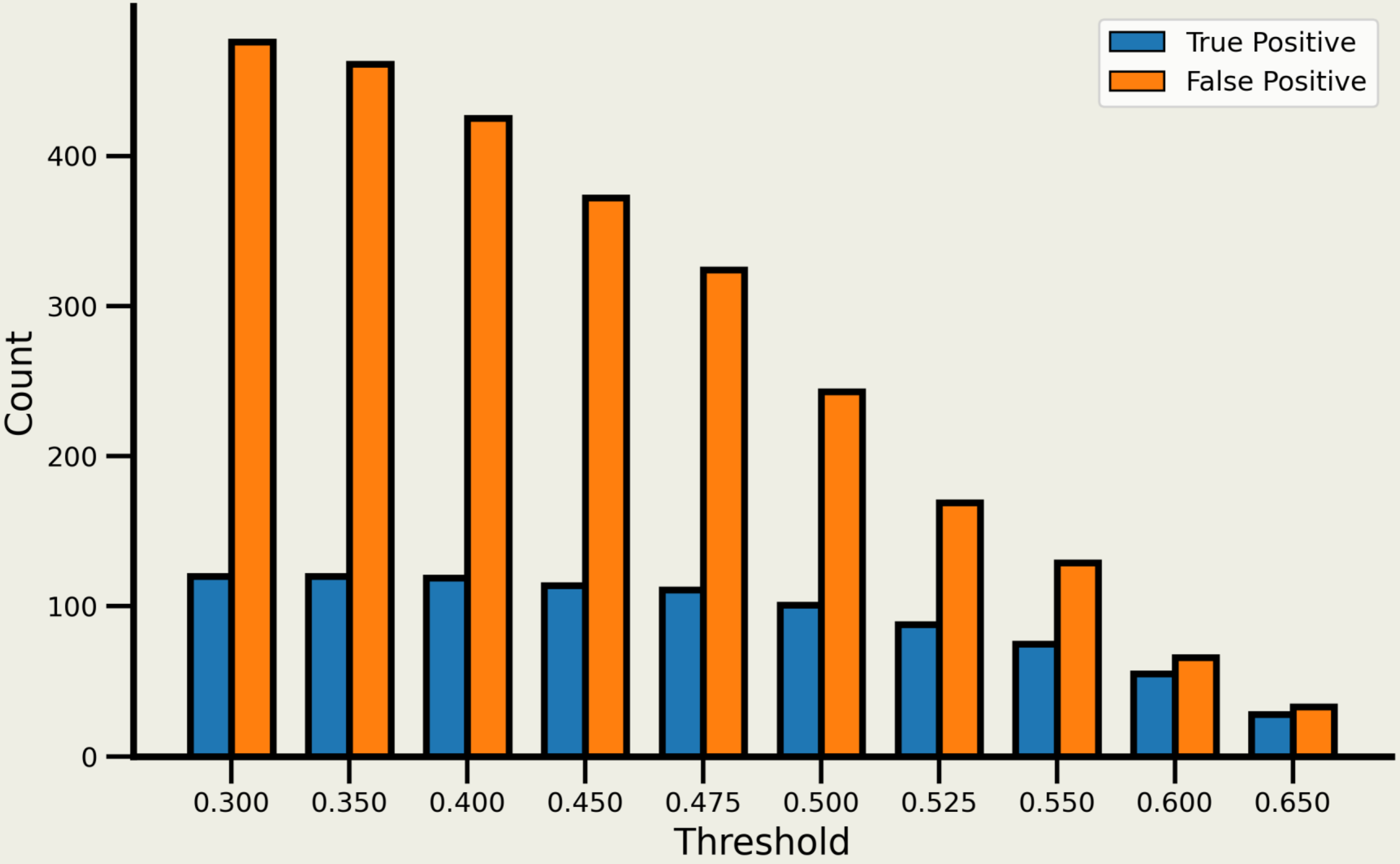
Example 5: Quarry blasts

Quarry blasts are used to evaluate the performance of the monitoring scheme. Predictions are matched to seismic ground-truth information and mine confirmations on a daily basis, following the same method used for labeling.

It should be noted that many quarry blasts are not detected. For context, the ratio of true positives to total detections is 85.7%. In addition, a few events lack corresponding ground-truth data.

Evaluation of automatic infrasound signal classification via Machine Learning deployed at the Central and Eastern European Infrasound Network

Marcell Pásztor, Tereza Sindelarova, Daniela Ghica, Ulrike Mitterbauer, Oleksandr Liashchuk, István Bondár



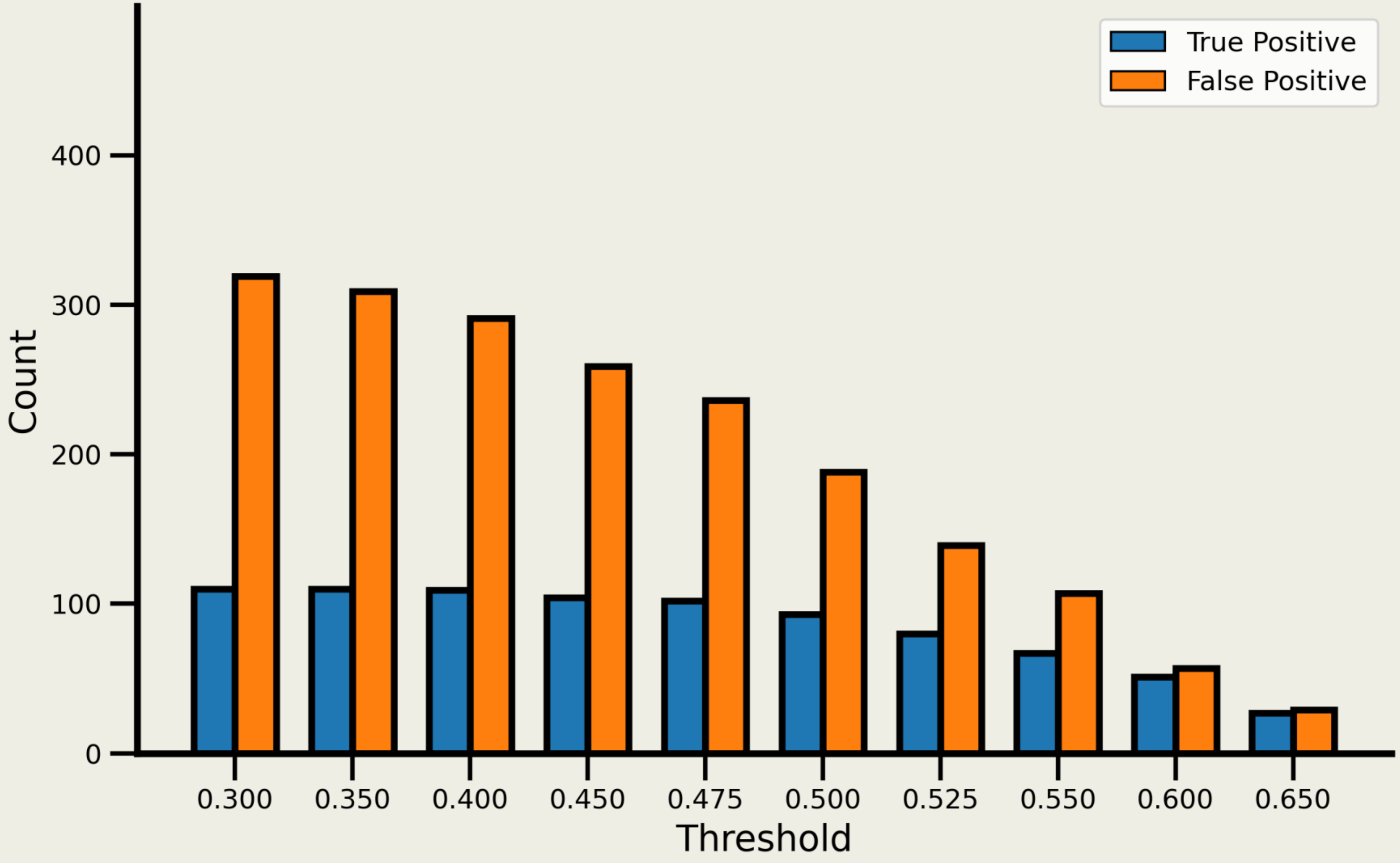
Example 5: Quarry blasts

Quarry blasts are used to evaluate the performance of the monitoring scheme. Predictions are matched to seismic ground-truth information and mine confirmations on a daily basis, following the same method used for labeling.

For context, the ratio of true positives to total detections is 85.7%. In addition, a few events lack corresponding ground-truth data. **At present, the class is determined by taking the argument of the maximum from the probability vector. The figure shows the effect of the different thresholds.**

Evaluation of automatic infrasound signal classification via Machine Learning deployed at the Central and Eastern European Infrasound Network

Marcell Pásztor, Tereza Sindelarova, Daniela Ghica, Ulrike Mitterbauer, Oleksandr Liashchuk, István Bondár



Example 5: Quarry blasts

Can we apply constraints similar to those used for Etna and the oil refineries?

Here, predictions are filtered between 6 AM and 15 PM.

Evaluation of automatic infrasound signal classification via Machine Learning deployed at the Central and Eastern European Infrasound Network

Marcell Pásztor, Tereza Sindelarova, Daniela Ghica, Ulrike Mitterbauer, Oleksandr Liashchuk, István Bondár

Conclusions

The presented monitoring scheme has shown good potential towards a machine learning based filtering method to highlight signals of interest (e.g., quarry blasts, volcanic eruptions) and reduce the number of detections that are originated from sources considered as noise (e.g., oil refineries).

However for reliable evaluation, longer testing period is required. Additional constraints after predictions can be used and thresholds can be adjusted to customize the model's output for the particular task.

ModelsV2

Since March 2025, the dataset has been updated, which should allow for the training of improved models. The new ensemble model (based on the Random Forest_v2 and CNN_v2) are operational since September 01, but due to the short period, a reliable comparison can't be made.

Future plans

In addition to testing the new models, we plan expanding the dataset with more classes (e.g., rocket launches) and incorporating additional stations. Upon interest, please feel free to contact us (pasztor.marcell@epss.hun-ren.hu) for collaboration—after all, in machine learning, more data is always better.

Once the single-array approach has been shown to be successful, we aim to extend the models to network-level processing as well.