LEARNING THE ACOUSTIC BACKGROUND FOR WEAK VESSEL OBSERVABILITY THROUGH TEMPORAL PROBABILISTIC DEVIATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Rather than learning what vessels sound like, we model the acoustic background itself. This motivates a distinction between *vessel presence* and *vessel observability*: the former is a geometric condition, whereas the latter depends on whether the received acoustic signal is distinguishable from the prevailing soundscape. We introduce Probabilistic Background Support Modelling (PBSM), a background-first framework that learns the local operational acoustic background from no-contact recordings and identifies statistically unusual 30-second observations as *acoustic departures*. Rather than assigning these departures to a predefined acoustic class, PBSM reduces large passive-acoustic datasets to short-duration observations warranting further investigation. An Empirical Background Anomaly Percentile (EBAP) expresses departure relative to the learned background, while a matched 30-second no-contact control accounts for departures caused by temporal representation. Acoustic scoring is blind to vessel position, range, Closest Point of Approach (CPA), and identity. Identified departures are subsequently examined for substantive spectral excess, with AIS-derived vessel geometry introduced only afterward as an independent physical reference. Across eight station-season datasets, short-duration background variability can itself produce substantial acoustic departure, demonstrating that departure alone cannot be interpreted as vessel detection. The prevalence and discriminative value of PBSM departures vary across operational environments, while spectral characterization can retain meaningful acoustic structure even where background departure is prevalent. Where PBSM departure, substantive spectral excess, and vessel proximity converge, increasingly strong acoustic evidence is observed as CPA decreases. The principal contribution is their integration into a background-first evidence chain for acoustic observability: learn the local operational background, identify unusual observations blindly, characterize their spectral structure, and use independent vessel geometry to interpret the resulting evidence.

1 INTRODUCTION

Passive acoustic vessel detection in real-world underwater environments is complicated by highly variable background noise and low signal-to-noise conditions. Vessel-generated acoustic energy may overlap substantially with the ambient acoustic field, such that the physical presence of a vessel does not necessarily imply that its acoustic contribution is distinguishable from background conditions (Kinsler et al., 2000). This motivates a distinction between *vessel presence* and *vessel observability*: the former indicates that a vessel is physically present, whereas the latter depends on whether its received acoustic signal is distinguishable from the prevailing soundscape.

Rather than approaching weak vessel detection as a conventional supervised classification problem, we ask whether an acoustic observation remains statistically consistent with a background distribution learned exclusively from periods in which no qualifying Automatic Identification System (AIS)-confirmed vessel is present. These *no-contact* periods are not assumed to be acoustically silent, but represent the operational underwater acoustic background encountered at the sensor in the absence of a qualifying AIS contact.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

We introduce Probabilistic Background Support Modelling (PBSM), which models this operational background using Power Spectral Density (PSD) representations and a Gaussian Mixture Model (GMM). An independent background model is learned for each station and observation period from full-duration no-contact recordings, and consecutive 30-second acoustic windows are assigned a PBSM score representing their statistical support under the corresponding local background model. This formulation is closely related to out-of-distribution (OOD) detection, with the learned no-contact background defining the reference distribution against which low-support observations are identified. Low-support observations are interpreted as *acoustic departures*: observations that are statistically unusual relative to the corresponding learned background distribution, thereby reducing the acoustic search space to observations warranting further investigation rather than assignment to a predefined acoustic class.

We also introduce the Empirical Background Anomaly Percentile (EBAP), which expresses each PBSM score relative to the empirical background score distribution. Specifically, EBAP is the percentage of full-duration background scores less than or equal to the evaluated score, so low EBAP indicates unusually low support under the learned background. EBAP is not a probability that a vessel is present.

A central methodological issue is temporal representation. PBSM learns the background from full-duration no-contact PSDs but evaluates consecutive 30-second windows, which may exhibit greater variability even in the absence of a qualifying AIS contact. We therefore apply the same 30-second analysis to an independent held-out no-contact control set, providing a matched control for determining whether one-contact recordings exhibit additional departure beyond this temporal-representation effect.

PBSM scoring does not use vessel geometry relative to the sensor. AIS is used to construct the no-contact and one-contact populations, while vessel geometry is introduced only after acoustic scoring as an independent physical reference. Spectral characterization then determines whether identified departures contain *substantive spectral excess* relative to background conditions. AIS-derived geometry is then used as an independent physical reference to determine whether acoustic departure and spectral evidence vary systematically with vessel proximity, thereby evaluating whether the independently identified acoustic evidence is associated with known vessel presence.

The objective of this work is therefore to identify background-relative acoustic departures and determine whether they contain substantive spectral excess that varies systematically with independently derived vessel geometry, providing evidence of acoustic observability associated with known vessel presence.

The principal contributions of this work are:

- PBSM, a background-first framework for identifying *acoustic departures* from a locally learned operational background;
- EBAP, which expresses PBSM scores as percentiles of the corresponding empirical background-score distribution;
- a matched 30-second no-contact control for the effect of temporal representation;
- spectral characterization of identified departures using *substantive spectral excess*; and
- independent evaluation of PBSM acoustic departure and *substantive spectral excess* against AIS-derived vessel geometry introduced only after PBSM acoustic scoring.

2 Related Work

Passive acoustic vessel detection has commonly been approached as a classification problem in which acoustic observations are used to distinguish vessels from background conditions or among predefined vessel classes (Aslam et al., 2024). Santos-Domínguez et al. (Santos-Domínguez et al., 2016) constructed the ShipsEar dataset and demonstrated detection and classification using cepstral features and Gaussian mixture models (GMMs), with separate models trained for four vessel-size classes and a background-noise class. DeepShip extended this supervised setting to more than 47 hours of recordings from 265 vessels spanning four commercial vessel classes, with AIS used to retain recordings containing a single vessel within 2 km of the hydrophone (Irfan et al., 2021).

More recent supervised approaches have extended vessel-acoustic classification through increasingly sophisticated spectral representations and operational passive-acoustic monitoring. Najamuddin et al. (Najamuddin et al., 2025) combined coherent spectral estimation with a Vision Transformer on the ShipsEar and DeepShip benchmarks, while da Silva et al. (Silva et al., 2025) evaluated supervised machine-learning classifiers using passive-sonar recordings from 12 vessel classes. DeepPAM (Topple et al., 2024) uses deep neural networks to detect and classify individually labelled vessels in Arctic acoustic data. Jedari-Eyvazi et al. (Jedari-Eyvazi et al., 2026) similarly evaluate low-frequency vessel tonals and a convolutional neural network trained for binary ship-noise detection across multiple Western Canadian Arctic monitoring sites. Despite differences in architecture and deployment setting, these approaches rely on vessel-specific acoustic characteristics or labelled vessel observations. PBSM instead learns only the operational no-contact background during acoustic model training.

A closely related but independently developed background-modelling approach was reported by Mihailov (Mihailov, 2025), who fitted a GMM exclusively to background-noise observations and evaluated test samples using their log-likelihood under the learned background distribution. A lower-tail threshold was then used to identify potential acoustic anomalies. This provides independent evidence for probabilistic modelling of the underwater acoustic background. The development of PBSM preceded our awareness of this work and was undertaken independently.

PBSM additionally examines the effect of temporal representation between full-duration background observations and 30-second test windows using temporally matched no-contact controls. EBAP expresses these departures relative to the empirical background score distribution. Related work on out-of-distribution and one-class inference provides statistical context for this interpretation. Liang et al. (Liang et al., 2022) describe an inductive conformal formulation in which conformity scores are interpreted through their relative rank among held-out inlier scores. This is closely related to the motivation for EBAP, although EBAP is used here as a descriptive empirical percentile without the finite-sample validity or false-discovery guarantees associated with conformal p-values.

3 METHODOLOGY

A common processing and analysis pipeline is applied unchanged across all station-season datasets, with fixed preprocessing, model configuration, decision criteria, and analysis procedures and no station- or season-specific tuning to improve individual results (Trochim et al., 2016).

3.1 AIS-CONDITIONED ACOUSTIC DATASET

The acoustic dataset is constructed using the AIS-acoustic integration pipeline described in Renaud et al. (2026), which temporally and spatially aligns fixed-duration acoustic recordings with AIS vessel observations. Two experimental populations are defined:

- **No-contact recordings**: recordings containing no qualifying AIS vessel contact during the acoustic interval. These provide the operational background used to train PBSM. No-contact indicates the absence of qualifying AIS-correlated vessel traffic, not the absence of acoustic activity.
- **One-contact recordings**: recordings containing exactly one qualifying AIS vessel, identified by MMSI. Recordings containing more than one qualifying unique MMSI are excluded from this population.

AIS conditioning defines the experimental populations rather than the acoustic detection result. One-contact establishes vessel presence but not acoustic observability or attribution of particular acoustic structure to that vessel.

3.2 ACOUSTIC REPRESENTATION

Background training and temporal evaluation use different acoustic representations. Each no-contact training recording is represented by a single Power Spectral Density (PSD) calculated over its full approximately 11-minute duration, whereas one-contact recordings are divided into consecutive,

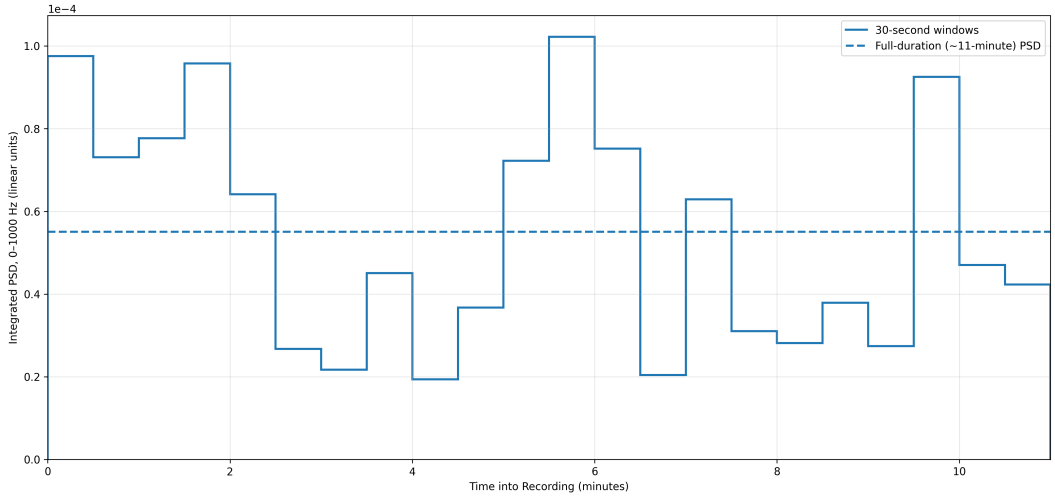


Figure 1: Integrated 0–1000 Hz PSD for a representative no-contact recording. Consecutive 30-second windows vary substantially around the full-duration approximately 11-minute representation, demonstrating that temporal representation alone can produce short-duration acoustic departure.

non-overlapping 30-second windows. Only complete 30-second windows are retained; any remaining audio shorter than 30 seconds at the end of a recording is discarded.

PSD provides a frequency-domain representation of signal power Ifeachor & Jervis (2002). Estimates are calculated using Welch's method with 2048-sample segments over 0–1000 Hz. The same procedure is used for full-duration and 30-second observations; only their temporal duration differs.

This asymmetry allows the full-duration PSD to represent the characteristic background state after short-duration fluctuations have been averaged, while 30-second windows expose finer temporal variation. Consequently, even 30-second no-contact observations can depart substantially from their full-duration representation, as illustrated in Figure 1.

Held-out no-contact recordings therefore provide an explicit matched control for the change in temporal representation. The 30-second no-contact (NC) windows are therefore subject to the same temporal representation as the 30-second one-contact windows. Comparing these two 30-second populations establishes whether PBSM acoustic departure in the one-contact population exceeds that attributable to temporal representation alone.

Linear PSD is retained during acoustic preprocessing and converted to a logarithmic representation before probabilistic modelling. Dimensionality reduction and feature standardization are learned exclusively from the full-duration no-contact training observations.

3.3 Probabilistic Background Support Modelling (PBSM)

Probabilistic Background Support Modelling (PBSM) applies established Gaussian Mixture Model (GMM) density estimation to the operational *no-contact* acoustic background (Murphy, 2023; Bishop & Bishop, 2024; Dempster et al., 1977). Log-PSD observations are reduced using Principal Component Analysis (PCA), fitted exclusively to the full-duration no-contact training observations, and standardized using parameters estimated from the same population. The frozen configuration retains 64 principal components.

The background density is

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}\left(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right), \tag{1}$$

where π_k, $\boldsymbol{\mu}_k$, and $\boldsymbol{\Sigma}_k$ denote the mixture weight, mean, and covariance of component k. The frozen configuration uses $K = 12$ full-covariance GMM components, estimated from no-contact training observations using expectation–maximization (EM) (Dempster et al., 1977). The same configuration is fitted independently for each of the eight station-seasons, producing a local rather than geographically transferred background model.

The 12 GMM components collectively provide a flexible density model of the multimodal structure present within the local operational background. During model development, multiple PCA dimensionalities and GMM configurations were trialled, with their behaviour examined relative to independently derived vessel CPA. This development process resulted in the PCA-64, GMM 12-component full-covariance configuration used here. The configuration was subsequently frozen before the reported evaluation and applied without station- or season-specific tuning.

Each 30-second observation is evaluated against its local background model; low-support observations are identified as *acoustic departures* rather than assigned to an acoustic class.

3.4 EBAP and Acoustic Departure Detection

PBSM is applied blindly to consecutive 30-second one-contact windows, without AIS-derived vessel position, timing, range, CPA, or identity. Each window is transformed using the PCA and standardization parameters learned from the full-duration no-contact training population and assigned the background log-likelihood

$$s_t = \log p(\mathbf{x}_t \mid \mathcal{B}), \qquad t = 1, \ldots, T, \tag{2}$$

where $\mathbf{x}_t$ is the frozen 64-dimensional PCA representation of the 30-second PSD observation in window t, $\mathcal{B}$ is the fitted GMM probability-density model representing the distribution of the full-duration no-contact training PSD observations, s_t is the resulting background log-likelihood score, and T is the number of complete 30-second windows in the recording. Lower values of s_t correspond to observations that are statistically less consistent with the learned background distribution.

Because the numerical scale of GMM log-likelihood depends on the fitted model, the raw value of s_t does not by itself provide a directly interpretable measure of extremity. PBSM therefore expresses each score relative to the empirical score distribution of the full-duration no-contact background observations. Let

$$S_i^B = \log p(\mathbf{x}_i^B \mid \mathcal{B}), \qquad i = 1, \ldots, n, \tag{3}$$

denote the GMM log-likelihood scores of the n full-duration no-contact background observations. Their empirical cumulative distribution is

$$\widehat{F}_B(s) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{S_i^B \leq s\}, \tag{4}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function. This provides an empirical rank-based interpretation related to methods using reference-score distributions (Bates et al., 2023). The Empirical Background Anomaly Percentile (EBAP) of an evaluated score s is then

$$\mathrm{EBAP}(s) = 100\, \widehat{F}_B(s). \tag{5}$$

Thus, $\mathrm{EBAP}(s)$ is the percentage of empirical full-duration background scores less than or equal to s. Low EBAP therefore identifies an observation as statistically unusual relative to the empirical full-duration background distribution. EBAP is a descriptive background-relative percentile, not a formally calibrated conformal p-value or probability of vessel presence.

A 30-second observation is identified as an acoustic departure when EBAP $\leq 5\%$. Sensitivity checks using alternative EBAP thresholds did not materially change the qualitative results, and the 5% criterion was retained as the fixed threshold for the reported analysis. A recording contains a PBSM

acoustic departure when at least one window satisfies this criterion. To examine sensitivity to this recording-level criterion, departure-window counts and within-recording departure persistence were also examined. These analyses preserved separation where PBSM departure was informative and remained saturated in station-seasons where the recording-level measure was saturated, indicating that the observed saturation was not solely a consequence of the at-least-one-window criterion.

Importantly, EBAP expresses a 30-second score relative to the empirical full-duration background-score distribution; it does not remove the temporal-representation difference between the approximately 11-minute background observations and 30-second evaluation windows. Consecutive 30-second windows may also exhibit temporal dependence; PBSM evaluates each window independently and does not explicitly model dependence between adjacent observations. The matched comparison is therefore between identically represented 30-second no-contact (i.e., NC) and one-contact (i.e., OC) observations. For corresponding 30-second representations,

$$s_{30s}^{NC} = \log p(\mathbf{x}_{30s}^{NC} \mid \mathcal{B}), \qquad s_{30s}^{OC} = \log p(\mathbf{x}_{30s}^{OC} \mid \mathcal{B}), \tag{6}$$

where both $\mathbf{x}_{30s}^{NC}$ and $\mathbf{x}_{30s}^{OC}$ are obtained using the same frozen PCA and standardization transformation and are scored against the same local background model $\mathcal{B}$. Their corresponding EBAP values are therefore compared directly.

The held-out 30-second no-contact population measures departure attributable to the change in temporal representation under operational no-contact conditions. Additional departure in the one-contact population therefore represents departure beyond this matched control (Figure 2). The lowest-support window in each recording is retained for subsequent temporal comparison with AIS-derived CPA.

Complete EBAP capture behaviour across all eight station-season datasets is provided in Appendix B.

3.5 Spectral Characterization

A low EBAP establishes statistical departure but not substantive spectral structure. The exact 30-second windows identified by PBSM are therefore characterized spectrally by comparing their linear PSD with the median full-duration no-contact background PSD. *Spectral excess* is the per-frequency-bin difference in decibels between these PSDs, evaluated at 3 dB and 6 dB. These levels provide complementary measures of spectral excess, corresponding approximately to twofold and fourfold increases in linear spectral power relative to the median background, respectively.

Substantive spectral excess at either level is defined as at least 10 frequency bins exceeding the median background by that amount. These criteria characterize the magnitude and spectral extent of PBSM departures. For one-contact recordings, substantive spectral excess among PBSM acoustic departures is summarized as a function of AIS-derived CPA and compared with its prevalence among held-out no-contact recordings exhibiting PBSM acoustic departure.

Selected departure windows are additionally examined using time–frequency spectrograms of the original audio as a targeted qualitative inspection of the acoustic structure underlying the quantitative results. Complete spectral-evidence results across all eight station-season datasets are provided in Appendix D.

3.6 AIS Validation and Physical Interpretation

AIS-derived vessel geometry is introduced only after blind acoustic scoring. One-contact recordings are evaluated as a function of independently calculated CPA using 0–5, 5–10, 10–15, 15–20, and greater-than-20 km bins. These bins provide a common geometric reference for determining whether the prevalence of PBSM acoustic departure and substantive spectral excess varies systematically with vessel proximity.

AIS determines the no-contact and one-contact experimental populations, but vessel position, range, CPA, aspect, course, speed, and identity are not PBSM features and do not determine which windows are identified as departures. Geometry therefore provides an independent physical reference for acoustic departure and substantive spectral excess. The held-out no-contact population provides the

corresponding operational control. Vessel-proximity results across all eight station-season datasets are provided in Appendix C.

3.7 Study Site and Dataset

The dataset comprises four fixed-sensor stations (Stations 2, 5, 6, and 20) evaluated across the 2015–2016 and 2016–2017 observation periods, yielding eight station-season datasets. The station-season datasets are not balanced: the amount and quality of available acoustic data vary across stations and observation periods. This heterogeneity is retained deliberately; datasets are not balanced, subsampled, or selected according to acoustic quality, detectability, or PBSM performance. The resulting evaluation therefore reflects the operational data available at each station-season rather than a curated benchmark population. Locations are shown in Appendix A (Figure 3).

Within each station-season, no-contact recordings are partitioned at the FLAC recording level into training and held-out sets before model fitting, preventing observations from the same recording from contributing to both. A separate background is learned for each station-season under the same frozen experimental configuration and analysis procedure. This standardization provides a methodological control against analysis-dependent variation and supports repeatable comparison across datasets (Trochim et al., 2016).

4 Results

4.1 Background-Relative Acoustic Departure

For Station 2 (2015–2016), 5.99% of held-out full-duration no-contact recordings fell below the 5% EBAP criterion. When the same held-out recordings were evaluated as consecutive 30-second windows, 56.86% of the individual windows were identified as acoustic departures, compared with 64.31% of one-contact windows. Thus, temporal representation accounts for a substantial component of short-duration departure, while the one-contact population exhibits additional departure beyond the temporally matched no-contact control.

At the recording level, requiring at least one 30-second acoustic departure, 74.86% of matched held-out no-contact recordings and 81.08% of one-contact recordings were captured. Figure 2 shows this relationship across the examined EBAP range. Full-duration held-out no-contact recordings remain closest to the empirical background reference, while the matched 30-second no-contact population exhibits substantially greater capture.

4.2 Acoustic Characterization and Vessel Proximity

PBSM detection decreased systematically with independently derived vessel CPA, from 100.0% within 5 km to 95.4% at 5–10 km, 91.4% at 10–15 km, 85.5% at 15–20 km, and 78.6% beyond 20 km.

Among PBSM-detected one-contact recordings within 5 km, 96.8% satisfied the 3-dB spectral-excess criterion and 95.2% satisfied the 6-dB criterion. The 6-dB prevalence decreased with CPA to 91.1% at 5–10 km, 83.5% at 10–15 km, 73.0% at 15–20 km, and 54.7% beyond 20 km. Among PBSM-detected held-out no-contact recordings, the corresponding acoustic reference was 67.8% at 3 dB and 50.9% at 6 dB. Thus, substantive spectral excess is not unique to AIS-confirmed vessel recordings. However, its substantially greater prevalence among one-contact PBSM departures, particularly at smaller CPA, shows that these departures contain stronger spectral evidence than identically processed no-contact departures. The spectral comparison therefore provides additional acoustic discrimination beyond PBSM departure alone.

Complete spectral-evidence results across all eight station-season datasets are provided in Appendix D.

Representative spectrograms of PBSM-identified departures are provided in Appendix E for targeted qualitative interpretation of the acoustic structure underlying the population-level results.

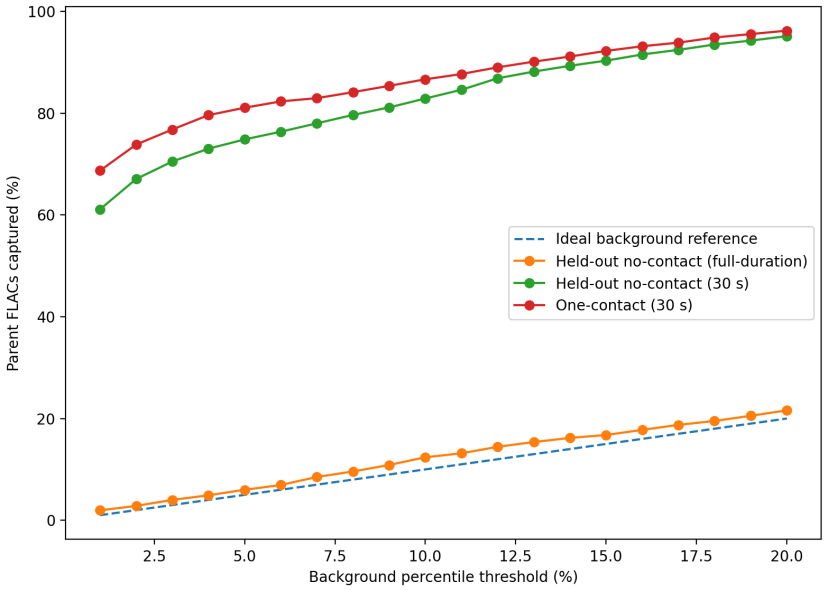


Figure 2: Recording-level EBAP capture as a function of the background percentile threshold for the hybrid-blind PBSM experiment. The full-duration held-out no-contact population remains close to the empirical background reference. Representing the same held-out no-contact recordings as consecutive 30-second windows substantially increases the proportion of recordings containing at least one acoustic departure, demonstrating the effect of temporal representation.

Table 1: Cross-station and cross-season comparison of PBSM acoustic departure at the fixed 5% EBAP operating point. Full NC reports the percentage of held-out full-duration no-contact recordings identified as departures. 30-s NC and one-contact window columns report departure prevalence among individual 30-second windows. The recording-level columns report the percentage of recordings containing at least one 30-second departure. Enrichment is the ratio of one-contact to matched 30-second no-contact recording-level departure prevalence.

Station	Season	Full NC (%)	30-s NC windows (%)	One-contact windows (%)	30-s NC recordings (%)	One-contact recordings (%)	Enrichment
2	2015–16	5.99	56.86	64.31	74.86	81.08	1.083×
2	2016–17	6.30	64.08	70.22	83.68	88.71	1.060×
5	2015–16	7.42	99.35	99.70	100.00	100.00	1.000×
5	2016–17	6.36	16.43	37.85	25.37	49.07	1.935×
6	2015–16	6.17	100.00	100.00	100.00	100.00	1.000×
6	2016–17	6.88	99.98	99.96	100.00	100.00	1.000×
20	2015–16	6.33	97.71	98.21	100.00	100.00	1.000×
20	2016–17	5.38	86.68	93.57	98.85	99.81	1.010×

4.3 Cross-Station and Cross-Season Comparison

To evaluate the behaviour of the frozen PBSM framework across heterogeneous operational environments, the same experimental configuration was applied independently to all eight station-season datasets. Table 1 summarizes background-relative departure prevalence at the fixed 5% EBAP operating point for the full-duration held-out no-contact population, the matched 30-second no-contact population, and the one-contact population. Complete EBAP capture curves across the examined thresholds are provided in Appendix B.

Across all station-season datasets, the held-out full-duration no-contact population remains comparatively close to the nominal 5% EBAP operating point. In contrast, departure prevalence under the 30-second representation varies substantially among station-seasons, demonstrating that short-duration background variability is strongly dependent on the local operational acoustic environment.

The degree of additional departure observed in the one-contact population also varies across datasets. Station 5 (2016–2017) exhibits the strongest recording-level separation between one-contact and matched 30-second no-contact observations, while several station-seasons approach complete short-duration departure saturation. In these saturated datasets, PBSM departure alone provides little discrimination between the two populations. This motivates the subsequent evidence stages: spectral characterization determines whether one-contact departures contain greater substantive spectral excess than matched no-contact departures, while independently derived vessel geometry determines whether that acoustic evidence varies systematically with vessel proximity.

5 Discussion

Temporal representation materially affects background-relative departure. Short-duration no-contact observations can depart substantially from a background learned from full-duration recordings, making the matched 30-second no-contact population the necessary reference for identifying additional departure in the one-contact population.

The station-season results demonstrate that this behaviour depends on the local operational acoustic environment. One plausible hypothesis is that PBSM separation depends on the acoustic propagation and background conditions present during a particular deployment, including sensor placement and depth, local bathymetry, weather and sea state, and other sources of acoustic activity (Urick, 1983). The marked change at Station 5 is particularly informative: the same station changes from near-complete short-duration saturation in 2015–2016 to the strongest one-contact versus matched no-contact separation in 2016–2017, despite the same frozen experimental procedure. This indicates that saturation is not simply a fixed property of the station and instead points toward station-season-specific acoustic conditions.

Two directions follow directly from these results. First, the marked station-season variation at Station 5 motivates investigation of the acoustic propagation, environmental, and deployment conditions associated with PBSM separation. Second, substantive spectral excess provides additional acoustic discrimination where background-relative departure alone is insufficient, motivating deeper investigation of the frequency structure, magnitude, and persistence of this spectral evidence and its relationship with vessel proximity. Richer acoustic representations, including learned encoder embeddings and contemporary out-of-distribution methods, could further test whether this information can be captured more effectively than by the present PCA-based representation. A further statistical extension is formal calibration of EBAP using a separate background calibration partition and finite-sample rank-based or conformal formulations.

6 Conclusion

This work demonstrates a background-first approach for identifying acoustic observations that warrant further investigation within operational passive-acoustic datasets. PBSM models the local no-contact acoustic background using a Gaussian Mixture Model and identifies short-duration observations that depart from its learned statistical support. EBAP provides an interpretable measure of their extremity relative to the empirical background, while the matched no-contact control explicitly accounts for the substantial effect of temporal representation.

PBSM acoustic departure is meaningful when interpreted together with independent acoustic and physical evidence. Substantive spectral excess establishes that identified departures can contain measurable spectral structure beyond the operational background, while AIS-derived vessel geometry determines whether that evidence varies with vessel proximity. Where these measures converge, increasingly strong acoustic evidence is observed as vessel CPA decreases, despite vessel geometry playing no role in PBSM scoring.

The principal contribution is therefore a background-first framework for acoustic observability: learn the local operational background, identify departures blindly, and determine whether those departures are supported by independent spectral and physical evidence.

REPRODUCIBILITY STATEMENT

The experimental procedure was designed to support reproducibility across independent station-season datasets (Trochim et al., 2016). All datasets are processed using the same preprocessing procedures, frozen PBSM configuration, EBAP criterion, spectral characterization criteria, and analysis pipeline, with no station- or season-specific tuning. A separate background model is learned for each station-season using only its no-contact training observations. The methodological details required to reproduce the analysis are provided in the main paper, with complete station-season results and supporting analyses provided in the appendices.

AI USE STATEMENT

Generative AI (OpenAI's ChatGPT) was used during the development of this work as a programming, analytical, and writing assistant. It assisted with Python implementation and revision, debugging, development and review of experimental analysis routines, discussion and interpretation of experimental results, literature search and review, and drafting and editing portions of the manuscript. It was also used to critically examine aspects of the experimental methodology and presentation of the results.

The research direction, development of the PBSM and EBAP framework, experimental design and methodological decisions, execution and validation of the experiments, interpretation of the final results, and scientific conclusions remain the responsibility of the author. All AI-assisted code, analysis, claims, citations, and written material were reviewed and, where applicable, tested or independently verified by the author. The author takes responsibility for the final content of this work, including all material produced with the assistance of generative AI.

REFERENCES

Muhammad Azeem Aslam, Lefang Zhang, Xin Liu, Muhammad Irfan, Yimei Xu, Na Li, Ping Zhang, Jiangbin Zheng, and Yaan Li. Underwater sound classification using learning based methods: A review. *Expert Systems with Applications*, 255:124498, 2024. doi: 10.1016/j.eswa.2024.124498.

Stephen Bates, Emmanuel Candès, Lihua Lei, Yaniv Romano, and Matteo Sesia. Testing for outliers with conformal p-values. *The Annals of Statistics*, 51(1), February 2023. ISSN 0090-5364. doi: 10.1214/22-aos2244. URL http://dx.doi.org/10.1214/22-AOS2244.

Christopher M. Bishop and Hugh Bishop. *Deep Learning: Foundations and Concepts*. Springer, 1 edition, 2024. ISBN 978-3-031-45467-7. doi: 10.1007/978-3-031-45468-4.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1): 1–22, 09 1977. ISSN 0035-9246. doi: 10.1111/j.2517-6161.1977.tb01600.x. URL https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.

E.C. Ifeachor and B.W. Jervis. *Digital Signal Processing: A Practical Approach*. Electronic systems engineering series. Prentice Hall, 2 edition, 2002. ISBN 9780201596199. URL https://books.google.ca/books?id=slHJIO9xmcEC.

Muhammad Irfan, Zheng Jiangbin, Shahid Ali, Muhammad Iqbal, Zafar Masood, and Umar Hamid. Deepship: An underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification. *Expert Systems with Applications*, 183:115270, 2021. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2021.115270. URL https://www.sciencedirect.com/science/article/pii/S0957417421007016.

Farid Jedari-Eyvazi, Fabio Soares Frazao, William D. Halliday, Romina Gehrmann, Laurelie Menelon, Jacob T. Dingwall, Chris Whidden, and Michael Dowd. Advancing passive acoustic ship detection in the western canadian arctic: Signal processing and deep learning approaches. *The Journal of the Acoustical Society of America*, 159(6):4843–4861, 2026. doi: 10.1121/10.0043937.

Lawrence E. Kinsler, Austin R. Frey, Alan B. Coppens, and James V. Sanders. *Fundamentals of Acoustics*. Wiley, 4th edition, 2000.

Ziyi Liang, Matteo Sesia, and Wenguang Sun. Integrative conformal p-values for powerful out-of-distribution testing with labeled outliers, 2022. URL https://arxiv.org/abs/2208.11111.

Maria Emanuela Mihailov. Characterization and automated classification of underwater acoustic environments in the western black sea using machine learning techniques. *Journal of Marine Science and Engineering*, 13(7), 2025. ISSN 2077-1312. doi: 10.3390/jmse13071352. URL https://www.mdpi.com/2077-1312/13/7/1352.

Kevin P. Murphy. *Probabilistic Machine Learning: Advanced Topics*. MIT Press, 2023. ISBN 9780262048439.

Najamuddin Najamuddin, Usman Ullah Sheikh, and Ahmad Ameri. Capse-vit: A lightweight framework for underwater acoustic vessel classification using coherent spectral estimation and modified vision transformer. *Archives of Acoustics*, 50(2), 05 2025. doi: 10.24425/aoa.2025.153662.

Wayne Renaud, Priyanka Aravindan, and Gabriel Spadon. From continuous deployment to queryable dataset: Terabyte-scale ais-aligned passive acoustic labelling, 2026. URL https://arxiv.org/abs/2607.13840.

David Santos-Domínguez, Soledad Torres-Guijarro, Antonio Cardenal-López, and Antonio Pena-Gimenez. Shipsear: An underwater vessel noise database. *Applied Acoustics*, 113:64–69, 2016. ISSN 0003-682X. doi: https://doi.org/10.1016/j.apacoust.2016.06.008. URL https://www.sciencedirect.com/science/article/pii/S0003682X16301566.

Allyson A. Silva, Lisandro Lovisolo, and Tadeu N. Ferreira. Machine Learning Ship Classifiers for Signals from Passive Sonars. *Applied Sciences*, 15(13):6952, 2025. ISSN 2076-3417. doi: 10.3390/app15136952.

Jessica M. Topple, Dugald J. Thomson, Jeffrey R. Bates, Jinshan Xu, Aaron J. L. Webstey, and Carolyn M. Binder. Passive Acoustic Monitoring of Arctic Surface Ships with DeepPAM. In *OCEANS'24 - Halifax*, pp. 1–6, DRDC-ARC, 2024. Oceans'24. doi: 10.1109/OCEANS55160.2024.10754032.

William M. K. Trochim, James P. Donnelly, and Kanika Arora. *Research Methods: The Essential Knowledge Base*. Cengage Learning, Boston, MA, 2 edition, 2016. URL https://www.cengage.ca/. Cengage eTextbook edition, accessed via Cengage Dashboard (licensed resource).

R.J. Urick. *Principles of Underwater Sound*. Peninsula Publishing, Los Altos CA, 3 edition, 1983.

A Study Site and Sensor Locations

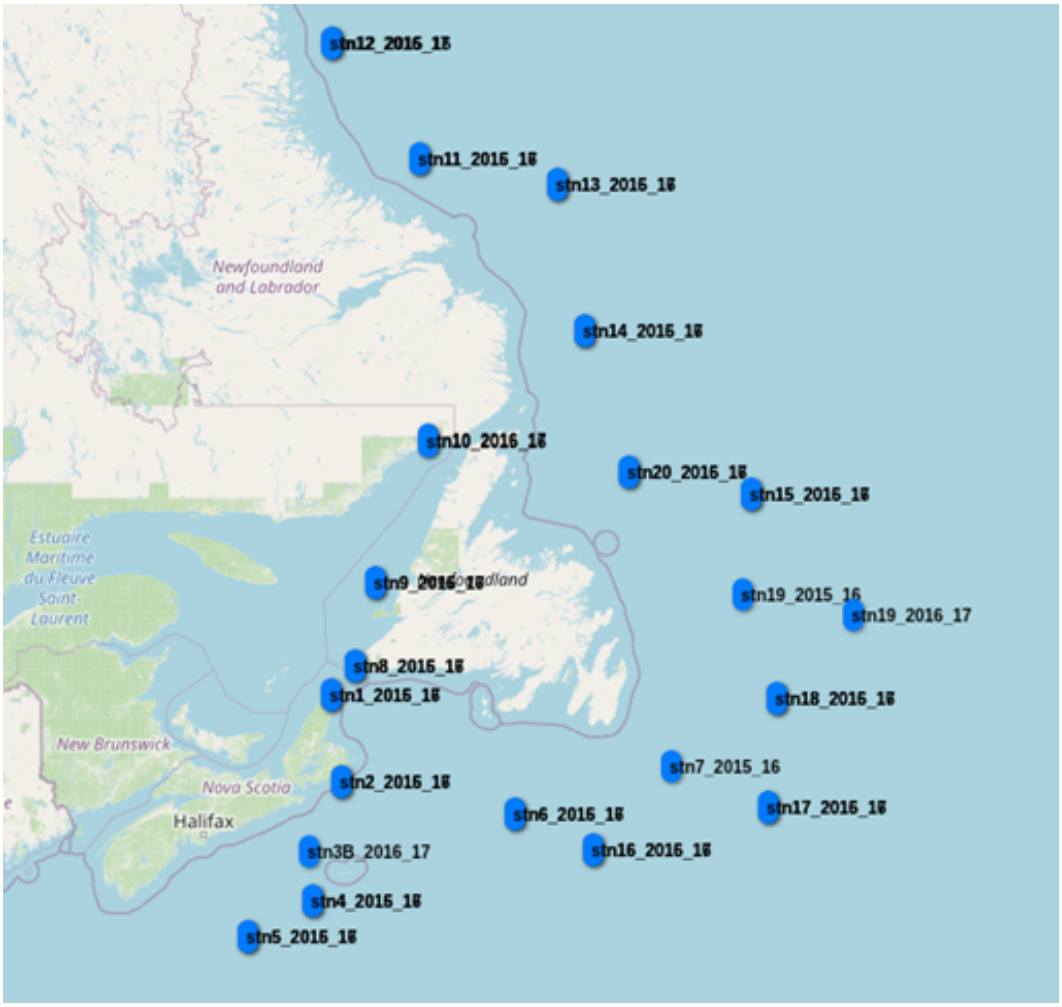
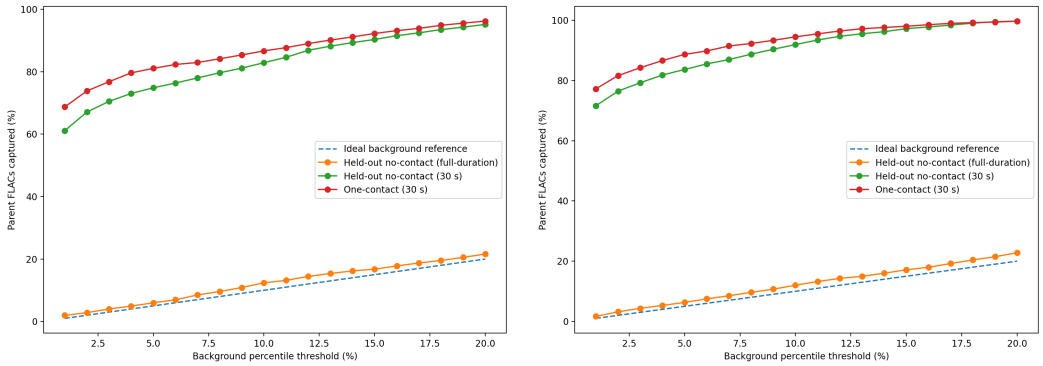


Figure 3: Geographic distribution of the passive acoustic monitoring stations in the study region. Stations 2, 5, 6, and 20 were selected to provide geographically distinct monitoring locations and are evaluated across the 2015–2016 and 2016–2017 observation periods.

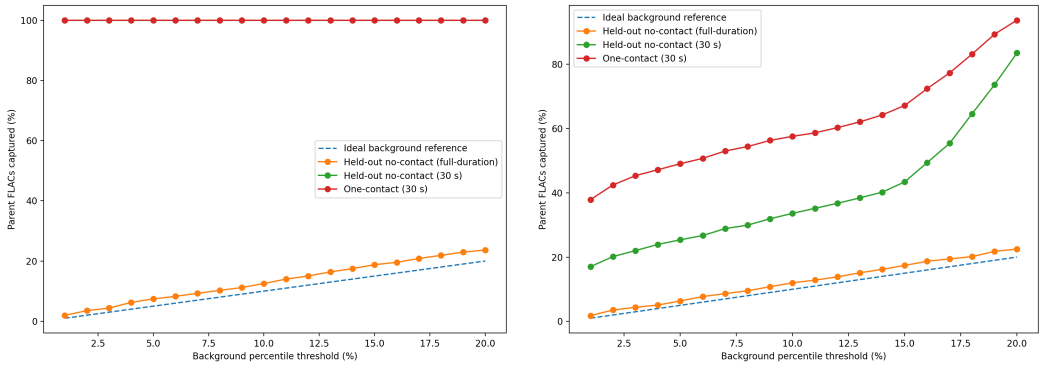
B EBAP

Figures 4–7 show the complete EBAP capture behaviour for all eight station-season datasets. The full-duration held-out no-contact population remains close to the empirical background reference across datasets, while the 30-second populations reveal substantial station-season dependence. Station 2 exhibits consistent separation between one-contact and matched 30-second no-contact recordings in both observation periods. Station 5 changes markedly between seasons: the 2015–2016 result is effectively saturated, whereas 2016–2017 exhibits strong separation. Station 6 is saturated in both observation periods, providing little separation between the two 30-second populations. Station 20 is similarly saturated in 2015–2016 and approaches saturation in 2016–2017, although some separation remains at lower EBAP thresholds. These results show that short-duration background departure is strongly dependent on the local operational acoustic environment and that PBSM departure alone does not provide uniform separation across station-seasons.



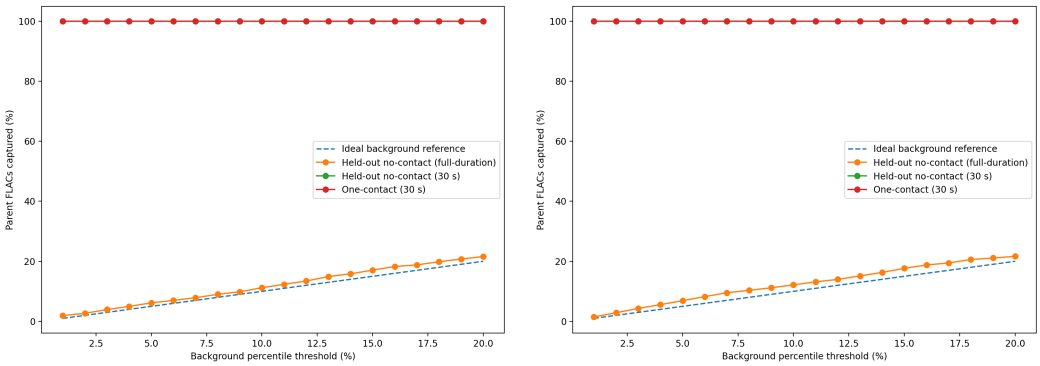
(a) Station 2 2015–2016 **(b) Station 2 2016–2017**

Figure 4: EBAP capture behaviour for Station 2. Both observation periods show separation between the one-contact and matched 30-second no-contact populations, while the full-duration held-out no-contact population remains close to the empirical background reference.



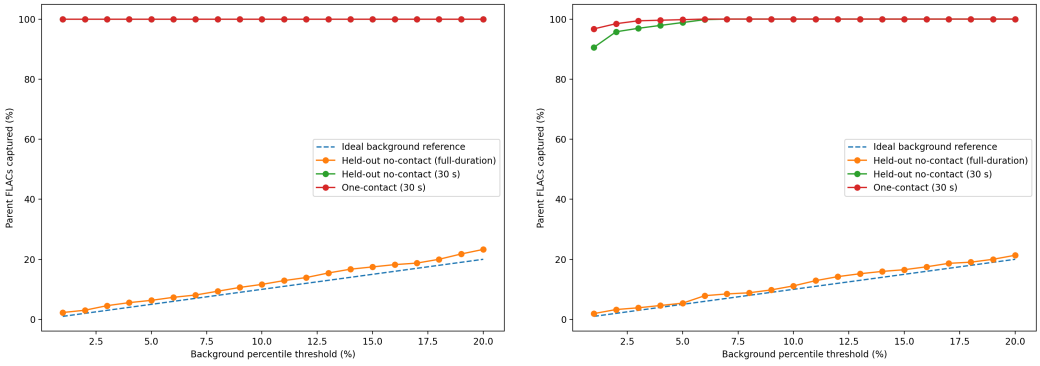
(a) Station 5 2015–2016 **(b) Station 5 2016–2017**

Figure 5: EBAP capture behaviour for Station 5. The 2015–2016 30-second populations are effectively saturated and provide little separation, whereas 2016–2017 exhibits strong separation between one-contact and matched 30-second no-contact recordings.



(a) Station 6 2015–2016 **(b) Station 6 2016–2017**

Figure 6: EBAP capture behaviour for Station 6. In both observation periods, the matched 30-second no-contact and one-contact populations are effectively saturated, providing little separation from PBSM departure alone.



(a) Station 20 2015–2016 **(b) Station 20 2016–2017**

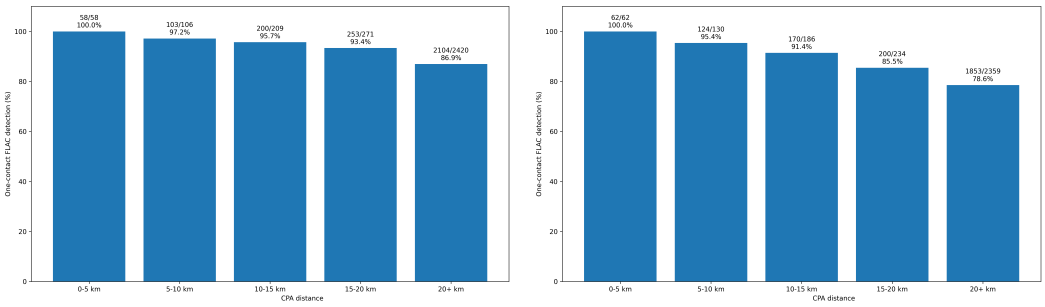
Figure 7: EBAP capture behaviour for Station 20. The 2015–2016 30-second populations are effectively saturated. The 2016–2017 populations also approach saturation, although separation between one-contact and matched no-contact observations remains visible at lower EBAP thresholds.

C VESSEL PROXIMITY AND PBSM ACOUSTIC DEPARTURES

Vessel proximity analysis provides an independent physical reference for interpreting the acoustic departures identified by PBSM. Whereas EBAP establishes whether an acoustic observation departs from the locally learned background, vessel proximity provides external geometric context for those departures using independently derived AIS information. The analysis therefore examines whether the strongest blindly identified acoustic departures exhibit a physical relationship with AIS reported vessel proximity.

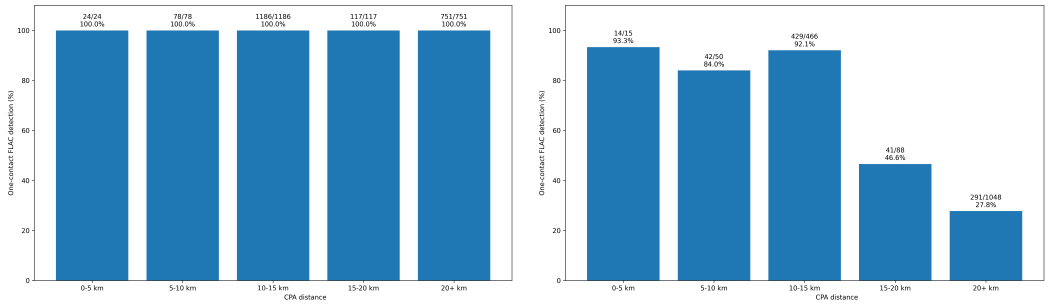
Figures 8–11 examine this relationship for the one-contact population across all eight station-season datasets. For each one-contact recording, the 30-second window exhibiting the greatest acoustic departure is identified, and the centre time of that window is associated with the nearest AIS vessel report in time. The vessel position from that report is then dead reckoned to the acoustic-departure time, providing an approximate vessel-to-sensor distance when the greatest acoustic departure occurred. This acoustic-departure distance is compared with the independently calculated closest point of approach (CPA), which represents the minimum distance between the AIS-derived vessel trajectory and the sensor. The distance bins provide a common geometric reference for these two quantities. For example, correspondence in the 0–5 km bin indicates that both the vessel distance associated with the strongest acoustic departure and the independently calculated CPA place the vessel within that distance range. The percentages reported for each bin quantify how often this correspondence occurs across the one-contact recordings.

The resulting correspondence must be interpreted in the context of the EBAP behaviour shown in Appendix B. Station 2 retains substantial variation in PBSM acoustic departure, providing a more selective comparison with independently calculated vessel proximity. Station 5 shows contrasting behaviour between observation periods, while Station 6 and Station 20 are saturated or near-saturated at the fixed EBAP operating point. In these saturated datasets, acoustic departures occur throughout most one-contact recordings, limiting the information provided by departure occurrence alone. The proximity analysis nevertheless establishes whether the strongest blindly identified acoustic departure occurs at a vessel distance geometrically consistent with the independently calculated closest approach. Across the station-seasons, these results therefore provide an independent physical check on PBSM acoustic departure while also demonstrating where short-duration EBAP saturation limits the strength of that evidence.



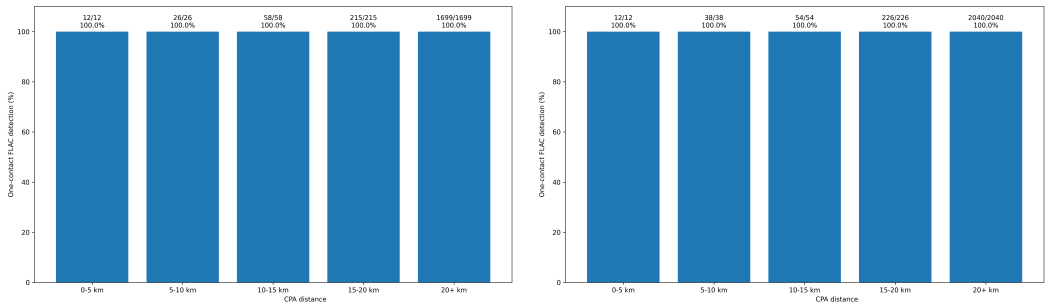
(a) Station 2 2015–2016 **(b) Station 2 2016–2017**

Figure 8: PBSM recording-level departure prevalence as a function of independently derived vessel CPA distance for the Station 2 one-contact population. Both observation periods exhibit decreasing PBSM departure prevalence with increasing CPA, providing a physical proximity gradient where the fixed EBAP operating point retains variation.



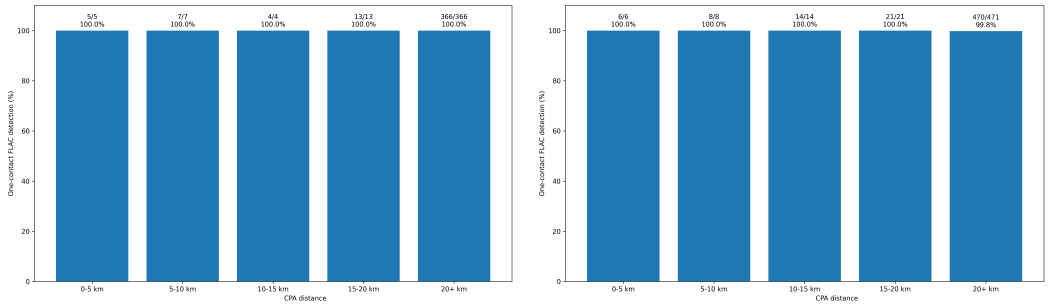
(a) Station 5 2015–2016 **(b) Station 5 2016–2017**

Figure 9: PBSM recording-level departure prevalence as a function of independently derived vessel CPA distance for the Station 5 one-contact population. The 2015–2016 result remains nearly constant across CPA because PBSM departure is already saturated, whereas 2016–2017 exhibits a pronounced relationship between departure prevalence and vessel proximity.



(a) Station 6 2015–2016 **(b) Station 6 2016–2017**

Figure 10: PBSM recording-level departure prevalence as a function of independently derived vessel CPA distance for the Station 6 one-contact population. Departure prevalence remains at or near 100% across CPA bins in both observation periods because the fixed EBAP criterion is already saturated. The flat CPA response therefore reflects saturation of PBSM departure rather than evidence of uniform vessel detectability across distance.



(a) Station 20 2015–2016 **(b) Station 20 2016–2017**

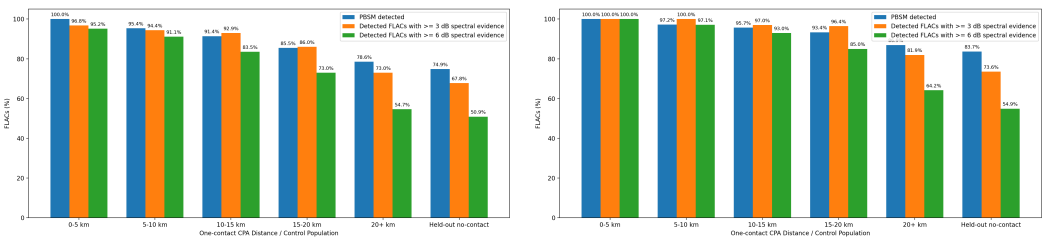
Figure 11: PBSM recording-level departure prevalence as a function of independently derived vessel CPA distance for the Station 20 one-contact population. Both observation periods are saturated or near-saturated at the fixed EBAP operating point, leaving little remaining variation for CPA to stratify. Values near 100% therefore represent the prevalence of PBSM acoustic departure within the one-contact population rather than a 100% vessel-detection rate.

D SPECTRAL EVIDENCE OF PBSM ACOUSTIC DEPARTURES

Spectral evidence provides an independent acoustic reference for interpreting the acoustic departures identified by PBSM. Whereas EBAP establishes whether an acoustic observation departs from the locally learned background, spectral analysis examines whether those departures contain *substantive spectral excess* relative to the no-contact background. The analysis therefore characterizes the acoustic structure present within blindly identified PBSM departures.

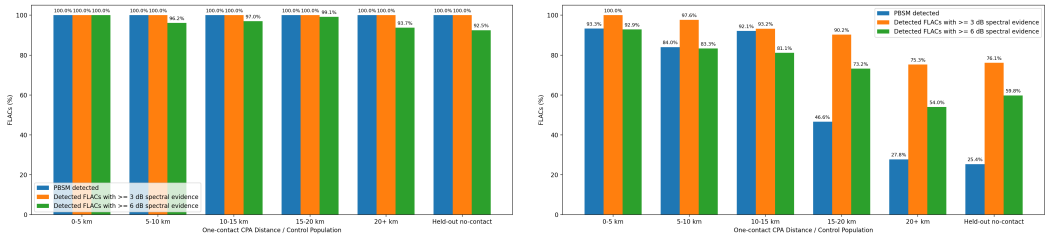
Figures 12–15 extend the vessel-proximity analysis of Appendix C across all eight station-season datasets. The blue bars use the same recording-level PBSM acoustic-departure measure reported in the preceding analysis: within each CPA distance bin, they show the percentage of one-contact recordings containing at least one 30-second acoustic departure at the fixed EBAP operating point. Among those recordings containing a PBSM acoustic departure, the 3 dB and 6 dB bars report the percentage that also exhibit substantive spectral excess at the corresponding threshold. Thus, a value of 90% for the 3 dB criterion means that 90% of the one-contact recordings containing a PBSM acoustic departure in that CPA bin also contain a departure window satisfying the 3 dB spectral criterion. Substantive spectral excess is defined as at least 10 frequency bins exceeding the median full-duration no-contact background by the corresponding 3 dB or 6 dB threshold. The spectral percentages are therefore conditional on PBSM acoustic departure and characterize the acoustic content of those departures rather than constituting additional detection criteria.

The final three bars provide the held-out no-contact reference for interpreting the one-contact results. They report the same three quantities: PBSM acoustic departure prevalence and, among those departures, the percentage satisfying the 3 dB and 6 dB spectral criteria. Because the held-out population contains no qualifying AIS contact, these values establish how frequently substantive spectral excess occurs among PBSM departures under operational no-contact conditions. The 6 dB held-out value is therefore an important reference for the 6 dB one-contact results. For example, a one-contact CPA bin in which 100% of PBSM departures satisfy the 6 dB criterion is substantially more informative when only 65.2% of held-out no-contact departures satisfy the same criterion. This difference shows that, even where PBSM departure itself is saturated, the one-contact departures can contain substantially stronger spectral excess than is observed in the held-out no-contact population. The held-out comparison therefore establishes whether the spectral content of PBSM departures provides additional acoustic evidence beyond background departure alone.



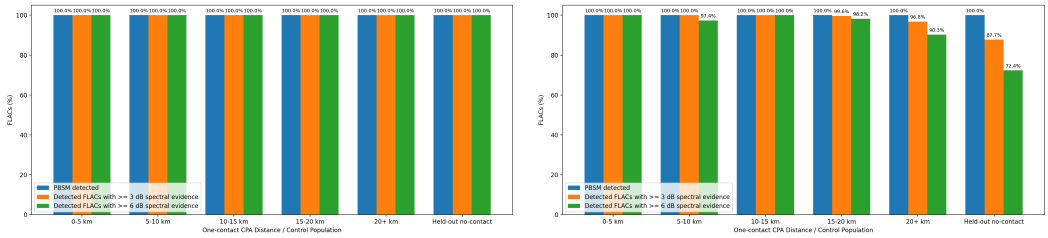
(a) Station 2 2015–2016 **(b) Station 2 2016–2017**

Figure 12: PBSM acoustic departure and spectral evidence for the Station 2 one-contact population across independently derived vessel CPA distance. The blue bars report recording-level PBSM acoustic-departure prevalence; the 3 dB and 6 dB bars report substantive spectral excess among those recordings. The held-out no-contact bars provide the corresponding aggregate acoustic reference.



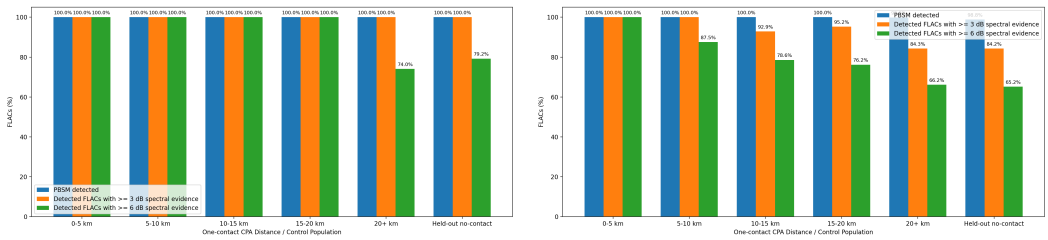
(a) Station 5 2015–2016 **(b) Station 5 2016–2017**

Figure 13: PBSM acoustic departure and spectral evidence for the Station 5 one-contact population across independently derived vessel CPA distance. The held-out no-contact reference distinguishes spectral excess associated with one-contact PBSM departures from spectral excess also present among operational no-contact departures.



(a) Station 6 2015–2016 **(b) Station 6 2016–2017**

Figure 14: PBSM acoustic departure and spectral evidence for the Station 6 one-contact population across independently derived vessel CPA distance. Although PBSM acoustic departure is saturated or near-saturated, comparison with the held-out no-contact spectral reference determines whether the identified departures retain additional measurable spectral structure.



(a) Station 20 2015–2016 **(b) Station 20 2016–2017**

Figure 15: PBSM acoustic departure and spectral evidence for the Station 20 one-contact population across independently derived vessel CPA distance. Comparison with the held-out no-contact reference remains informative despite high PBSM departure prevalence by showing whether substantive spectral excess within one-contact departures exceeds that observed among held-out no-contact departures.

E SPECTROGRAMS OF PBSM ACOUSTIC DEPARTURES

The preceding analyses progressively characterize the acoustic departures identified by PBSM. Appendix B establishes their background-relative statistical behaviour, Appendix C examines their relationship with independently derived vessel proximity, and Appendix D determines whether those departures contain substantive spectral excess relative to the no-contact background. The spectrograms presented here provide a final qualitative examination of that evidence by showing the acoustic structure contained within selected PBSM departure windows.

Examples are selected to represent both strong and weak spectral evidence within the one-contact and held-out no-contact populations. Strong examples satisfy the 6 dB substantive spectral-excess criterion, whereas weak examples are PBSM acoustic departures that do not satisfy that criterion. This distinction is particularly useful because PBSM departure and substantive spectral excess are not equivalent: an observation may depart statistically from the learned background without exhibiting strong spectral excess under the 6 dB criterion. Similarly, the held-out no-contact analysis shows that substantive spectral excess can also occur in operational no-contact conditions.

The spectrograms therefore provide representative examples of the acoustic content underlying the quantitative results. Comparing strong and weak one-contact departures illustrates the spectral differences underlying the 6 dB percentages reported in Appendix D, while the corresponding held-out no-contact examples illustrate the acoustic variability represented by the control population. Together, these examples provide a direct visual interpretation of what PBSM acoustic departure and substantive spectral excess represent in the recorded acoustic data.

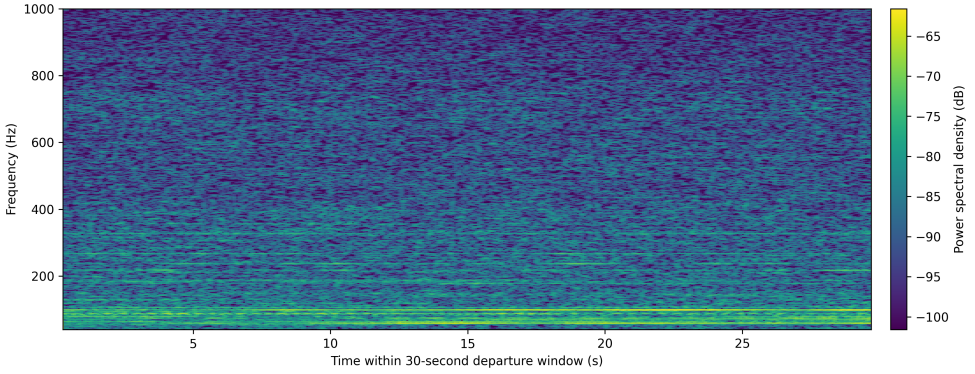


Figure 16: Representative strong one-contact PBSM acoustic departure from Station 2 (2015–2016). The 30-second window has EBAP 0.270%, is associated with a vessel CPA of 3.252 km, and contains 257 frequency bins exceeding the median full-duration no-contact background by at least 6 dB. This example illustrates the acoustic structure associated with a PBSM departure satisfying the 6 dB substantive spectral-excess criterion.

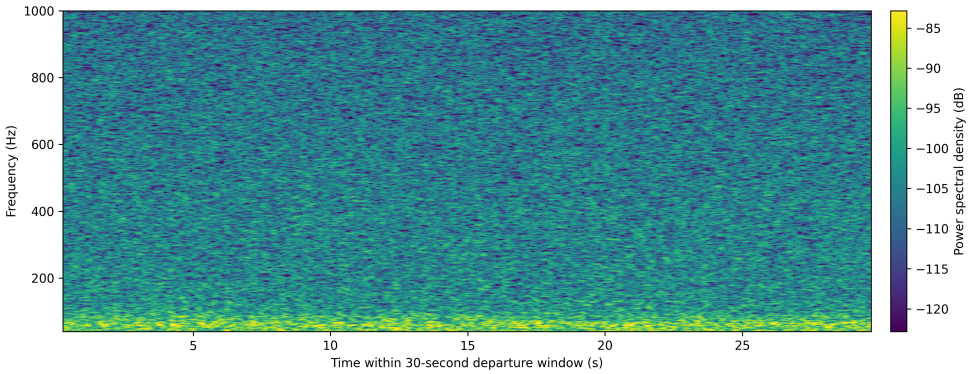


Figure 17: Representative weak one-contact PBSM acoustic departure from Station 6 (2015–2016). The 30-second window has EBAP 0.000%, is associated with a vessel CPA of 0.799 km, and contains no frequency bins exceeding the median full-duration no-contact background by at least 3 dB or 6 dB. Its maximum spectral excess is 1.847 dB. This example illustrates that a strong background-relative PBSM acoustic departure can occur without substantive spectral excess.

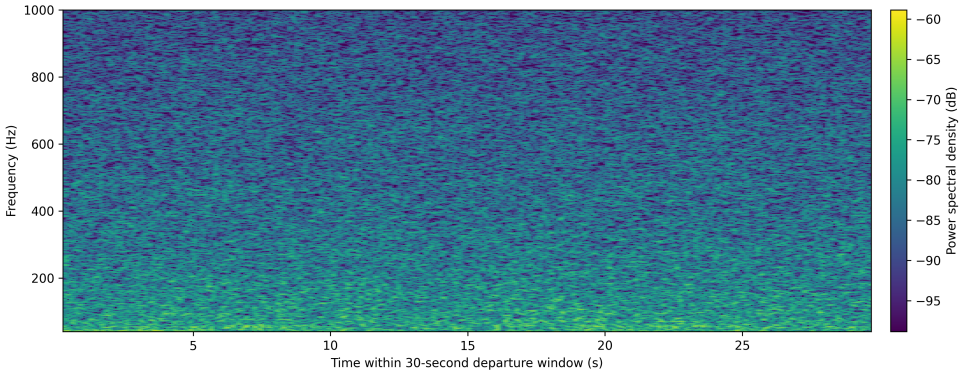


Figure 18: Representative strong held-out no-contact PBSM acoustic departure from Station 20 (2016–2017). The 30-second window has EBAP 2.107% and contains 257 frequency bins exceeding the median full-duration no-contact background by at least 6 dB, with a maximum spectral excess of 16.373 dB. Despite satisfying both the PBSM acoustic-departure and substantive spectral-excess criteria, the spectrogram contains no conspicuous spectral feature. This example illustrates that strong quantitative departure from the learned operational background need not correspond to visually obvious spectral structure in an individual spectrogram.

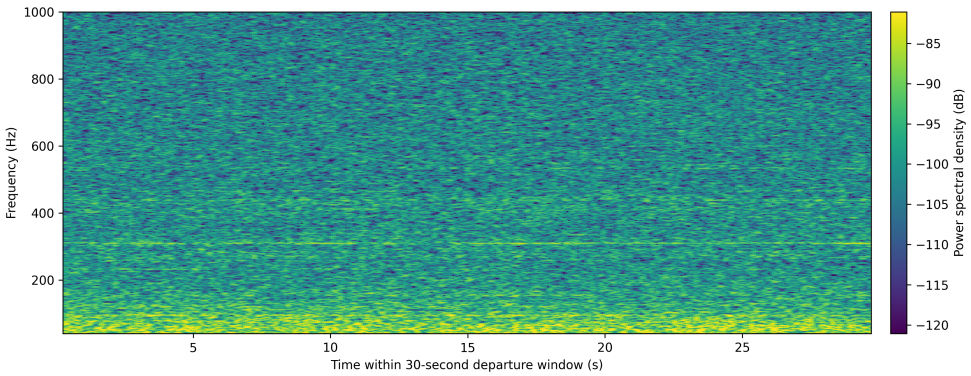


Figure 19: Representative held-out no-contact window from Station 5 (2016–2017) not identified as an acoustic departure by PBSM. The 30-second window has EBAP 5.083%, placing it just above the fixed 5% acoustic-departure operating point. Despite remaining consistent with the learned background under PBSM, the spectrogram contains visible spectral structure, including persistent low-frequency energy and a narrow band near 300 Hz. This example illustrates that no-contact background observations need not be acoustically featureless and that the presence of spectral structure alone does not imply background departure.