Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Azimuthal Dependence in Empirical Seismic Station Detection Threshold Models

Dorian Golriz

doriango@soreq.gov.il

National Data Center, Soreq Nuclear Research Center

David M. Steinberg

Tel Aviv University

Yochai Ben-Horin

National Data Center, Soreq Nuclear Research Center

Yael Radzyner

National Data Center, Soreq Nuclear Research Center

Research Article

Keywords: Seismic detectability, Azimuthal variability, International Monitoring System

Posted Date: May 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9534277/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Pure and Applied Geophysics on August 10th, 2026. See the published version at https://doi.org/10.1007/s00024-026-04047-3.

Azimuthal Dependence in Empirical Seismic Station Detection Threshold Models

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

[1]National Data Center, Soreq Nuclear Research Center, 81800, Yavne, Israel.
[2]Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel.

*Corresponding author(s). E-mail(s): doriango@soreq.gov.il;
Contributing authors: dms@tauex.tau.ac.il;
yochai@ndc.soreq-ndc.gov.il; yaelr@ndc.soreq-ndc.gov.il;

Abstract

Empirical models used to estimate seismic station detection thresholds commonly assume azimuthal symmetry by relying only on event magnitude and source–station distance. This simplification neglects directional variability arising from source radiation and propagation-path heterogeneity. We quantify the impact of including azimuth in empirical station-specific detection probability models using approximately two decades of International Monitoring System (IMS) seismic observations from 17 three-component primary stations.

Station-level detection is defined using time-defining phases in the IDC Late Event Bulletin (LEB) under verified waveform availability conditions. For each station, Random Forest classifiers are trained to estimate the detection probability using (i) magnitude and distance and (ii) magnitude, distance, and azimuth. Detection thresholds are defined as the minimum magnitude at which estimated probability exceeds 75% within cells containing sufficient observations.

Including azimuth systematically modifies empirical magnitude thresholds and reduces optimistic bias inherent in distance-only models. In some directions, threshold differences approach one magnitude unit. These results demonstrate that azimuthal variability is operationally relevant for station performance assessment and should be considered in empirical monitoring frameworks supporting CTBT verification.

Keywords: Seismic detectability, Azimuthal variability, International Monitoring System

1 Introduction

Reliable assessment of seismic detection capability is central to nuclear explosion monitoring under the Comprehensive Nuclear-Test-Ban Treaty (CTBT) and also supports regional and global earthquake monitoring. Although event declaration ultimately depends on network-level criteria, station-level detectability provides the physical and statistical basis upon which network performance rests (Woessner and Wiemer 2005). Empirical detection threshold models, alongside physics-based approaches, increasingly support station performance evaluation and network monitoring optimization (Gibbons et al. 2008; D'Alessandro et al. 2013; Mignan and Chouliaras 2014). Most such models describe detection probability primarily as a function of event magnitude and source–station distance, thereby imposing radial symmetry (Kagan 2003; Schorlemmer and Woessner 2008; Mignan et al. 2011; Schultz et al. 2015).

This symmetry assumption does not reflect established seismological principles since seismic amplitudes may vary with source radiation and propagation path. Radiation patterns depend on focal mechanism and azimuth relative to fault geometry (Boatwright and Choy 1986; Choy and Boatwright 1995), and for larger events may also reflect rupture directivity effects (Spudich and Chiou 2008; Kane et al. 2013). In addition, seismic waves propagate through laterally heterogeneous crust and mantle structures, introducing direction-dependent attenuation and potentially shadow zones (Kennett et al. 1995; Dalton et al. 2008). These mechanisms introduce inherent azimuthal variability in observed amplitudes and, consequently, in detectability.

Some operational monitoring developments acknowledge aspects of such directional variability. Regional correction efforts such as Regional Seismic Travel Time (RSTT) incorporate three-dimensional crustal structure to improve path predictions (Myers et al. 2015), and Bayesian monitoring frameworks such as NET-VISA embed spatial dependence in station-level priors derived from historical seismicity (Le Bras et al. 2021). Beyond empirical approaches, physics-based Threshold Monitoring methods estimate expected signal amplitudes at individual stations and compare them with measured background noise levels to determine detectability (Gibbons et al. 2008).

The central question addressed here is whether empirical station detection threshold models should explicitly incorporate azimuth, or whether magnitude and distance alone sufficiently characterize detectability. Beyond establishing its relevance, we seek to quantify the influence of azimuth on empirical threshold estimates. We investigate this question using long-term IMS observations and data-driven probabilistic detection models, and we measure how inclusion of azimuth modifies estimated magnitude thresholds and model performance.

2 IMS Data and Station-Level Detection Modeling

We analyze three-component IMS primary seismic stations using between 15 and 24 years of data from the IDC Late Event Bulletin (LEB), depending on station availability (Table 1). Event parameters, including origin time, location, and magnitude, are obtained from the `origin` table of the IDC database. For each origin in the bulletin, we define station-level detection by the presence of a time-defining phase associated

with that origin. If the station reports no such phase, we label the station–event pair as non-detection.

To ensure that non-detections reflect absence of observable signal rather than missing waveform data, we evaluate waveform availability using the `wfdisc` table. For each station-day, we compute daily coverage from reported sample counts and nominal sampling rate. We retain only days with availability between 90% and 110%. The lower bound removes periods with significant data gaps, while the upper bound filters days with overlapping waveform segments, which may indicate data integrity issues. After applying these filters, we discard station–event pairs that fall outside validated availability windows. Each station dataset ultimately contains between approximately 490,000 and 850,000 labeled station–event pairs.

We estimate detection probability using Random Forest classifiers (Breiman 2001; Biau and Scornet 2016). The model provides a continuous estimate of the probability of detection given event magnitude, distance, and (where included) azimuth, without imposing parametric assumptions. We do not convert these probabilities into binary classifications using a fixed cutoff. Rather, the model output is used directly to derive empirical detection thresholds.

For each station, we train two models independently. The first model uses event magnitude (m_b) and source–station distance (km) as predictors, representing the conventional radially symmetric formulation (hereafter the distance model). The second model augments these predictors with the azimuth between the station and the event source, denoted by ϕ. To preserve circular continuity and avoid artificial discontinuities at 360°, azimuth is encoded using $\sin(\phi)$ and $\cos(\phi)$, which together represent event direction in a continuous manner (hereafter the distance–azimuth model).

We divide each dataset into 85% training and 15% test subsets using stratified sampling to preserve the proportion of detections and non-detections. For each station, we tune three Random Forest hyperparameters: the number of trees, maximum tree depth, and minimum samples per leaf, using repeated cross-validation.

Model performance is evaluated on the independent test subset using the area under the receiver operating characteristic curve (ROC AUC) and log loss.

$$\mathrm{LL} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p_i) + (1-y_i)\log(1-p_i)\right], \tag{1}$$

where $y_i \in \{0,1\}$ denotes the observed detection outcome and $p_i \in (0,1)$ is the estimated detection probability.

To compare the two models in operationally meaningful terms, we estimate empirical magnitude thresholds rather than relying solely on classification metrics. We partition the globe relative to each station into 15 radial distance cells extending to 10,000 km and 36 azimuth sectors of 10° each, yielding 540 distance–azimuth cells. Radial cell edges follow a power-law spacing defined as

$$r_i = R_{\max}\left(\frac{i}{K}\right)^{\alpha} \tag{2}$$

where r_i denotes the radial cell edge, i is the radial cell index, K is the total number of radial cells, $R_{\max}$ is the maximum distance considered, and $\alpha = 0.85$ controls

the spacing of the cells. This formulation produces coarser sampling near the station and finer spacing at larger distances. We retain only cells containing at least 250 events to ensure statistically meaningful estimates. Within each cell, we assume that detection probability increases monotonically with magnitude. This assumption reflects the general physical expectation that larger earthquakes generate larger signal amplitudes at a given source–station geometry. Although factors such as focal depth can influence observed amplitudes, the overall trend of increasing detectability with earthquake magnitude (Noda and Ellsworth 2016) and explosion yield (Ringdal et al. 1992; Peacock et al. 2017) is well established in empirical detection studies.

Table 1 Summary of station datasets and model performance. Cells(%) is the percentage of distance–azimuth cells that satisfy the minimum event-count requirement ($N \geq 250$) and reach the $P = 75\%$ detection threshold. Columns AUC_D and LL_D correspond to the distance model, while AUC_{DA} and LL_{DA} correspond to the distance–azimuth model. LL denotes log loss. Stations highlighted in bold are shown as examples in Figure 1.

Station	StartDate	DataAvail(%)	Events	Detect(%)	CellsUsed(%)	AUC_D	AUC_{DA}	ΔAUC	LL_D	LL_{DA}	ΔLL
BDFB	**2004-02-19**	**98.2**	**600214**	**7.6**	**15.6**	**0.967**	**0.975**	**+0.008**	**0.227**	**0.158**	**-0.069**
BOSA	2005-03-25	95.4	720396	7.5	18.0	0.951	0.969	+0.018	0.270	0.169	-0.101
CPUP	2003-08-29	94.7	720211	8.3	17.2	0.960	0.968	+0.008	0.255	0.182	-0.073
DBIC	2004-04-30	97.2	759377	8.4	20.0	0.948	0.966	+0.018	0.284	0.185	-0.099
KBZ	**2009-02-18**	**98.8**	**655837**	**11.2**	**27.4**	**0.940**	**0.961**	**+0.021**	**0.306**	**0.213**	**-0.093**
KEST	2006-07-29	98.9	739402	4.9	23.5	0.964	0.974	+0.010	0.224	0.129	-0.095
KMBO	2002-06-07	95.7	695498	3.9	22.6	0.962	0.971	+0.009	0.229	0.137	-0.092
LPAZ	2003-05-10	95.5	770126	14.2	19.6	0.962	0.972	+0.010	0.246	0.190	-0.056
MAW	2003-07-04	98.0	809816	7.7	18.7	0.910	0.940	+0.030	0.367	0.248	-0.119
NRIK	2011-01-07	96.6	491215	8.2	24.3	0.944	0.958	+0.014	0.286	0.202	-0.084
PLCA	2004-04-23	96.0	733371	10.2	17.0	0.948	0.964	+0.016	0.287	0.200	-0.087
PPT	2001-09-21	98.5	849681	0.8	24.4	0.959	0.965	+0.006	0.231	0.165	-0.066
ROSC	2003-05-10	98.2	748178	3.2	19.6	0.983	0.986	+0.003	0.149	0.123	-0.026
SCHQ	2003-05-08	99.0	817003	6.3	23.0	0.951	0.969	+0.018	0.271	0.165	-0.106
STKA	2002-12-03	99.1	816263	26.4	24.3	0.890	0.926	+0.036	0.413	0.336	-0.077
ULM	**2003-06-01**	**98.3**	**804701**	**8.1**	**22.4**	**0.957**	**0.972**	**+0.015**	**0.259**	**0.166**	**-0.093**
VNDA	**2004-03-23**	**97.9**	**723307**	**11.4**	**16.9**	**0.900**	**0.930**	**+0.030**	**0.395**	**0.298**	**-0.097**

For each cell, we determine the minimum magnitude at which the estimated detection probability (P) exceeds 75% and define this value as the empirical detection threshold. We select the 75% level because some cells do not reach higher probabilities within the available magnitude range. Cells that do not reach the $P = 75\%$ threshold are assigned undefined values. Finally, we quantify azimuthal effects by computing threshold differences as the distance–azimuth model threshold minus the distance model threshold. Positive differences indicate that the radially symmetric model underestimates the magnitude required for reliable detection.

3 Quantifying Azimuthal Effects on Magnitude Thresholds

Figure 1 illustrates empirical magnitude thresholds derived from the two detection models for four representative IMS stations: BDFB (Brasilia, Brazil), KBZ (Khabaz, Russian Federation), VNDA (Vanda, Antarctica), and ULM (Lac du Bonnet, Canada). Each row corresponds to one station. The left column shows thresholds estimated from the distance model, which by construction produces radially symmetric threshold fields. The middle column shows thresholds obtained from the distance–azimuth model. In contrast to the radial structure of the distance model, the distance–azimuth model reveals clear directional variability in detection capability, with cells at similar distances but different azimuths often exhibiting different threshold magnitudes.

For all four stations, the distance–azimuth model produces structured spatial patterns in which thresholds vary substantially with direction. The right panels of Figure 1 show these differences explicitly, defined as $\Delta m_b = m_b^{DA} - m_b^{D}$. Positive values dominate most cells across stations, indicating that the distance model often underestimates the magnitude required for reliable detection. In several azimuthal sectors, threshold differences approach one magnitude unit relative to the radially symmetric model. At the same time, some sectors show negative differences, indicating that the effect of azimuth is not uniform and may either increase or decrease the estimated threshold depending on direction.

This behavior can be observed in multiple tectonic regions sampled by the stations. At BDFB, cells corresponding to events along the South American subduction zone and the Mid-Atlantic Ridge show predominantly positive differences, whereas cells in the Caribbean show negative differences despite lying at broadly comparable distances from the station. At ULM, cells toward the western United States and Alaska are mostly positive, while cells toward the Mexican subduction zone and farther south toward Panama and the northern Andes are mostly negative. At KBZ, many cells toward China, Nepal, and the Tibetan Plateau are positive, whereas cells toward the North Atlantic near Greenland are negative. These examples illustrate that events located at similar distances but different azimuths do not share equivalent detectability, and that empirical detection models capture spatial heterogeneity that cannot be represented by distance alone.

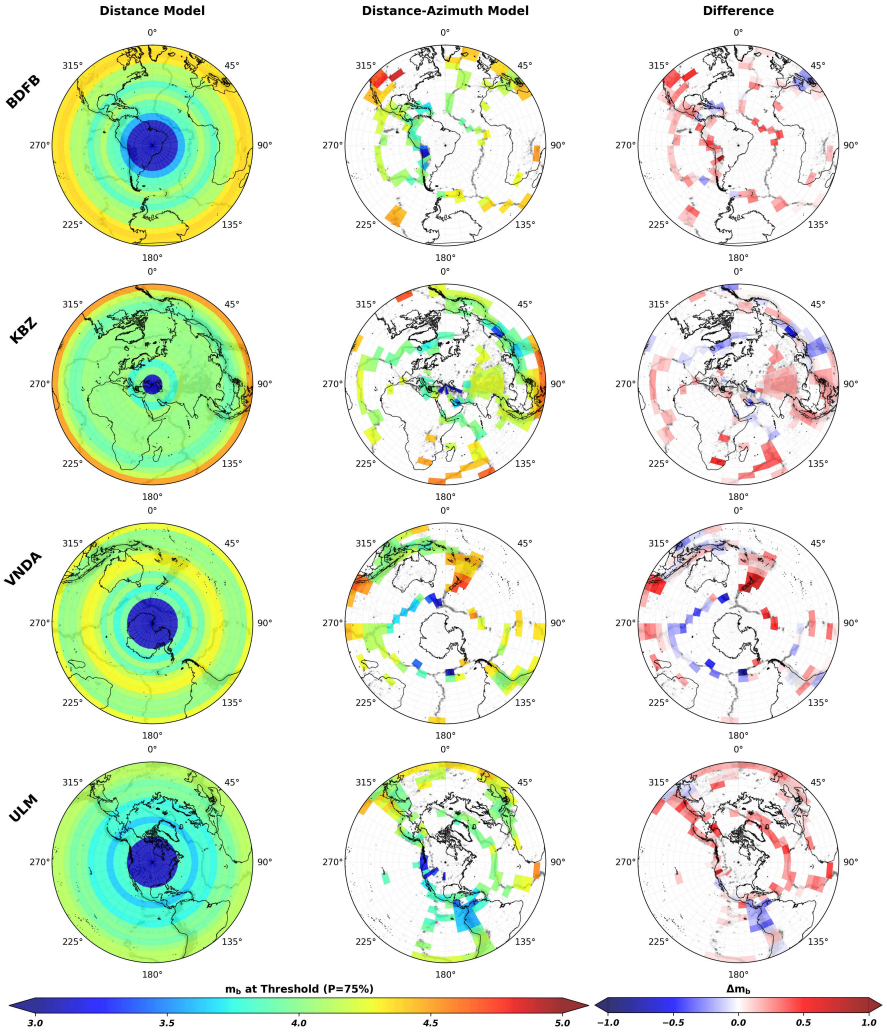


Fig. 1 Empirical magnitude detection thresholds for four representative IMS stations. Left panels show thresholds from the distance model, middle panels show thresholds from the distance–azimuth model, and right panels show their differences ($\Delta m_b = m_b^{DA} - m_b^{D}$). Colors indicate the minimum magnitude at which estimated detection probability exceeds $P = 75\%$. Event locations used in the analysis are shown as light grey points.

While spatial threshold patterns provide intuitive visualization of directional effects, model performance metrics provide a quantitative assessment of the statistical contribution of azimuth. Figure 2 compares model performance of the two models across all analyzed stations using independent test datasets that were not used during model training, while Table 1 summarizes the corresponding values.

Two complementary metrics are used. The area under the receiver operating characteristic curve (ROC AUC) measures the model's ability to discriminate between detections and non-detections across all probability thresholds, with larger values

indicating better classification performance. Log loss evaluates the quality of probability estimates by penalizing confident but incorrect estimates. Lower log loss values therefore indicate better calibrated or more reliable probability estimates.

For all 17 analyzed three-component IMS primary stations, the distance–azimuth model yields higher ROC AUC values and lower log loss values than the distance model (Figure 2; Table 1). The magnitude of these improvements is modest for individual stations (typically about 0.005–0.03 in ROC AUC and about 0.05–0.12 in log loss), but their consistency across the entire station set demonstrates that azimuth contributes additional predictive information beyond magnitude and distance alone.

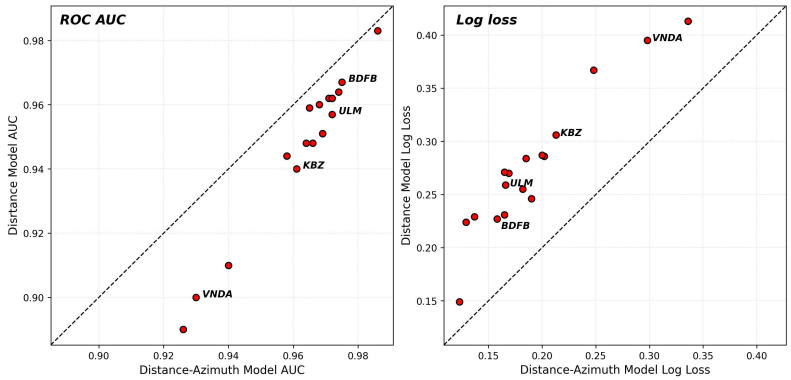


Fig. 2 Comparison of model performance between the distance and distance–azimuth models for all 17 analyzed stations. The left panel shows ROC AUC values and the right panel shows log loss values computed on independent test datasets. The black dashed line indicates the one-to-one reference. Points below the line in the ROC AUC panel and above the line in the log loss panel indicate improved performance of the distance–azimuth model.

4 Implications for CTBT Verification

Our results demonstrate that azimuth has a measurable effect on empirical seismic station detection thresholds. Models that rely solely on magnitude and distance impose artificial radial symmetry and can misrepresent directional variations in station performance. In most analyzed cells, thresholds derived from the distance–azimuth model exceed those from the distance model, indicating systematic optimistic bias when azimuth is ignored. In some sectors, threshold differences between the two models approach one magnitude unit. The distance–azimuth model also reveals that cells at similar distances but different azimuths can exhibit different detection thresholds.

For CTBT monitoring, such station-level asymmetry has implications for network-level capability assessments. Because station detection performance contributes directly to network detection capability, directional asymmetry propagates into network-level monitoring evaluations. Empirical frameworks used for threshold monitoring and station performance evaluation are therefore expected to benefit from accounting for directional variability in signal behavior.

This study provides a proof-of-concept demonstration that incorporating azimuth improves empirical station detection models. Although improvements in ROC AUC and log loss are modest, they are consistent across all 17 analyzed three-component IMS primary stations. These results indicate that incorporating azimuth reduces bias in empirical magnitude threshold estimates and improves the statistical representation of station detectability.

Acknowledgements. The authors acknowledge the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) for providing access to International Data Center (IDC) Late Event Bulletin (LEB) used in this study.

References

Biau G, Scornet E. A random forest guided tour. Test. 2016;25(2):197–227.

Boatwright J, Choy GL. Teleseismic estimates of the energy radiated by shallow earthquakes. Journal of Geophysical Research: Solid Earth. 1986;91(B2):2095–2112.

Breiman L. Random forests. Machine learning. 2001;45(1):5–32.

Choy GL, Boatwright JL. Global patterns of radiated seismic energy and apparent stress. Journal of geophysical research: Solid earth. 1995;100(B9):18205–18228.

Dalton CA, Ekström G, Dziewoński AM. The global attenuation structure of the upper mantle. Journal of Geophysical Research: Solid Earth. 2008;113(B9).

D'Alessandro A, Badal J, D'Anna G, Papanastassiou D, Baskoutas I, Özel NM. Location performance and detection threshold of the Spanish National Seismic Network. Pure and Applied Geophysics. 2013;170(11):1859–1880.

Gibbons SJ, Ringdal F, Kværna T. Detection and characterization of seismic phases using continuous spectral estimation on incoherent and partially coherent arrays. Geophysical Journal International. 2008;172(1):405–421.

Kagan YY. Accuracy of modern global earthquake catalogs. Physics of the Earth and Planetary Interiors. 2003;135(2-3):173–209.

Kane DL, Shearer PM, Goertz-Allmann BP, Vernon FL. Rupture directivity of small earthquakes at Parkfield. Journal of Geophysical Research: Solid Earth. 2013;118(1):212–221.

Kennett BL, Engdahl E, Buland R. Constraints on seismic velocities in the Earth from traveltimes. Geophysical Journal International. 1995;122(1):108–124.

Le Bras R, Arora N, Kushida N, Mialle P, Bondár I, Tomuta E, et al. NET-VISA from cradle to adulthood. A machine-learning tool for seismo-acoustic automatic association. Pure and Applied Geophysics. 2021;178(7):2437–2458.

Mignan A, Werner M, Wiemer S, Chen CC, Wu YM. Bayesian estimation of the spatially varying completeness magnitude of earthquake catalogs. Bulletin of the Seismological Society of America. 2011;101(3):1371–1385.

Mignan A, Chouliaras G. Fifty years of seismic network performance in Greece (1964-2013): spatiotemporal evolution of the completeness magnitude. Seismological Research Letters. 2014;85(3):657–667.

Myers SC, Simmons NA, Johannesson G, Matzel E. Improved regional and teleseismic P-wave travel-time prediction and event location using a global 3D velocity model. Bulletin of the Seismological Society of America. 2015;105(3):1642–1660.

Noda S, Ellsworth WL. Scaling relation between earthquake magnitude and the departure time from P wave similar growth. Geophysical Research Letters. 2016;43(17):9053–9060.

Peacock S, Douglas A, Bowers D. Joint maximum-likelihood magnitudes of presumed underground nuclear test explosions. Geophysical Journal International. 2017;210(2):621–644.

Ringdal F, Marshall PD, Alewine RW. Seismic yield determination of Soviet underground nuclear explosions at the Shagan River test site. Geophysical Journal International. 1992;109(1):65–77.

Schorlemmer D, Woessner J. Probability of detecting an earthquake. Bulletin of the Seismological Society of America. 2008;98(5):2103–2117.

Schultz R, Stern V, Gu YJ, Eaton D. Detection threshold and location resolution of the Alberta Geological Survey earthquake catalogue. Seismological Research Letters. 2015;86(2A):385–397.

Spudich P, Chiou BS. Directivity in NGA earthquake ground motions: Analysis using isochrone theory. Earthquake Spectra. 2008;24(1):279–298.

Woessner J, Wiemer S. Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty. Bulletin of the Seismological Society of America. 2005;95(2):684–698.