Characterizing Seismic Stations’ Detection Capabilities with Supervised Learning

Dorian Golriz[1], David M. Steinberg[2], Yochai Ben-Horin[1] and Yael Radzyner[1]

[1] National Data Center, Soreq Nuclear Research Center, Yavne, Israel

[2] Department of Statistics and Operations Research, Tel-Aviv University, Tel-Aviv, Israel

SOREQ
Soreq Nuclear Research Center

TEL AVIV UNIVERSITY

INTRODUCTION AND MAIN RESULTS

Which station–event characteristics determine whether an event is detected or missed? Using supervised learning on global IDC data, we build baseline models of IMS station detection thresholds. We develop a Heat Score, capturing observed–predicted detection probability per epicentral region, and find that adding this score to the baseline model, along with time-of-day, can improve predictive skill and highlight conditions that control station sensitivity.

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

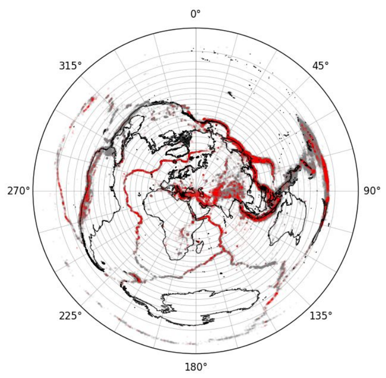
Characterizing Seismic Stations' Detection Capabilities with Supervised Learning

Dorian Golriz, David M. Steinberg, Yochai Ben-Horin and Yael Radzyner

P4.1-206

Introduction

Monitoring earthquakes and underground explosions is essential both for **hazard early warning** and for **verifying compliance with the Comprehensive Nuclear-Test-Ban Treaty (CTBT)**. The detection capability of a seismic station is primarily determined by the **event source**, its **distance and propagation path** to the station, and the **site characteristics**, which include the **background noise environment** and the **recording instrument**. Quantifying these detection thresholds provides a framework to assess **station performance** and to identify the conditions that limit sensitivity.



Example station KBZ (Khabaz, Russian Federation): Events plotted in polar coordinates relative to the station. Red points are detected events; grey points are undetected. This illustrates the geographic distribution of the dataset used for modeling.

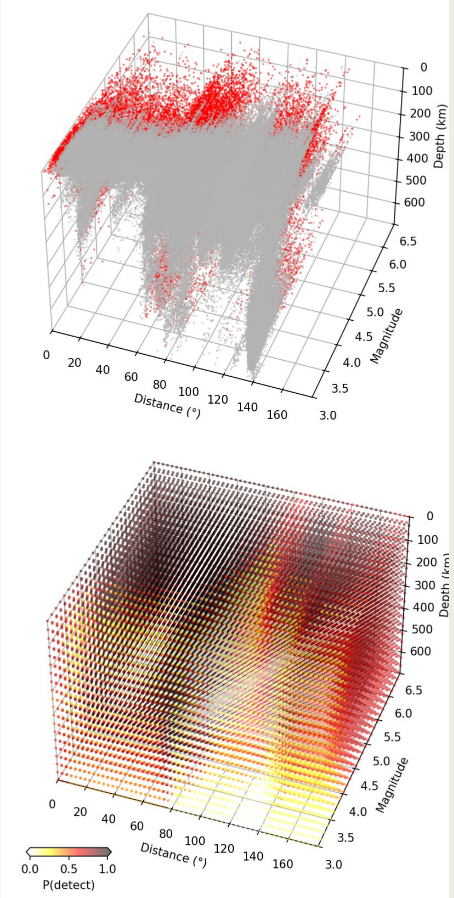
In this study, we develop **supervised learning models** to characterize the detection thresholds of **International Monitoring System (IMS) stations** using global event data reported by the **International Data Centre (IDC)**. Our objectives are to evaluate how well individual stations detect seismic events, to identify regions or conditions where models succeed or fail, and to test whether additional features, beyond a baseline of magnitude, distance, and depth, improve predictive skill. Beyond event parameters, we also want to investigate in the future whether **station-specific characteristics** such as instrument depth, installation type, or local geology carry predictive value for detection thresholds, either enhancing or limiting station sensitivity.

Methods

We analyze **~420,000 events** from the **IDC Late Event Bulletin (LEB)** covering **January 2014 – August 2025**. Four example IMS stations are presented here: **ILAR, KBZ, CMAR, and FINES**. Events associated with a time-defining phase at a station are treated as detected by the station.

An initial **Random Forest** model was trained separately for each station, using **event magnitude (mb)**, **station–event distance**, and **event depth** as input features. Models were trained on **85% of the data**, with the remaining **15% reserved for independent testing**. Hyperparameters were tuned per station using 10 replicates of 5-fold cross-validation.

To improve model skill beyond magnitude, distance, and depth, we introduced an additional feature we call the **Heat Score**. The idea is simple: we divide the globe into large azimuth–distance cells (~1M km²). For each cell, we compare the **observed detection probability** (fraction of events detected) with the **predicted probability** from the initial model. Their difference (Observed – Predicted) becomes the Heat Score. This value is then assigned to all events in that cell and used as an extra feature when training the new model (in contrast to the baseline model).



Training dataset and baseline model for KBZ. Top: event distribution in magnitude–distance–depth space (red = detected, grey = undetected). Bottom: predicted detection probability from the baseline Random Forest model using the same features, shown as continuous probability fields.

Soreq Nuclear Research Center
TEL AVIV UNIVERSITY

Characterizing Seismic Stations' Detection Capabilities with Supervised Learning

Dorian Golriz, David M. Steinberg, Yochai Ben-Horin and Yael Radzyner

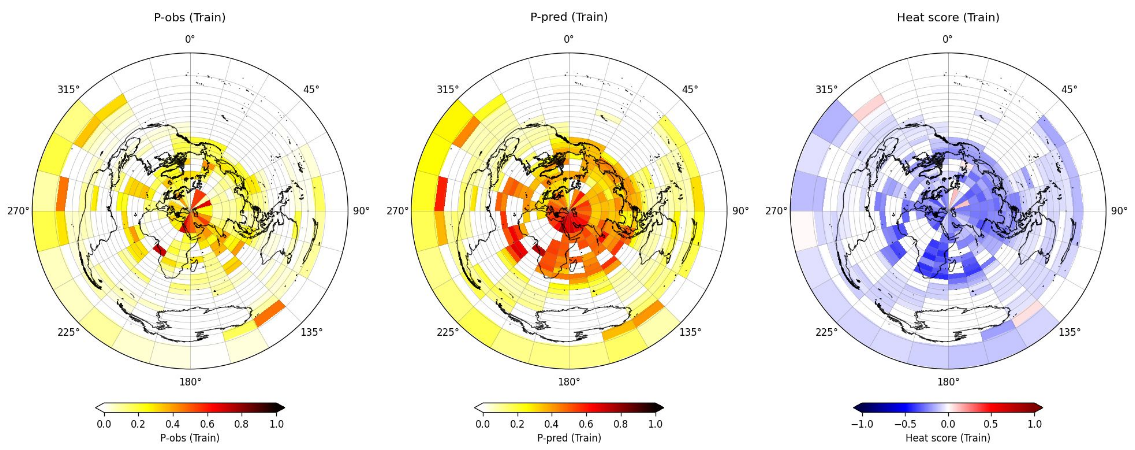
In this way, the Heat Score highlights regions where the station performs **better than expected ("hot") or worse ("cold")**, effectively encoding spatial residuals of the baseline model.

Preliminary Results

Across all four test stations, the new model that included the **Heat Score** feature showed **consistent, though modest, gains** over the baseline model on independent test data. Baseline AUC values ranged from **0.887–0.952** and increased to **0.907–0.964** with Heat Score (Δ=0.012–0.028). F1 scores rose from **0.650–0.792** to **0.688–0.819** (Δ=0.019–0.038), while LogLoss dropped from **0.259–0.431** to **0.205–0.373** (Δ=–0.041 to –0.065). Feature rankings confirm Heat Score contributes alongside magnitude and distance, while event depth remains **minor**.

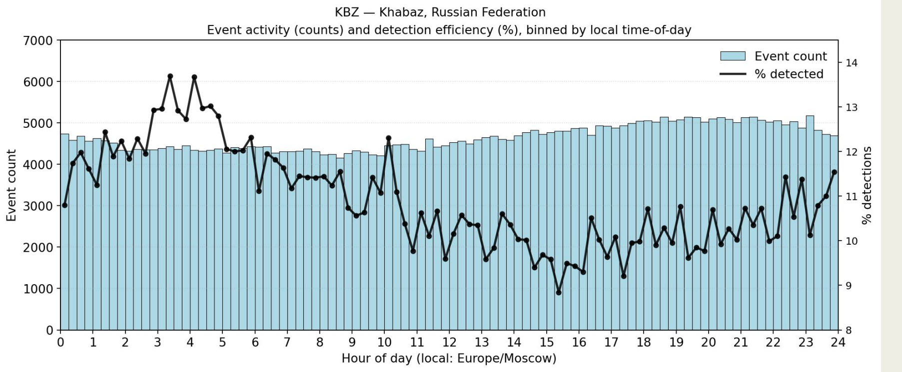
At **KBZ**, and similarly at other IMS stations, local **time-of-day analysis** reveals **lower detection fractions during local daytime hours**, consistent with elevated **cultural noise**. This supports the idea that including **temporal features** may further refine **station-specific models**.

Overall, these results demonstrate that integrating **derived features** reflecting station-specific **azimuthal performance (Heat Score)** and **environmental factors (local time-of-day)** may enhance our ability to predict detection outcomes.



Heat Score for KBZ. Left: observed detection probability (P-obs) across azimuth–distance bins. Middle: model-predicted probability (P-pred). Right: residuals (P-obs – P-pred, "Heat Score"), highlighting regions where the station is more effective (red) or less effective (blue) than expected.

Such enhanced models can ultimately inform improvements in **monitoring strategies** and guide evaluation of **site characteristics** that affect network performance.



Diurnal variation in KBZ detection efficiency. Histogram shows number of total events by local time of day. Black curve: percentage of detected events. Reduced detection fraction during daytime hours may reflect higher cultural noise, motivating exploration of local time-of-day as a model feature.

Future Directions

Ongoing work will expand the feature set to include **station-specific characteristics** (instrument type, installation depth, local geology) and **environmental conditions**. By systematically modeling detection thresholds across the IMS, we aim to provide a framework for evaluating **sensitivity**, identifying **weaknesses**, and guiding **network optimization**.

SOREQ
Soreq Nuclear Research Center
TEL AVIV UNIVERSITY