WHEN IS RESIDUAL EOF1 CERTIFIABLE?

Anonymous authors
Paper under double-blind review

ABSTRACT

We study pipelines that compare an analysis window with a background reference and interpret the leading analysis mode as an anomaly—for example, a seismic event against ambient noise or a climate pattern against climatology. Write Σ for the analysis-window covariance, C for a quiet reference covariance in the same feature space, and $\Sigma - C$ for their background-adjusted contrast. We show that Σ alone does not identify an anomaly: the same matrix admits pure-background and anomaly-plus-background decompositions, identical at second order when the anomaly lies along the background axis. Under partial contamination, the calibrated direction remains fixed while the whitened eigen-gap required for finite-sample certification shrinks to zero at same-record, where the direction is unidentified. Under independent sub-Gaussian sampling from the rank-one covariance model, a quiet held-out background recovers a residual spike above concentration scale, but that spike need not be EOF1 of Σ. We prove observational equivalence and a sharp gap law, and introduce a Contrast Audit Procedure (CAP) that separates declared guard passes from conditional, theorem-backed direction certificates and otherwise abstains. An optional held-out-control calibration bounds per-record null returns under exchangeability without assuming independent guards. On the public INSTANCE benchmark, with station keys held out and guards frozen, both guards pass $240/299$ earthquakes and $13/264$ noise records. Known-direction injections into the 264 held-out real-noise backgrounds reach 98.1% accurate CAP returns at twice the frozen strength threshold; on real LOCATA recordings, frozen guards return $11/13$ sources versus $1/13$ inactive controls with 2.1° median azimuth error. We release our code, prespecified splits, results, and reproduction instructions in our anonymous repository for replication.

1 INTRODUCTION

Background adjustment is meant to support a consequential decision: whether a reported direction is evidence of an event or merely structured background. A stable leading mode does not resolve that distinction. The reference determines what can be concluded. Analysis covariance alone and same-record calibration cannot certify the label. A quiet held-out reference supports a direction claim only after spike, gap, and sampling conditions clear under thresholds fixed in advance.

EOF/PCA analyses report leading variance modes (Hotelling, 1933; Lorenz, 1956; Preisendorfer, 1988; von Storch & Zwiers, 1999; Jolliffe, 2002). We study the additional interpretive step of labelling such a mode as an anomaly after background removal; those references motivate the leading-mode operation, not that label. Contrastive dimension reduction seeks variation unique to a foreground and now includes a subspace-inclusion test and dimension estimator (Abid et al., 2018; Hawke et al., 2024). CAP instead audits the supplied background's provenance and reports a leading contrast direction only after observable spike and gap margins clear. Reject-option methods control predictive risk or coverage (Chow, 1970; Geifman & El-Yaniv, 2017). Spectral-projector inference quantifies uncertainty around an eigenspace of a posited covariance (Koltchinskii & Lounici, 2017; Naumov et al., 2019). Neither asks whether the reference is quiet or whether both contrast margins are resolved, so CAP withholds the unlabelled direction when either is unsupported. Ambient-noise processing and event detection distinguish seismic background from events, but do not address when a held-out background covariance identifies a residual direction (Campillo & Paul, 2003; Shapiro & Campillo, 2004; Bensen et al., 2007; Mousavi et al., 2020). Climate analyses remove a climatology and report residual EOF1 as a mode (Hannachi et al., 2007; Bretherton et al., 1992; von Storch

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.

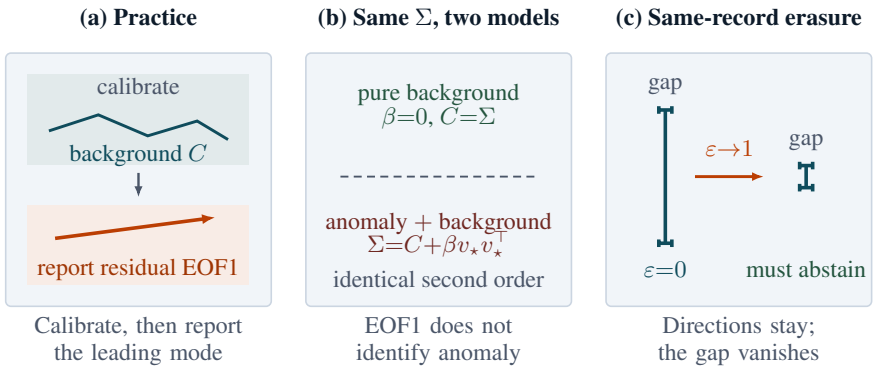

Figure 1: Residual EOF1 is not a certifiable anomaly from covariance alone. **(a)** Remove a background, report residual EOF1 as the anomaly; **(b)** one Σ admits pure-background and anomaly-plus-background decompositions; **(c)** same-record contamination shrinks the calibrated eigen-gap under the gap law of Thm. 3.1, so CAP abstains from certifying a residual direction here.

Table 1: Certification table for residual reports: none and same-record abstain; quiet held-out C passes only after spike and gap guards clear, and certifies only with theorem-backed thresholds.

Side information	Separable?	Aligned case	Finite-sample call
None (Σ only)	unidentified	collapsed	**must abstain**
Same-record ($\hat{C}{=}\hat{\Sigma}$)	unidentified	collapsed	**gap erasure / abstain**
Quiet held-out C	$\Sigma{-}C$ identifies $v_\star$	$\Sigma{-}C$ still recovers $v_\star$	**guard pass / theorem certificate**

& Zwiers, 1999), with classical caveats on EOF sampling (North et al., 1982). For a symmetric $A \in \mathbb{R}^{n\times n}$, order its eigenvalues as $\lambda_1(A) \geq \cdots \geq \lambda_n(A)$ and let $\mathrm{top}(A)$ denote any unit leading eigenvector. Direction claims below require a simple leading eigenvalue; we may plot tied EOFs but never certify a direction from the tie alone. Throughout, Σ is the covariance of the analysis window after domain preprocessing, not an already subtracted covariance; C is a separately supplied reference, and $\Sigma - C$ performs the background adjustment. We retain the application term residual EOF1 for $u_{\mathrm{var}} = \mathrm{top}(\Sigma)$; under the rank-one model the quiet-reference target is $\mathrm{top}(\Sigma - C)$. Figure 1 shows the failure and its consequence for certification: the same analysis covariance supports pure-background and anomaly-plus-background decompositions, while same-record contamination can preserve the plotted direction yet erase the whitened gap. CAP therefore abstains without identifying side information or a resolved gap. Table 1 distinguishes those cases from a quiet-background guard pass, which becomes a theorem-backed certificate only when the finite-sample assumptions and radii hold. The report then states the corresponding angle bound, probability, and assumptions.

Related work. Reporting leading EOFs as variance modes is standard in climate and related fields (Hotelling, 1933; Lorenz, 1956; Preisendorfer, 1988; von Storch & Zwiers, 1999; Hannachi et al., 2007; Jolliffe, 2002; Bretherton et al., 1992; North et al., 1982). Spiked PCA studies detection after a given spike/noise split (Johnstone, 2001; Baik et al., 2005; Baik & Silverstein, 2006; Paul, 2007; Benaych-Georges & Nadakuditi, 2011; Nadler, 2008; Chen et al., 2021). Contrastive PCA extracts $\mathrm{top}(\Sigma - \alpha C)$ (Abid et al., 2018); covariance tests ask whether $\Sigma = C$ (Cai et al., 2013); and optimal fingerprinting uses prescribed background covariance (Hasselmann, 1993; Hegerl et al., 1996; Allen & Tett, 1999; Allen & Stott, 2003; Ribes et al., 2013). CAP leaves the eigensolver unchanged: at $\alpha = 1$ its quiet-reference direction matches contrastive PCA when the leading contrast is positive, and the generalised eigenproblem measures spectral separation. Its contribution is a joint decision rule: reference provenance, residual strength, and background-normalised separation determine whether CAP reports a direction or abstains under uncertainty (Table 4).

Research gap. Analysis covariance Σ and reference estimate $\hat{C}$ can both look well estimated even when no unique anomaly label exists. Residual EOF1 is still often plotted as the anomaly even when quiet C would certify only $\mathrm{top}(\Sigma - C)$; Table 1 states the certify-versus-abstain cut.

Our contributions. **(1)** We formalise analysis/reference covariance decompositions (§2) and prove observational non-uniqueness with aligned collapse (§3). **(2)** We show same-record gap erasure with an explicit gap law $\delta_\star(\varepsilon)$ and finite-sample gap control with Davis–Kahan angle bounds (§3). **(3)** Under the rank-one covariance model and independent sub-Gaussian sampling, we prove quiet certification of $\mathrm{top}(\Sigma - C)$; at population level we distinguish EOF1–CAP disagreement Δ_{CAP} from EOF1–SNR disagreement Δ_{SNR} (§4). **(4)** Our methodological contribution is a decision and reporting layer, not a new spectral estimator: CAP separates fixed-threshold returns from theorem-backed certificates, adds an exchangeable-control calibration, and abstains when either condition is unresolved (§5). **(5)** Independent simulations test contamination and null specificity; held-out seismic, optically tracked acoustic, NOAA/Etna, and TROPOMI studies evaluate frozen guards out of sample across independent domains and measurement systems (§6, App. C).

2 Problem formulation: analysis and reference covariances

We formalise the pipeline in Fig. 1 and Table 1: practice reports a leading analysis-window mode, then interprets it against a background reference. We distinguish three directions. Residual EOF1 $u_{\mathrm{var}} = \mathrm{top}(\Sigma)$ maximises variance. Under the theorem conditions, $\mathrm{top}(\Sigma - C)$ is the population residual direction and CAP target. When the maximiser is unique up to sign, $u_{\mathrm{snr}}(C)$ maximises $u^\top \Sigma u / (u^\top C u)$ over unit vectors. The latter two need not coincide with u_{var}.

Definition 2.1 (Residual decomposition)**.** Let the ambient dimension satisfy $n \geq 2$. A *decomposition* is a tuple $(\beta, v_\star, C)$ with $C \succ 0$, $\beta \geq 0$, and $\|v_\star\| = 1$ satisfying the covariance identity

$$\Sigma \;=\; \beta\, v_\star v_\star^\top \;+\; C. \tag{1}$$

C is the quiet-reference covariance in the same feature space; $v_\star$ is the model's anomaly direction when $\beta > 0$; residual EOF1 is the application label for $u_{\mathrm{var}} = \mathrm{top}(\Sigma)$ in this paper.

Practice pipeline and example. Analysts estimate an analysis covariance Σ on the window of interest and, optionally, a background covariance C. EOF/PCA practice reports u_{var} as a leading variance mode (Jolliffe, 2002; Hannachi et al., 2007); we study the additional, application-dependent step of labelling that mode as the anomaly after background removal. Take $\Sigma = \mathrm{diag}(3, 1)$. One decomposition is pure background ($\beta = 0$, $C = \Sigma$); another is an isotropic spike ($C = I$, $\beta = 2$, $v_\star = e_1$), available because the bulk eigenvalue is flat. Both reproduce the same Σ with residual EOF1 equal to e_1. The label—anomaly or pure background structure—is therefore not determined by second moments of Σ, even though both decompositions return the same leading variance direction.

Quiet versus same-record backgrounds. We distinguish backgrounds by the sample that supplies C. A *quiet* background is a held-out C that does not contain the anomaly of interest (e.g., ambient noise before an event). Practice often instead reuses the analysis window or a climatology that still contains that anomaly (*same-record*). We model contamination by the explicit path

$$C_\varepsilon \;=\; C + \varepsilon\beta\, v_\star v_\star^\top, \qquad \Sigma \;=\; C + \beta\, v_\star v_\star^\top, \qquad \varepsilon \in [0, 1], \tag{2}$$

so $\Sigma = C_\varepsilon + (1-\varepsilon)\beta\, v_\star v_\star^\top$: as $\varepsilon \to 1$, the residual anomaly strength relative to the contaminated background shrinks to zero. Exact same-record is $\varepsilon = 1$ ($C \leftarrow \Sigma$); quiet held-out background is $\varepsilon = 0$. Table 1 contrasts quiet $C = I$ with same-record $C \leftarrow \Sigma$ in the worked example.

Spiked background. Aligned collapse needs a background with a preferred axis v_c and strength $\gamma > 0$, written $C = I + \gamma v_c v_c^\top$. Let $\alpha := |\langle v_\star, v_c \rangle|$ be the share of the anomaly on that axis, and $s := \sqrt{1 - \alpha^2}$ the orthogonal share; full alignment is $s = 0$. We use this spiked form only for Prop. 3.2; Prop. 3.1 already gives two decompositions of every covariance $\Sigma \succ 0$.

Possible outcomes. CAP distinguishes a fixed-threshold guard pass, a theorem-backed certificate, no residual at the declared resolution, and abstention. It abstains when Σ alone cannot identify the anomaly or when the whitened gap is unresolved; only the certificate carries the stated finite-sample guarantee. Sections 3–5 establish the three cases in Table 1, and App. A gives the full conditions, inferential bounds, and reporting obligations for each regime used throughout the paper.

3 Non-identifiability and same-record gap erasure

Covariance is non-identifying (Props. 3.1–3.2); contamination preserves the direction while erasing the gap (Thms. 3.1–3.2); and concentration determines finite-sample resolution (Thm. 3.3).

Observational non-uniqueness. From the eigendecomposition we obtain the two decompositions in Prop. 3.1; in the spiked family, the second-moment ambiguity is exact at full alignment ($s = 0$).

Proposition 3.1 (Non-uniqueness). *For any $\Sigma \succ 0$ there exist at least two decompositions reproducing Σ:* (i) *pure background $\beta = 0$, $C = \Sigma$;* (ii) *top-eigenvalue split $\beta = \lambda_{\max}(\Sigma)/2$, $v_\star = u_{\mathrm{var}}$, and $C = \Sigma - \beta u_{\mathrm{var}} u_{\mathrm{var}}^\top$ (hence $C \succ 0$). These differ in β for every $\Sigma \succ 0$. If $\lambda_{\min}(\Sigma) < \lambda_{\max}(\Sigma)$ and $\lambda_2(\Sigma) = \cdots = \lambda_n(\Sigma) = \lambda_{\min}(\Sigma)$, the choice $\beta = \lambda_{\max}(\Sigma) - \lambda_{\min}(\Sigma)$ recovers the isotropic spike $C = \lambda_{\min}(\Sigma)\, I$. If Σ is a multiple of the identity, every direction is an EOF1. In all cases Σ alone does not determine whether EOF1 is anomaly or pure background structure.*

Proposition 3.2 (Aligned collapse). *If $\beta > 0$, $v_\star = \pm v_c$, and $C = I + \gamma v_c v_c^\top$, then $\Sigma = I + (\gamma + \beta) v_c v_c^\top$: the aligned anomaly is second-order identical to a larger pure-background spike.*

For $\Sigma = \mathrm{diag}(3, 1)$, both decompositions have direction e_1; quiet held-out C separates them when $\beta > 0$.

Same-record gap erasure. If $\hat{C}$ reuses the analysis window or a contaminated climatology, (2) interpolates to exact same-record at $\varepsilon = 1$. The SNR direction $u_{\mathrm{snr}}(C) \propto C^{-1} v_\star$ remains fixed for $\varepsilon < 1$, but the gap collapses and the exact same-record direction is unidentified (Fig. 1c).

Theorem 3.1 (Contaminated generalised eigenproblem). *Under (2), let $C \succ 0$, $\varepsilon \in [0, 1)$, $\beta > 0$, and $Q := v_\star^\top C^{-1} v_\star > 0$. Define $M_\varepsilon := C_\varepsilon^{-1/2} \Sigma C_\varepsilon^{-1/2}$ and $\lambda_\varepsilon := \lambda_1(M_\varepsilon)$. Then $u_{\mathrm{snr}}(C_\varepsilon) \propto C_\varepsilon^{-1} v_\star \propto C^{-1} v_\star$, and*

$$\lambda_\varepsilon \;=\; 1 + \frac{(1-\varepsilon)\beta Q}{1 + \varepsilon \beta Q}, \qquad \delta_\star(\varepsilon) \;=\; \lambda_\varepsilon - 1 \;=\; \frac{(1-\varepsilon)\beta Q}{1 + \varepsilon \beta Q}, \tag{3}$$

decreases in ε with $\delta_\star(\varepsilon) \to 0$ as $\varepsilon \to 1$. For every threshold $\tau \in (0, \beta Q)$ the gap falls through τ at $\varepsilon(\tau) = (\beta Q - \tau)/(\beta Q(1 + \tau))$. The calibrated direction is invariant for $\varepsilon < 1$ and has a stable limit as $\varepsilon \to 1^-$; at $\varepsilon = 1$, the whitened matrix is I and no direction is identified.

Theorem 3.2 (Same-record non-certifiability). *If $C_\star = \Sigma$, then $C_\star^{-1/2} \Sigma C_\star^{-1/2} = I$ and the population gap is zero. Any deterministic binary rule based only on (Σ, Σ) that must label the observation as anomaly or pure background and never abstains fails on some decomposition in Prop. 3.1. If $\hat{C} = \hat{\Sigma} \succ 0$, then $\hat{M} = I$, so $\hat{\delta} = g = 0$ and every $\tau > 0$ rejects the eigen-gap guard.*

Corollary 3.1 (Aligned collapse needs side information). *Under Prop. 3.2, even infinite samples cannot certify anomaly versus a larger background spike from second-order statistics of Σ alone. Some side information beyond Σ is necessary to separate those two decompositions; quiet held-out C suffices whenever $\beta > 0$, because $\Sigma - C$ then recovers the anomaly direction $v_\star$.*

Finite-sample gap control. Write $M_\star = C_\star^{-1/2} \Sigma C_\star^{-1/2}$ and $\delta_\star = \lambda_1(M_\star) - \max(1, \lambda_2(M_\star))$; Thm. 3.3 yields an empirical guard with explicit thresholds from this population separation.

Assumption 3.1 (Sub-Gaussian residuals). Analysis and calibration samples are independent, centred, i.i.d. sub-Gaussian vectors with covariances Σ and $C_\star$, common covariance-relative norm bound K, and known-zero-mean sample covariances $\hat{\Sigma}$ and $\hat{C}$. Exact same-record input is excluded here and handled by Thm. 3.2; App. A.1 states the full inequalities and estimated-mean Gaussian variant.

Under Assumption 3.1, covariance concentration gives both $\|\hat{\Sigma} - \Sigma\|_{\mathrm{op}} \le \eta_\Sigma(T, t)$ and $\|\hat{C} - C_\star\|_{\mathrm{op}} \le \eta_C(T_c, t)$ simultaneously with probability at least $1 - 2e^{-t}$ (Vershynin, 2026). If $C_\star \succeq \mu I$ and $\eta := \max(\eta_\Sigma, \eta_C) \le \mu/4$, the resolvent representation of the inverse square root yields $\|\hat{M} - M_\star\|_{\mathrm{op}} \le \eta_M := 2(\eta/\mu)(1 + \|\Sigma\|_{\mathrm{op}}/\mu)$ (App. A). For (2) with $C_\star = C_\varepsilon$, write $\delta_\star := \delta_\star(\varepsilon)$ from (3), the population gap that any finite-sample threshold must clear in CAP.

Theorem 3.3 (Finite-sample gap control). *Under Assumption 3.1, (2), $C_\varepsilon \succeq \mu I$, and $\eta \le \mu/4$, if $\delta_\star \ge 4\eta_M$, then with probability at least $1 - 2e^{-t}$,*

$$|\hat{\delta} - \delta_\star| \le 2\eta_M, \qquad \hat{\delta} \ge \delta_\star/2, \qquad \sin \angle(\hat{u}, u_\star) \le 2\eta_M/\delta_\star.$$

Thus a fixed gap guard passes when $\delta_\star \ge \tau + 2\eta_M$ and rejects when $\delta_\star < \tau - 2\eta_M$. Conversely, under uniform conditioning and finite $\bar{\eta}_M$, every fixed $\tau > \bar{\eta}_M$ rejects with the same probability once $\varepsilon < 1$ is sufficiently large. The angle uses whitened coordinates; App. A.7 proves these claims and gives the contamination band and observable Gaussian radii used by CAP below.

Remark 3.1 (Stable direction, vanishing gap). For $\varepsilon < 1$, the direction stays fixed while $\delta_\star(\varepsilon) \to 0$; exact same-record identifies no direction. Thus a stable plot can coexist with a failed gap guard. If a spike remains, CAP abstains; otherwise it returns `no_detectable_residual`. In either case, CAP withholds the direction despite a stable principal axis in the same covariance plot.

4 QUIET BACKGROUNDS ALLOW CERTIFICATION

For a quiet held-out C ($\varepsilon = 0$), $\Sigma - C = \beta v_\star v_\star^\top$ identifies $v_\star$ when $\beta > 0$. This side information distinguishes the decompositions in §3, provided both the spike and whitened gap survive sampling. Even at population level, $\mathrm{top}(\Sigma - C)$ can disagree with residual EOF1 of Σ.

Quiet residual spike. Let $R := \Sigma - C = \beta v_\star v_\star^\top$ and $\hat{R} := \mathrm{PSD}(\hat{\Sigma} - \hat{C})$, where PSD truncates negative eigenvalues. Once β clears covariance error, the Davis–Kahan / Wedin perturbation bound (Davis & Kahan, 1970; Wedin, 1972; Yu et al., 2015; Vershynin, 2026) controls the leading angle used in CAP's certificate for the returned residual eigenvector under quiet calibration.

Theorem 4.1 (Quiet residual spike)**.** *Assume $\varepsilon = 0$, $C_\star = C$, $\Sigma = C + \beta v_\star v_\star^\top$ with $\beta > 0$, and Assumption 3.1. Fix $t \geq 1$ and write $\eta := \max(\eta_\Sigma(T,t), \eta_C(T_c,t))$. If $\beta > 4\eta$, then with probability at least $1 - 2e^{-t}$: $\left|\|\hat{R}\|_{\mathrm{op}} - \beta\right| \leq 2\eta$ and the leading eigenvector $\hat{v}$ of $\hat{R}$ satisfies $\sin\angle(\hat{v}, v_\star) \leq 4\eta/\beta$, by Lem. A.5 on a perturbation of size 2η under quiet held-out C.*

Together, Thms. 3.3 and 4.1 require both $\beta > 4\eta$ and $\delta_\star \geq 4\eta_M$; these are distinct axes, since contamination may erase background-normalised separation while leaving the residual spike visible (App. Prop. A.1), so neither guard substitutes for the other without additional spectral bounds.

EOF1 versus certified direction under quiet C. Residual EOF1 u_{var} maximises variance; the SNR direction $u_{\mathrm{snr}}(C) \propto C^{-1} v_\star$ ranks signal against background; CAP returns $v_\star$. We therefore distinguish variance-to-CAP disagreement from variance-to-SNR disagreement:

$$\Delta_{\mathrm{CAP}}(\Sigma; C) \;:=\; 1 - \left(u_{\mathrm{var}}^\top v_\star\right)^2, \qquad \Delta_{\mathrm{SNR}}(\Sigma; C) \;:=\; 1 - \left(u_{\mathrm{var}}^\top u_{\mathrm{snr}}(C)\right)^2. \tag{4}$$

Proposition 4.1 (EOF1–CAP disagreement under quiet C)**.** *In the spiked-background family of §2 with quiet C, there is an open set of (α, β, γ) on which $\Delta_{\mathrm{CAP}}(\Sigma; C) > 0$: u_{var} can nearly align with the background axis v_c while $\Sigma - C$ still recovers $v_\star$ (App. A.9). A concrete witness is $(\alpha, \beta, \gamma) = (0.8, 1, 8)$ in $n = 2$. There residual EOF1 of Σ tracks v_c, the EOF1–CAP gap is $\Delta_{\mathrm{CAP}} \approx 0.31$, the distinct EOF1–SNR gap is $\Delta_{\mathrm{SNR}} \approx 0.96$, and $\mathrm{top}(\Sigma - C) = v_\star$ keeps a nontrivial innovation component. Under high alignment, the leading variance direction of Σ is mostly background, yet the residual spike points elsewhere. Thus CAP's theorem-backed quiet path targets $\mathrm{top}(\Sigma - C)$ rather than residual EOF1 of Σ under the stated side information and rank-one model.*

5 CONTRAST AUDIT PROCEDURE

CAP implements Table 1 as an abstain-by-default rule. With a quiet held-out C, it may plot u_{var} but certifies only $\mathrm{top}(\hat{R})$; optional physical interpretations lie outside the guarantee.

Certification target. CAP audits evidence for an anomaly claim, not its semantics. Theorem mode certifies only $\mathrm{top}(\hat{R})$ as a separated residual direction under the declared reference, provenance, thresholds, and model; declared mode reports two frozen positive margins. Neither establishes causality or importance; every return inherits the reference's side information and inferential scope.

Statistical interpretation. Fixed cutoffs can establish that both guards pass; only theorem mode yields the certificate in Cor. 5.1. An optional empirical calibration gives fixed-threshold mode a finite-sample null-return bound under exchangeability, but no direction-coverage guarantee. Otherwise CAP reports no detectable residual or abstains; the associated margins state how far the observation lies from either boundary and identify which prerequisite failed.

Exchangeable-control calibration. For m exchangeable quiet controls, upper-tail rank p-values p_G, p_S for the gap and spike obey $\Pr\{\max(p_G, p_S) \leq \alpha\} \leq \alpha$ without guard independence (App. Prop. B.1). The bound controls one null return but does not validate reference quietness, supply anomaly semantics, or direction coverage under drift or calibration shift.

Why two guards. The spike guard controls absolute recovery of CAP's contrast direction; the whitened gap controls background-relative separation and stability of a distinct generalised eigendirection. Neither implies the other without spectral bounds on C, so their conjunction is minimal for CAP's two-part report (App. Prop. A.1), not a claim of optimal event detection. Spike failure means no residual at the declared resolution; gap failure blocks certification but does not prove that $\mathrm{top}(\hat{R})$ is unstable or that the observed residual has vanished at the declared resolution.

Decision rule. Missing side information, reuse, or unstable whitening leads to abstention; a weak spike reports no detectable residual; a visible spike with unresolved gap also abstains; only two positive margins return $\mathrm{top}(\hat{R})$. Theorem mode also requires documented independence and model-based radii, setting $(\tau_{\mathrm{spike}}, \tau_{\mathrm{gap}}) = m_0(\bar{\eta}, \bar{\eta}_M)$ for $m_0 > 6$. Because the spike guard has covariance units and the gap is dimensionless, rescaling data changes only the spike threshold.

Table 2: **CAP's four outcomes.** Only theorem mode provides direction coverage; Prop. B.1 controls null returns when test nulls and controls are exchangeable, without guard independence.

Status	Trigger	Permitted statement
`abstain`	missing/reused/unstable reference, or failed relative-separation guard	no residual direction is certified; contrast stability is not negated
`no_detectable_residual`	residual spike below its threshold	no residual detected at the declared resolution
`guard_pass_direction`	both fixed-threshold margins are positive	residual direction; optional exchangeable-null return control, no anomaly label
`certified_direction`	theorem inputs valid and both theorem margins positive	residual direction plus finite-sample radii and failure probability

Algorithm 1 Contrast Audit Procedure (CAP) for background-adjusted leading directions

Require: $\hat{\Sigma}$; optional $\hat{C}$; mode; provenance; cutoffs or theorem inputs with independence evidence.

$u_{\mathrm{var}} \leftarrow \mathrm{top}(\hat{\Sigma})$ ▷ leading variance mode; not certified
if $m =$ `none` **then return** `abstain`, u_{var}
if provenance matches or $\hat{C} = \hat{\Sigma}$ **then return** `abstain` with `same_record`
if $\hat{C}$ is absent or fails conditioning guards **then return** `abstain`
Resolve declared or theorem thresholds; invalid inputs, independence evidence, or radii **return** `abstain`
$\hat{R} \leftarrow \mathrm{PSD}(\hat{\Sigma} - \hat{C})$; $\hat{M} \leftarrow \hat{C}^{-1/2}\hat{\Sigma}\hat{C}^{-1/2}$; $\hat{\delta} \leftarrow \lambda_1(\hat{M}) - \max(1, \lambda_2(\hat{M}))$
if $\|\hat{R}\|_{\mathrm{op}} < \tau_{\mathrm{spike}}$ **then return** `no_detectable_residual`, no direction
if $\hat{\delta} < \tau_{\mathrm{gap}}$ **then return** `abstain` with `gap_fail`
if declared-threshold mode **then return** `guard_pass_direction`, $\mathrm{top}(\hat{R})$, margins
else
    return `certified_direction`, $\mathrm{top}(\hat{R})$, margins / angle radii
end if ▷ return the direction with both margins

Reported quantities. Every positive return includes provenance, normalisation, sample sizes, thresholds, margins, and direction. Theorem mode adds model inputs, population lower bounds, angle radii, and failure probability; missing assumptions lead to abstention. App. B specifies these quantities, the optional branches, and their identification limits in each decision regime.

Corollary 5.1 (CAP guarantees)**.** (i) *CAP abstains deterministically without side information or on exact same-record input.* (ii) *Under Assumption 3.1, quiet $C \succeq \mu I$, $\eta \leq \mu/4$, $\beta > 4\eta$, and $\delta_\star \geq 4\eta_M$, oracle-interval thresholds accept $\mathrm{top}(\hat{R})$ with probability at least $1 - 2e^{-t}$.* (iii) *Uniform conditioning and any fixed $\tau_{\mathrm{gap}} > \bar{\eta}_M$ prevent a positive return under sufficiently large contamination; $\tau_{\mathrm{spike}} > 2\eta$ does so under the quiet null.* (iv) *Theorem-mode acceptance with $m_0 > 6$ implies both population separation conditions and carries observable angle bounds (App. A.10). Fixed cutoffs lack direction coverage; Prop. B.1's null-return bound requires target exchangeability.*

6 EMPIRICAL VALIDATION

We evaluate the theory from controlled mechanisms to independent physical targets: simulations isolate contamination and finite-sample effects; INSTANCE measures station-held-out event separation; LOCATA supplies optical direction truth; and NOAA and Etna probe low-signal regimes. All experiments use Alg. 1's $\hat{\delta} = \lambda_1 - \max(1, \lambda_2)$. Only the Gaussian study invokes theorem mode; real-data results are development-frozen declared-mode guard passes, without direction coverage or physical-anomaly guarantees beyond their calibration regimes or target populations.

Contamination and null specificity under independent sampling. For $T = T_c = 400$, we freeze $(\tau_{\text{gap}}, \tau_{\text{spike}}) = (0.290, 0.966)$ at the 0.95 quantiles of a disjoint 2,000-null pilot. Across 200 fresh independent $(\hat{\Sigma}, \hat{C}_\varepsilon)$ pairs per ε, guard-pass rates are $1.00, 0.945, 0.23, 0.02, 0.02$ at $\varepsilon = 0, 0.30, 0.60, 0.85, 0.95$. Separately, five fresh 200-null batches give $0.01, 0, 0.02, 0, 0$ (pooled 0.006; Fig. 2). Thus guard passing collapses even without covariance reuse while null specificity remains high; these fixed thresholds do not invoke the theorem-mode guarantee.

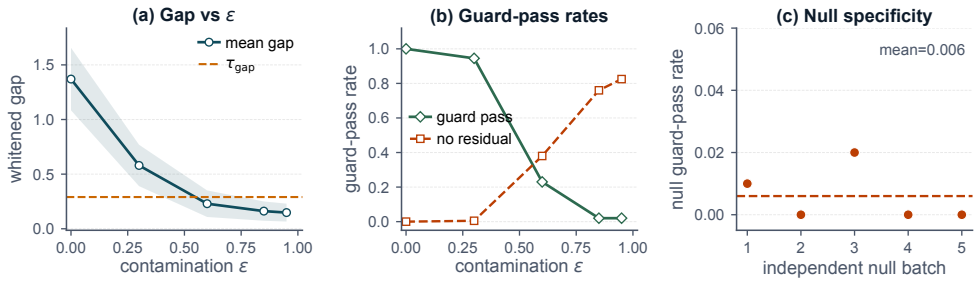

Figure 2: **Contamination erodes separation under independent sampling** (200/cell): **(a)** independent gap and frozen guard; **(b)** pass rate falls to 0.02; **(c)** five fresh null batches give pooled rate 0.006. Pilot, contamination grid, and null batches are mutually independent; the pilot determines the thresholds applied unchanged to every evaluation cell and replicate without refitting.

Sensitivity to nuisance conditions and model rank. Seven cells vary T/n, dimension, condition number, Gaussian versus Rademacher or t_5 draws, and a near-tied rank-two alternative outside the theorem. Each cell uses a 600-null pilot, then 250 fresh null and 250 alternative replicates. Matched fresh-null pass rates remain 0.004–0.020, but power falls to 0.076 at $T/n = 3$ and direction alignment to 0.820 under the near tie (Fig. 3). Transporting the baseline cutoff to $T/n = 3$ gives a 0.928 null-pass rate, so fixed thresholds require nuisance matching before reuse across sampling regimes.

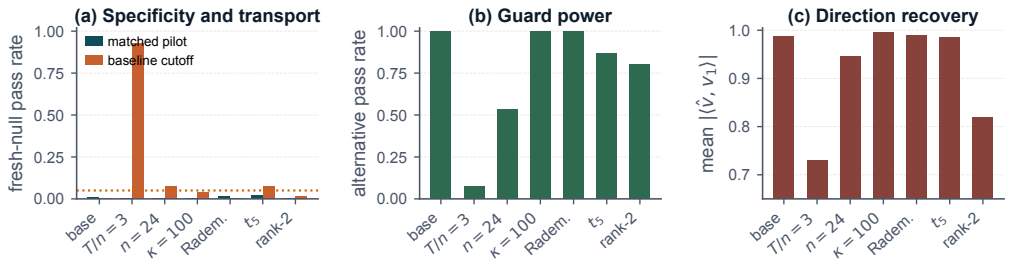

Figure 3: **Sensitivity to sample size, conditioning, tails, and model rank** (600 independent null-pilot, 250 fresh null, and 250 fresh alternative replicates per cell). **(a)** Matched cutoffs retain low fresh-null rates, whereas the baseline cutoff fails under low-T/n transport. **(b)** Low T/n and a near-tied rank-two residual reduce guard power. **(c)** Direction recovery also degrades; the rank-two case lies outside the rank-one theorem and cannot support recovery without additional model structure.

We run theorem mode with exact Gaussian Wishart covariances, $n = 2$, $C = I$, $\beta = 2$, $t = 3$, and no fitted constants. Across 200 fresh replicates per cell for $T = T_c \in \{0.25, 0.5, 1, 2, 4\}$ million and $\varepsilon \in \{0, .1, .2, .4, .6, .8, .95\}$, the empirical frontier expands with sample size beyond the conservative sufficient frontier (Fig. 4). All five null grids have zero positive returns, and no

acceptance violates its residual-angle radius, consistent with Cor. 5.1. The real-data analyses use fixed thresholds and carry no nominal finite-sample or direction-coverage guarantee.

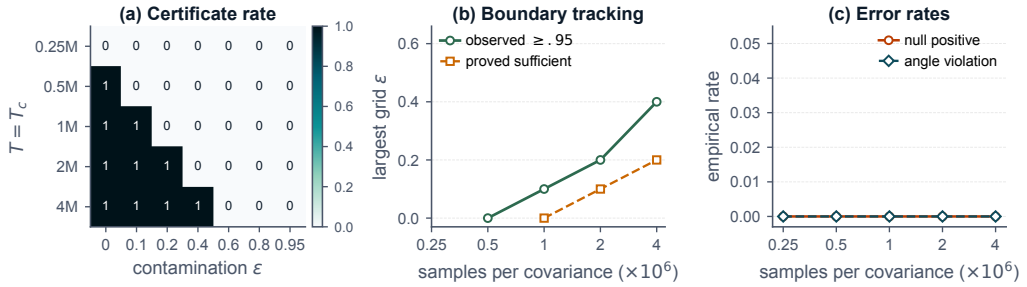

Figure 4: **Finite-sample certification under the Gaussian rank-one model** (200 replicates per cell). **(a)** Certification across contamination and sample size. **(b)** Largest tested ε with certification rate at least 0.95, compared with the proved sufficient boundary. **(c)** Null-positive and conditional angle-bound violation rates are zero in every tested cell across the full contamination and sample-size grid with unchanged theorem constants and the same nominal failure probability throughout.

Station-held-out real event/noise benchmark. Our fixed official INSTANCE v3 sample has 10,000 three-component earthquakes, 1,000 independently labelled noise traces, 120 seconds at 100 Hz, and catalog arrivals/magnitudes (Michelini et al., 2021). A prespecified outcome-blind station/channel split yields 196 development and 145 test keys. Disjoint 30-second noise windows supply $\hat{C}$ and negative controls; positives begin five seconds before catalog P arrival and pair with independent same-key noise. We filter 1–20 Hz, standardise by the reference window, retain three channels, and select one deterministic test trace per earthquake, leaving 299 source-level positives.

We freeze $(\tau_{\text{gap}}, \tau_{\text{spike}}) = (2.733, 5.030)$ at the marginal 0.95 quantiles of development noise. Without refitting, joint guards pass $13/264 = .049$ held-out noise and $240/299 = .803$ earthquakes (station-cluster 95% CIs $[.023, .080]$, $[.717, .878]$); development noise passes $16/381 = .042$ (Fig. 5a–b). Prop. B.1's rank rule at $\alpha = .05$ returns $13/264$ noise and $238/299$ events; its finite-sample level applies only if new nulls exchange with development controls, while station holdout tests transport. At the same $16/381$ development return count, Roy's largest-root and Gaussian log-determinant gates (Roy, 1953; Box, 1949) are strong event detectors: they return $254/299$ and $248/299$ events with $13/264$ and $7/264$ noise, versus CAP's $240/299$ and $13/264$; bootstrap stability returns $41/299$ and $11/264$. Yet at $1\times$ known-direction injection, accurate-return rates are .939 (bootstrap), .470 (CAP), .337 (Roy), and .027 (log-det); no gate dominates across labelled shifts, directional recovery, or physical transfer across these held-out regimes (Table 3, App. Fig. 7).

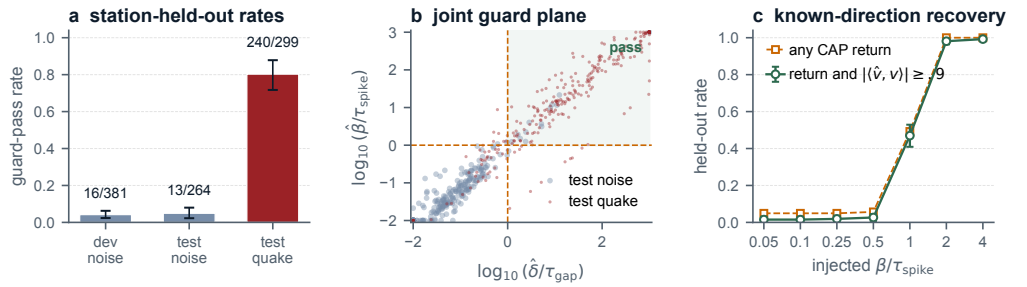

Figure 5: **Public INSTANCE validation.** **(a)** Frozen-guard rates with station-clustered 95% intervals. **(b)** Both normalised guards; the upper-right quadrant passes. **(c)** Semi-synthetic known-direction injection into held-out real backgrounds; accurate means CAP returns with cosine $\geq .9$.

Low-signal regimes across acoustics and infrasound. For NOAA/DCLDE right-whale acoustics, guards frozen on 34 development quiet windows pass $1/37$ held-out quiet files and only $4/37$ annotated calls; 32 calls return no residual and one fails the gap (DCLDE 2013, 2023). Weak labels therefore need not form an estimable rank-one residual under this representation. For Etna infrasound, development and held-out-array controls pass $2/30$ and $1/30$ blocks (De Angelis et al., 2021); high-activity windows pass 0.80 and 0.90 with repeatability 0.92 and 0.60, whereas the other

six pass $15/120$ and none clears the majority rule (Fig. 6). LOCATA frozen guards return $11/13$ sources versus $1/13$ inactive controls (Evers et al., 2020); returned azimuths have 2.1° median error against optical tracking, and every active recording exceeds its paired control on both statistics under the unchanged frozen protocol (App. Fig. 8). Here the gap guard is binding (joint equals gap-only): LOCATA evaluates returned directions against physical ground truth, whereas INSTANCE supplies complementary evidence that both guards affect decisions. Bootstrap returns all 13 sources but also $7/13$ inactive controls; Roy returns $7/13$ and $1/13$, while the log-det gate returns none, versus CAP's $11/13$ and $1/13$ under the identical frozen protocol and evaluation units (Table 3).

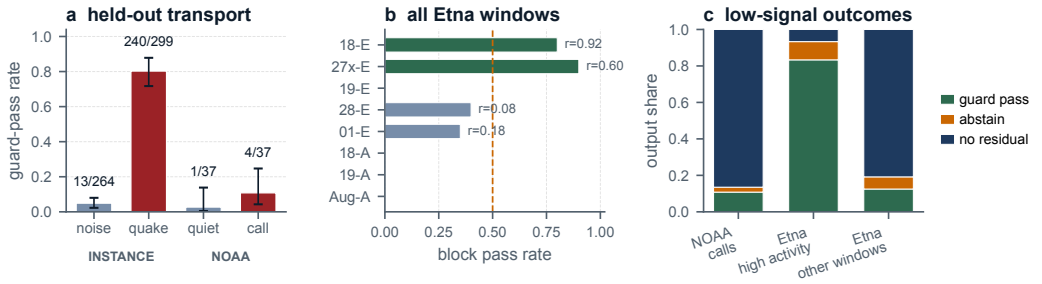

Figure 6: **Signal resolution across seismic, acoustic, and infrasound data.** **(a)** Held-out rates for INSTANCE and NOAA; **(b)** Etna pass rates and direction repeatability r under a fixed 0.5 rule; **(c)** the low-SNR domains usually return no direction across both recording systems.

Table 3: **Matched external gates.** INSTANCE cutoffs match CAP's $16/381$ development-null returns; LOCATA uses development maxima. All gates return the same contrast direction after selection, so accurate-return columns compare selection rules under matched development cutoffs.

Gate	INSTANCE noise / event	1× accurate	2× accurate	LOCATA inactive / active
Bootstrap stability	$11/264$ / $41/299$	.939	.962	$7/13$ / $13/13$
Roy largest root	$13/264$ / $254/299$	.337	.981	$1/13$ / $7/13$
Gaussian log-det	$7/264$ / $248/299$	.027	.042	$0/13$ / $0/13$
Whitened Frobenius	$12/264$ / $250/299$	.140	.951	$0/13$ / $6/13$
CAP	$13/264$ / $240/299$	.470	.981	$1/13$ / $11/13$

Cross-domain interpretation. The failures match the two theoretical margins: NOAA is spike-limited, LOCATA is gap-limited, and INSTANCE uses both guards. Matched baselines show that event-detection power need not transfer to directional recovery under held-out conditions.

Dependence and calibration transfer. Replication uses one INSTANCE trace per earthquake, dates for NOAA, and a second Etna array. A half-record cutoff passes $22/29$ and $6/8$ disjoint 2-second event pairs; all $100/100$ circular-lag surrogates pass. Thus cutoffs do not transfer across duration, and the representation detects covariance shifts, not temporal coherence (App. Table 5).

7 Conclusion and Future Work

We establish a sharp boundary: analysis covariance alone cannot identify an anomaly; same-record contamination erases background-normalised separation; and quiet held-out C identifies $\mathrm{top}(\Sigma - C)$, not EOF1 of Σ. The residual-spike and whitened-gap guards answer different questions; CAP requires both without claiming detector optimality or treating gap failure as proof of an unstable contrast. Exchangeable controls bound per-record null returns without a parametric covariance model, but not direction coverage or reference drift. Empirically, matched external gates expose task-specific tradeoffs: the Gaussian log-det and Roy scores strengthen event detection, bootstrap favors moderate injected directions, and CAP transfers most selectively to physical LOCATA directions. Theorem mode supplies conditional finite-sample coverage; the real studies test only the reproducible declared-mode decision rule. Limits include semi-synthetic directions, unmatched dates, dependence, small physical panels, and one room/array; future work should test preregistered multisite panels, higher-rank subspaces, and drifting backgrounds. CAP makes abstention reproducible by tracing every return to declared evidence, thresholds, and reference provenance. It converts an automatic residual eigenvector into a report exposing both margins and negative controls, distinguishing no detectable residual, an unstable direction, and a certified direction.

AI USE STATEMENT

In this work, we used generative AI tools to assist with grammar and readability, coding, and feedback on experimental design and the interpretation of results. The authors independently reviewed all AI-assisted contributions and take full responsibility for the methods, analyses, interpretations, and final manuscript.

ETHICS STATEMENT

This work studies statistical certification of residual modes in public seismic, acoustic, volcanic-infrasound, ocean, and satellite remote-sensing data. It does not involve human subjects and is not an operational warning system.

REPRODUCIBILITY STATEMENT

We state the assumptions underlying all theoretical claims and provide complete proofs in the appendix. The paper and appendix document the datasets, preprocessing, prespecified splits, frozen thresholds, evaluation units, comparators, uncertainty estimates, and additional replications used in our experiments. We release our code, prespecified splits, results, and reproduction instructions in our anonymous repository.

REFERENCES

Abubakar Abid, Martin J. Zhang, Vivek K. Bagaria, and James Zou. Exploring patterns enriched in a dataset with contrastive principal component analysis. *Nature Communications*, 9(1):2134, 2018. doi: 10.1038/s41467-018-04608-8.

Myles R. Allen and Peter A. Stott. Estimating signal amplitudes in optimal fingerprinting, part I: theory. *Climate Dynamics*, 21(5–6):477–491, 2003. doi: 10.1007/s00382-003-0313-9.

Myles R. Allen and Simon F. B. Tett. Checking for model consistency in optimal fingerprinting. *Climate Dynamics*, 15(6):419–434, 1999. doi: 10.1007/s003820050291.

Jinho Baik and Jack W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408, 2006. doi: 10.1016/j.jmva.2005.08.003.

Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005. doi: 10.1214/009117905000000233.

Steffen Beirle, Christian Borger, Adrian Jost, and Thomas Wagner. Improved catalog of NO_x point source emissions (version 2). *Earth System Science Data*, 15(7):3051–3073, 2023. doi: 10.5194/essd-15-3051-2023.

Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011. doi: 10.1016/j.aim.2011.02.007.

G. D. Bensen, M. H. Ritzwoller, M. P. Barmin, A. L. Levshin, F. Lin, M. P. Moschetti, N. M. Shapiro, and Y. Yang. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. *Geophysical Journal International*, 169(3):1239–1260, 2007. doi: 10.1111/j.1365-246X.2007.03374.x.

George E. P. Box. A general distribution theory for a class of likelihood criteria. *Biometrika*, 36 (3–4):317–346, 1949. doi: 10.1093/biomet/36.3-4.317.

Christopher S. Bretherton, Catherine A. Smith, and John M. Wallace. An intercomparison of methods for finding coupled patterns in climate data. *Journal of Climate*, 5(6):541–560, 1992. doi: 10.1175/1520-0442(1992)005⟨0541:AIOMFF⟩2.0.CO;2.

Tony Cai, Weidong Liu, and Yin Xia. Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association*, 108(501): 265–277, 2013. doi: 10.1080/01621459.2012.758041.

Michel Campillo and Anne Paul. Long-range correlations in the diffuse seismic coda. *Science*, 299 (5606):547–549, 2003. doi: 10.1126/science.1078551.

Yuxin Chen, Yuejie Chi, Jianqing Fan, and Cong Ma. Spectral methods for data science: A statistical perspective. *Foundations and Trends in Machine Learning*, 14(5):566–806, 2021. doi: 10.1561/2200000079.

C. K. Chow. On optimum recognition error and reject tradeoff. *IEEE Transactions on Information Theory*, 16(1):41–46, 1970. doi: 10.1109/TIT.1970.1054406.

Copernicus Sentinel-5P (processed by ESA). TROPOMI level 2 nitrogen dioxide total column products, version 01. European Space Agency, `https://doi.org/10.5270/S5P-s4ljg54`, 2018. Accessed 2026-09-25.

Chandler Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis*, 7(1):1–46, 1970. doi: 10.1137/0707001.

DCLDE 2013. Baleen whale annotations from the detection classification localization and density estimate conference in 2013. NOAA National Centers for Environmental Information, `https://doi.org/10.25921/zaea-1s39`, 2023. Accessed 2026-09-25.

Silvio De Angelis, Luciano Zuccarello, Sebastiano Rapisarda, and Vittorio Minio. Introduction to a community dataset from an infrasound array experiment at Mt. Etna, Italy. *Scientific Data*, 8(1): 247, 2021. doi: 10.1038/s41597-021-01030-6.

Christine Evers, Heinrich W. Löllmann, Heinrich Mellmann, Alexander Schmidt, Hendrik Barfuss, Patrick A. Naylor, and Walter Kellermann. The LOCATA challenge: Acoustic source localization and tracking. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:1620–1643, 2020. doi: 10.1109/TASLP.2020.2990485.

Matan Gavish and David L. Donoho. The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*, 60(8):5040–5053, 2014. doi: 10.1109/TIT.2014.2323359.

Yonatan Geifman and Ran El-Yaniv. Selective classification for deep neural networks. In *Advances in Neural Information Processing Systems*, volume 30, pp. 4878–4887, 2017.

A. Hannachi, I. T. Jolliffe, and D. B. Stephenson. Empirical orthogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology*, 27(9):1119–1152, 2007. doi: 10.1002/joc.1499.

Klaus Hasselmann. Optimal fingerprints for the detection of time-dependent climate change. *Journal of Climate*, 6(10):1957–1971, 1993. doi: 10.1175/1520-0442(1993)006⟨1957:OFFTDO⟩2.0.CO;2.

Sam Hawke, YueEn Ma, and Didong Li. Contrastive dimension reduction: When and how? In *Advances in Neural Information Processing Systems*, volume 37, pp. 74034–74057, 2024. doi: 10.52202/079017-2355.

Gabriele C. Hegerl, Hans von Storch, Klaus Hasselmann, Benjamin D. Santer, Ulrich Cubasch, and Philip D. Jones. Detecting greenhouse-gas-induced climate change with an optimal fingerprint method. *Journal of Climate*, 9(10):2281–2306, 1996. doi: 10.1175/1520-0442(1996)009⟨2281:DGGICC⟩2.0.CO;2.

Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 1933. doi: 10.1037/h0071325.

Iain M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, 29(2):295–327, 2001. doi: 10.1214/aos/1009210544.

I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, 2 edition, 2002. ISBN 9780387954424. doi: 10.1007/b98835.

Vladimir Koltchinskii and Karim Lounici. Normal approximation and concentration of spectral projectors of sample covariance. *The Annals of Statistics*, 45(1):121–157, 2017. doi: 10.1214/16-AOS1437.

Hans R. Künsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3):1217–1241, 1989. doi: 10.1214/aos/1176347265.

Edward N. Lorenz. Empirical orthogonal functions and statistical weather prediction. Scientific Report No. 1, Statistical Forecasting Project, Air Force Research Laboratories, Office of Aerospace Research, U.S. Air Force, Bedford, MA, 1956. URL https://wind.mit.edu/~emanuel/Lorenz/EdLorenz/Empirical_Orthogonal_Functions_1956.pdf.

Alberto Michelini, Spina Cianetti, Sonja Gaviano, Carlo Giunchi, Dario Jozinović, and Valentino Lauciani. INSTANCE—the Italian seismic dataset for machine learning. *Earth System Science Data*, 13(12):5509–5544, 2021. doi: 10.5194/essd-13-5509-2021.

S. Mostafa Mousavi, William L. Ellsworth, Weiqiang Zhu, Lindsay Y. Chuang, and Gregory C. Beroza. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11(1):3952, 2020. doi: 10.1038/s41467-020-17591-w.

Boaz Nadler. Finite sample approximation results for principal component analysis: A matrix perturbation approach. *The Annals of Statistics*, 36(6):2791–2817, 2008. doi: 10.1214/08-AOS618.

Alexey Naumov, Vladimir Spokoiny, and Vladimir Ulyanov. Bootstrap confidence sets for spectral projectors of sample covariance. *Probability Theory and Related Fields*, 174(3–4):1091–1132, 2019. doi: 10.1007/s00440-018-0877-2.

Gerald R. North, Thomas L. Bell, Robert F. Cahalan, and Fanthune J. Moeng. Sampling errors in the estimation of empirical orthogonal functions. *Monthly Weather Review*, 110(7):699–706, 1982. doi: 10.1175/1520-0493(1982)110⟨0699:SEITEO⟩2.0.CO;2.

Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, 17(4):1617–1642, 2007. URL https://www3.stat.sinica.edu.tw/statistica/j17n4/j17n418/j17n418.html.

Rudolph W. Preisendorfer. *Principal Component Analysis in Meteorology and Oceanography*, volume 17 of *Developments in Atmospheric Science*. Elsevier, Amsterdam, 1988. ISBN 9780444430144. Posthumously compiled and edited by Curtis D. Mobley.

Aurélien Ribes, Serge Planton, and Laurent Terray. Application of regularised optimal fingerprinting to attribution. Part I: method, properties and idealised analysis. *Climate Dynamics*, 41(11–12):2817–2836, 2013. doi: 10.1007/s00382-013-1735-7.

S. N. Roy. On a heuristic method of test construction and its use in multivariate analysis. *The Annals of Mathematical Statistics*, 24(2):220–238, 1953. doi: 10.1214/aoms/1177729029.

Nikolai M. Shapiro and Michel Campillo. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, 31(7):L07614, 2004. doi: 10.1029/2004GL019491.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*, volume 58 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2 edition, 2026. ISBN 9781009490641. doi: 10.1017/9781009490672.

Hans von Storch and Francis W. Zwiers. *Statistical Analysis in Climate Research*. Cambridge University Press, 1999. ISBN 9780521450713. doi: 10.1017/CBO9780511612336.

Per-Åke Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12(1):99–111, 1972. doi: 10.1007/BF01932678.

Yi Yu, Tengyao Wang, and Richard J. Samworth. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 2015. doi: 10.1093/biomet/asv008.

Yihan Zhang and Marco Mondelli. Matrix denoising with doubly heteroscedastic noise: Fundamental limits and optimal spectral methods. In *Advances in Neural Information Processing Systems*, volume 37, pp. 93060–93117, 2024. doi: 10.52202/079017-2955.

A Detailed Proofs of Theoretical Claims

We prove the paper's identification and certification claims in full. We move from observational equivalence and same-record gap erasure to quiet residual recovery, EOF1–anomaly disagreement, and CAP's decision behaviour. Notation follows the main text unless stated otherwise. Throughout, $\mathrm{top}(A)$ denotes a unit leading eigenvector of a symmetric matrix A, unique up to sign when the leading eigenvalue is simple. Also, $\mathrm{PSD}(A)$ denotes the positive-semidefinite part of A, obtained by setting negative eigenvalues to zero in an eigen-decomposition. We allow universal constants C, C' to change from line to line and absorb them into $\asymp$ where convenient.

Proof map. Shared lemmas (App. A.1) support Thms. 3.3–4.1; Apps. A.2–A.3 are algebraic.

A.1 Shared lemmas (concentration and whitening)

These lemmas supply the finite-sample bounds used in Thms. 3.3–4.1 and Cor. 5.1; the Gaussian specialisation expresses every constant in observable quantities.

Lemma A.1 (Covariance concentration)**.** *Under Assumption 3.1, there exists an absolute constant $c_1 > 0$ such that for every $t \geq 1$, with probability at least $1 - 2e^{-t}$, the following two bounds hold simultaneously:*

$$\|\hat{\Sigma} - \Sigma\|_{\mathrm{op}} \;\leq\; c_1 K^2 \|\Sigma\|_{\mathrm{op}} \Big(\sqrt{\tfrac{n+t}{T}} + \tfrac{n+t}{T} \Big) \;=:\; \eta_\Sigma(T,t), \tag{5}$$

and $\|\hat{C} - C_\star\|_{\mathrm{op}} \leq \eta_C(T_c, t)$ with the same form in $(T_c, C_\star)$ (Vershynin, 2026).

Proof. Apply the standard operator-norm covariance bound for covariance-relative sub-Gaussian vectors to each sample, assigning failure probability at most e^{-t} to each event (Vershynin, 2026). A union bound gives joint failure probability at most $2e^{-t}$. □

Lemma A.2 (Whitened perturbation)**.** *Assume $C_\star \succeq \mu I$ for some $\mu > 0$, and let $\eta := \max(\|\hat{\Sigma} - \Sigma\|_{\mathrm{op}}, \|\hat{C} - C_\star\|_{\mathrm{op}})$. If $\eta \leq \mu/4$, then $\hat{C} \succeq \frac{3}{4}\mu I$, the inverse square roots exist, and*

$$\|\hat{M} - M_\star\|_{\mathrm{op}} \;\leq\; 2\,\frac{\eta}{\mu}\Big(1 + \frac{\|\Sigma\|_{\mathrm{op}}}{\mu}\Big) \;=:\; \eta_M, \tag{6}$$

where $\hat{M} := \hat{C}^{-1/2}\hat{\Sigma}\hat{C}^{-1/2}$, $M_\star := C_\star^{-1/2}\Sigma C_\star^{-1/2}$, and one may take the explicit constant $c_2 = 2$.

Proof. Write $E_C := \hat{C} - C_\star$ and $E_\Sigma := \hat{\Sigma} - \Sigma$, so $\|E_C\|_{\mathrm{op}}, \|E_\Sigma\|_{\mathrm{op}} \leq \eta$. By Weyl and $C_\star \succeq \mu I$, we obtain $\lambda_{\min}(\hat{C}) \geq \mu - \eta \geq \frac{3}{4}\mu$, hence $\|\hat{C}^{-1}\|_{\mathrm{op}} \leq 4/(3\mu)$ and $\|\hat{C}^{-1/2}\|_{\mathrm{op}} \leq 2/\sqrt{3\mu}$. Likewise $\|C_\star^{-1/2}\|_{\mathrm{op}} \leq \mu^{-1/2}$. For every positive-definite matrix A,

$$A^{-1/2} = \frac{1}{\pi}\int_0^\infty t^{-1/2}(A + tI)^{-1}\,dt.$$

The resolvent identity therefore yields

$$\hat{C}^{-1/2} - C_\star^{-1/2} = \frac{1}{\pi}\int_0^\infty t^{-1/2}(\hat{C} + tI)^{-1}(C_\star - \hat{C})(C_\star + tI)^{-1}\,dt.$$

Writing $a := 3\mu/4$ and using both $\hat{C}, C_\star \succeq aI$,

$$\|\hat{C}^{-1/2} - C_\star^{-1/2}\|_{\mathrm{op}} \leq \frac{\eta}{\pi}\int_0^\infty \frac{t^{-1/2}}{(a+t)^2}\,dt = \frac{\eta}{2a^{3/2}} \leq C'\eta\mu^{-3/2}.$$

This argument does not assume that $C_\star$ and E_C commute. Expand

$$\hat{M} - M_\star = \hat{C}^{-1/2} E_\Sigma \hat{C}^{-1/2} + \big(\hat{C}^{-1/2} - C_\star^{-1/2}\big)\Sigma\hat{C}^{-1/2} + C_\star^{-1/2}\Sigma\big(\hat{C}^{-1/2} - C_\star^{-1/2}\big).$$

The first term is at most $4\eta/(3\mu)$. The last two are jointly at most $(8/9 + 4/(3\sqrt{3}))\eta\|\Sigma\|_{\mathrm{op}}/\mu^2 < 2\eta\|\Sigma\|_{\mathrm{op}}/\mu^2$. Thus (6) holds with $c_2 = 2$. □

Lemma A.3 (Observable theorem-mode radii)**.** *Let*

$$a_\Sigma := c_1 K^2\left(\sqrt{\frac{n+t}{T}} + \frac{n+t}{T}\right), \qquad a_C := c_1 K^2\left(\sqrt{\frac{n+t}{T_c}} + \frac{n+t}{T_c}\right),$$

and assume $\max(a_\Sigma, a_C) < 1$*. Define*

$$\bar{s}_\Sigma := \frac{\|\hat{\Sigma}\|_{\mathrm{op}}}{1 - a_\Sigma}, \quad \bar{s}_C := \frac{\|\hat{C}\|_{\mathrm{op}}}{1 - a_C}, \quad \bar{\eta} := \max(a_\Sigma \bar{s}_\Sigma, a_C \bar{s}_C).$$

On the event of Lem. A.1, the two covariance errors are at most $\bar{\eta}$ *and* $\|\Sigma\|_{\mathrm{op}} \le \bar{s}_\Sigma$*. If* $\bar{\eta} \le \mu/4$*, then*

$$\|\hat{M} - M_\star\|_{\mathrm{op}} \le \bar{\eta}_M := 2\,\frac{\bar{\eta}}{\mu}\left(1 + \frac{\bar{s}_\Sigma}{\mu}\right).$$

Proof. On the concentration event, $\|\hat{\Sigma}\|_{\mathrm{op}} \ge (1 - a_\Sigma)\|\Sigma\|_{\mathrm{op}}$ by the reverse triangle inequality, and likewise for $C_\star$. Thus the population norms and covariance errors are bounded by the displayed observable quantities. Applying Lem. A.2 with these upper bounds yields $\bar{\eta}_M$. □

Lemma A.4 (Explicit Gaussian covariance radii)**.** *Suppose the analysis and calibration samples are independent Gaussian samples. For* $N \in \{T, T_c\}$*, let* $d_N = N$ *for known mean, or* $d_N = N - 1$ *when the sample mean is estimated and the covariance is divided by* $N - 1$*. Set*

$$q_N := \frac{\sqrt{n} + \sqrt{2(t + \log 4)}}{\sqrt{d_N}}, \qquad a_N := 2q_N + q_N^2.$$

If $q_T, q_{T_c} < 1$*, then with probability at least* $1 - e^{-t}$*,* $\|\hat{\Sigma} - \Sigma\|_{\mathrm{op}} \le a_T\|\Sigma\|_{\mathrm{op}}$ *and* $\|\hat{C} - C_\star\|_{\mathrm{op}} \le a_{T_c}\|C_\star\|_{\mathrm{op}}$ *simultaneously. If additionally* $a_T, a_{T_c} < 1$ *(equivalently* $q_T, q_{T_c} < \sqrt{2} - 1$*), then Lem. A.3 applies with* $(a_\Sigma, a_C) = (a_T, a_{T_c})$ *and* $c_2 = 2$*.*

Proof. After whitening, each known-mean covariance is $d_N^{-1} ZZ^\top$ for an $n \times d_N$ standard Gaussian matrix; the unbiased centred covariance has the same Wishart law with $d_N = N - 1$. The Gaussian singular-value inequalities place every singular value of $Z/\sqrt{d_N}$ in $[1 - q_N, 1 + q_N]$ except with probability $2e^{-(t+\log 4)}$ per matrix (Vershynin, 2026). A union bound over two matrices costs at most e^{-t}, and squaring gives relative covariance error at most $\max\{1-(1-q_N)^2, (1+q_N)^2-1\} = 2q_N + q_N^2$. The observable conclusion follows from Lems. A.3–A.2. □

Lemma A.5 (Davis–Kahan for a simple leading spike)**.** *Let* A_0, A *be real symmetric with* $\|A - A_0\|_{\mathrm{op}} \le \eta$*. Suppose* A_0 *has a simple leading eigenvalue* $\lambda_1(A_0)$ *with unit eigenvector* u_0 *and gap* $\delta_0 := \lambda_1(A_0) - \lambda_2(A_0) > 0$*. If* $\eta < \delta_0/2$*, then the leading eigenvalue of* A *is simple, and any unit leading eigenvector* u *of* A *satisfies*

$$\sin\angle(u, u_0) \;\le\; \frac{2\eta}{\delta_0}. \tag{7}$$

Proof. By Weyl, $|\lambda_j(A) - \lambda_j(A_0)| \le \eta$, so $\lambda_1(A) - \lambda_2(A) \ge \delta_0 - 2\eta > 0$ and the leading eigenspace of A is one-dimensional. The Davis–Kahan $\sin\Theta$ theorem in the form of (Yu et al., 2015) (building on Davis & Kahan (1970); Wedin (1972)) yields $\|\sin\Theta\|_{\mathrm{F}} \le 2\eta/\delta_0$ for the angle between the leading invariant subspaces; in dimension one this is (7). □

A.2 NON-UNIQUENESS (PROP. 3.1)

Residual EOF1 from Σ alone is not a certified anomaly label. The same matrix admits both a pure-background and an anomaly-plus-background decomposition (Table 1, first row).

Restatement (Prop. 3.1). For any $\Sigma \succ 0$ there exist at least two decompositions reproducing Σ: *(i)* pure background $\beta = 0$, $C = \Sigma$; *(ii)* top-eigenvalue split $\beta = \lambda_{\max}(\Sigma)/2$, $v_\star = u_{\mathrm{var}}$, and $C = \Sigma - \beta u_{\mathrm{var}} u_{\mathrm{var}}^\top$ (hence $C \succ 0$). These differ in β for every $\Sigma \succ 0$. If $\lambda_{\min}(\Sigma) < \lambda_{\max}(\Sigma)$ and $\lambda_2(\Sigma) = \cdots = \lambda_n(\Sigma) = \lambda_{\min}(\Sigma)$, the choice $\beta = \lambda_{\max}(\Sigma) - \lambda_{\min}(\Sigma)$ recovers the isotropic spike $C = \lambda_{\min}(\Sigma)\, I$. If Σ is a multiple of the identity, every direction is an EOF1. In all cases Σ alone does not determine whether EOF1 is anomaly or pure background structure.

Proof. Let $\Sigma = \sum_{i=1}^n \lambda_i u_i u_i^\top$ be an eigendecomposition with $\lambda_1 \geq \cdots \geq \lambda_n > 0$ and $u_{\mathrm{var}} = u_1$. Decomposition (i): set $\beta = 0$ and $C = \Sigma$. Then (1) holds and $C \succ 0$. Decomposition (ii): set $\beta := \lambda_1/2 \in (0, \lambda_1)$, $v_\star := u_1$, and $C := \Sigma - \beta u_1 u_1^\top$. On the eigenbasis, C has eigenvalues $\lambda_1/2, \lambda_2, \ldots, \lambda_n$, each positive, so $C \succ 0$. By construction $\beta v_\star v_\star^\top + C = \Sigma$, and $\beta > 0$ makes (ii) distinct from (i). The same argument gives a valid split for every $\beta \in (0, \lambda_1)$: the eigenvalue of C along u_1 is $\lambda_1 - \beta > 0$, and every other eigenvalue stays positive. If $\lambda_2 = \cdots = \lambda_n$ and $\beta = \lambda_1 - \lambda_n > 0$, that leading eigenvalue equals λ_n and $C = \lambda_n I$, the isotropic spike. If $\lambda_1 = \lambda_n$, then $\Sigma = \lambda_1 I$ and every unit vector is an EOF1; decomposition (ii) remains distinct because $\beta = \lambda_1/2 > 0$. In all cases, the map $\Sigma \mapsto$"EOF1 is the anomaly" is not a function of Σ. □

A.3 ALIGNED COLLAPSE (PROP. 3.2)

Aligned anomalies collapse into the background spike, so second-order statistics of Σ alone cannot separate them (Table 1, aligned column).

Restatement (Prop. 3.2). If $\beta > 0$, $v_\star = \pm v_c$, and $C = I + \gamma v_c v_c^\top$, then $\Sigma = I + (\gamma + \beta) v_c v_c^\top$, a pure-background decomposition of strength $\gamma + \beta$.

Proof. Substitute $v_\star = \pm v_c$ into (1):

$$\Sigma = \beta v_c v_c^\top + I + \gamma v_c v_c^\top = I + (\gamma + \beta) v_c v_c^\top.$$

The right-hand side is the spiked-background covariance with anomaly strength 0 and background strength $\gamma + \beta$ along v_c. Hence the anomaly-plus-background decomposition $(\beta, v_c, I + \gamma v_c v_c^\top)$ and the pure-background decomposition $(0, v_c, I + (\gamma + \beta) v_c v_c^\top)$ induce the same Σ, so no functional of Σ alone can separate them. □

A.4 KNOWN BACKGROUND AXIS

With known v_c, identification holds only when the anomaly has a nontrivial orthogonal share ($s > 0$). Full alignment remains collapsed as in Prop. 3.2.

Claim (known v_c identification). In the spiked-background family $C = I + \gamma v_c v_c^\top$ and $v_\star = \alpha v_c + s v_\perp$ with $\alpha \in [0, 1]$, $s = \sqrt{1 - \alpha^2}$, and $v_\perp \perp v_c$ a unit vector, assume v_c is known. If $s > 0$ and $\beta > 0$, then (β, α, γ) are uniquely recoverable from Σ; if $s = 0$, then β and γ remain observationally collapsed as in Prop. 3.2.

Proof. First recover the unknown innovation direction. Let $P := I - v_c v_c^\top$. On $\mathrm{range}(P)$,

$$P \Sigma P = P + \beta s^2 v_\perp v_\perp^\top.$$

If $s > 0$ and $\beta > 0$, its unique direction with eigenvalue above 1 is $v_\perp$ up to sign; choose the sign so that $v_c^\top \Sigma v_\perp \geq 0$. We may now work in the identified plane $\mathrm{span}\{v_c, v_\perp\}$ (the remaining orthogonal complement is isotropic). Write $\Sigma_{cc} := v_c^\top \Sigma v_c$, $\Sigma_{\perp\perp} := v_\perp^\top \Sigma v_\perp$, and $\Sigma_{c\perp} := v_c^\top \Sigma v_\perp$. Expanding $\Sigma = C + \beta v_\star v_\star^\top$ gives

$$\Sigma_{cc} = 1 + \gamma + \beta\alpha^2, \tag{8}$$

$$\Sigma_{\perp\perp} = 1 + \beta s^2, \tag{9}$$

$$\Sigma_{c\perp} = \beta \alpha s. \tag{10}$$

If $s = 0$, then $\alpha = 1$ under the sign convention and $\Sigma = I + (\gamma + \beta)v_c v_c^\top$, so only the sum $\gamma + \beta$ is identified, recovering Prop. 3.2. If $s > 0$ and $\beta > 0$, then $\Sigma_{\perp\perp} > 1$, and

$$\frac{\alpha}{s} = \frac{\Sigma_{c\perp}}{\Sigma_{\perp\perp} - 1}, \qquad \beta = \frac{\Sigma_{\perp\perp} - 1}{s^2}, \qquad \gamma = \Sigma_{cc} - 1 - \beta\alpha^2,$$

with $s = \sqrt{1 - \alpha^2}$ recovered from the ratio α/s by $\alpha = \rho/\sqrt{1 + \rho^2}$ and $s = 1/\sqrt{1 + \rho^2}$ for $\rho := \Sigma_{c\perp}/(\Sigma_{\perp\perp} - 1)$. Thus (β, α, γ) are uniquely determined. (If $\beta = 0$ and $s > 0$, then $\Sigma_{\perp\perp} = 1$ and the ratio is undefined; there is no anomaly to identify.) □

A.5 Contaminated gap law (Thm. 3.1)

For partial contamination the calibrated direction is fixed while the whitened gap shrinks; at exact same-record the direction is unidentified (Fig. 1(c); Table 1, second row).

Restatement (Thm. 3.1). Under (2), let $C \succ 0$, $\varepsilon \in [0, 1)$, $\beta > 0$, and $Q := v_\star^\top C^{-1} v_\star > 0$. Define $M_\varepsilon := C_\varepsilon^{-1/2} \Sigma C_\varepsilon^{-1/2}$ and $\lambda_\varepsilon := \lambda_1(M_\varepsilon)$. Then $u_{\mathrm{snr}}(C_\varepsilon) \propto C_\varepsilon^{-1} v_\star \propto C^{-1} v_\star$, and

$$\lambda_\varepsilon = 1 + \frac{(1 - \varepsilon)\beta Q}{1 + \varepsilon\beta Q}, \qquad \delta_\star(\varepsilon) = \lambda_\varepsilon - 1 = \frac{(1 - \varepsilon)\beta Q}{1 + \varepsilon\beta Q},$$

decreases in ε with $\delta_\star(\varepsilon) \to 0$ as $\varepsilon \to 1$. For every threshold $\tau \in (0, \beta Q)$ the gap falls through τ at $\varepsilon(\tau) = (\beta Q - \tau)/(\beta Q(1 + \tau))$. The calibrated direction is invariant for $\varepsilon < 1$ and has a stable limit as $\varepsilon \to 1^-$; at $\varepsilon = 1$ it is unidentified.

Proof. From (2), $C_\varepsilon = C + \varepsilon\beta\, v_\star v_\star^\top$ and $\Sigma = C + \beta\, v_\star v_\star^\top = C_\varepsilon + (1 - \varepsilon)\beta\, v_\star v_\star^\top$. Sherman–Morrison yields

$$C_\varepsilon^{-1} = C^{-1} - \frac{\varepsilon\beta\, C^{-1} v_\star v_\star^\top C^{-1}}{1 + \varepsilon\beta Q},$$

hence

$$C_\varepsilon^{-1} v_\star = \frac{C^{-1} v_\star}{1 + \varepsilon\beta Q}. \tag{11}$$

In particular, $C_\varepsilon^{-1} v_\star$ is a positive multiple of $C^{-1} v_\star$. Thus the unit SNR directions $u_{\mathrm{snr}}(C_\varepsilon) \propto C_\varepsilon^{-1} v_\star$ and $u_{\mathrm{snr}}(C) \propto C^{-1} v_\star$ coincide up to sign for every $\varepsilon \in [0, 1)$.

It remains to identify the top generalised eigenpair of (Σ, C_ε). Whitening gives

$$M_\varepsilon := C_\varepsilon^{-1/2} \Sigma C_\varepsilon^{-1/2} = I + (1 - \varepsilon)\beta\, w_\varepsilon w_\varepsilon^\top, \qquad w_\varepsilon := C_\varepsilon^{-1/2} v_\star, \tag{12}$$

so the eigenvalues of M_ε are $1 + (1 - \varepsilon)\beta\|w_\varepsilon\|^2$ (multiplicity one) and 1 (multiplicity $n - 1$). From (11),

$$\|w_\varepsilon\|^2 = v_\star^\top C_\varepsilon^{-1} v_\star = \frac{Q}{1 + \varepsilon\beta Q},$$

and therefore

$$\lambda_\varepsilon := \lambda_1(M_\varepsilon) = 1 + \frac{(1 - \varepsilon)\beta Q}{1 + \varepsilon\beta Q}, \qquad \delta_\star(\varepsilon) := \lambda_\varepsilon - 1 = \frac{(1 - \varepsilon)\beta Q}{1 + \varepsilon\beta Q}.$$

The map $\varepsilon \mapsto \delta_\star(\varepsilon)$ is strictly decreasing on $[0, 1)$ because its derivative equals $-\beta Q(1 + \beta Q)/(1 + \varepsilon\beta Q)^2 < 0$, and $\delta_\star(\varepsilon) \to 0$ as $\varepsilon \to 1^-$. Since $\delta_\star(0) = \beta Q$, the equation $\delta_\star(\varepsilon) = \tau$ has the unique root $\varepsilon(\tau) = (\beta Q - \tau)/(\beta Q(1 + \tau))$ in $(0, 1)$ for every $\tau \in (0, \beta Q)$. The top eigenvector of M_ε is $w_\varepsilon/\|w_\varepsilon\|$, which maps back under $C_\varepsilon^{-1/2}$ to a multiple of $C_\varepsilon^{-1} v_\star$, i.e. to $u_{\mathrm{snr}}(C_\varepsilon)$. Thus the calibrated direction is invariant for $\varepsilon < 1$ and has a stable one-sided limit while the gap shrinks to zero; exact same-record has no distinguished eigenvector. □

A.6 Same-record non-certifiability (Thm. 3.2, Cor. 3.1)

Exact same-record ($C \leftarrow \Sigma$) drives the population gap to zero, so the gap never clears any $\tau > 0$. Aligned collapse needs side information even at infinite samples.

Restatement (Thm. 3.2). If $C_\star = \Sigma$, then $C_\star^{-1/2}\Sigma C_\star^{-1/2} = I$ and the population gap is zero. Any deterministic binary rule based only on (Σ, Σ) that must label the observation as anomaly or pure background and never abstains fails on some decomposition in Prop. 3.1. If $\hat{C} = \hat{\Sigma} \succ 0$, then $\hat{M} = I$, so $\hat{\delta} = g = 0$ and every $\tau > 0$ rejects the eigen-gap guard.

Proof. If $C_\star = \Sigma \succ 0$, then $M_\star = C_\star^{-1/2}\Sigma C_\star^{-1/2} = I$, so $\lambda_1(M_\star) = \lambda_2(M_\star) = 1$ and $\delta_\star = \lambda_1(M_\star) - \max(1, \lambda_2(M_\star)) = 0$.

Suppose a deterministic rule r maps every pair (Σ, Σ) to $\{\texttt{anomaly}, \texttt{background}\}$ (no abstention). Fix any $\Sigma \succ 0$. Prop. 3.1 establishes two decompositions of this same Σ: one with $\beta = 0$ (pure background) and one with $\beta = \lambda_{\max}(\Sigma)/2 > 0$ (top-eigenvalue split, $v_\star = u_{\text{var}}$). Any correct rule would have to return `background` on the first decomposition and `anomaly` on the second, which is impossible for a function of (Σ, Σ) alone. Hence any rule that never abstains fails on some decomposition in Prop. 3.1.

Finally, if $\hat{C} = \hat{\Sigma} \succ 0$, then $\hat{M} = I$ identically, so $\hat{\delta} = \lambda_1(\hat{M}) - \max(1, \lambda_2(\hat{M})) = 0$ and $g = \lambda_1(\hat{M}) - 1 = 0$. Thus every $\tau > 0$ rejects Alg. 1's eigen-gap guard. □

Restatement (Cor. 3.1). Under Prop. 3.2, even infinite samples cannot certify anomaly versus a larger background spike from second-order statistics of Σ alone. Some side information beyond Σ is necessary to separate those two decompositions; quiet held-out C suffices whenever $\beta > 0$, because $\Sigma - C$ then recovers the anomaly direction $v_\star$.

Proof. Prop. 3.2 exhibits two observationally identical second-order models for the same Σ: an anomaly of strength β on top of background strength γ, and a pure background of strength $\gamma + \beta$. Infinite samples identify Σ exactly, but not which of these two decompositions generated it. Any label from Σ alone is therefore incorrect on one of the two models. Some side information beyond Σ is necessary to break the equivalence. Quiet held-out C suffices whenever $\beta > 0$, because then $\Sigma - C = \beta v_\star v_\star^\top$ recovers the anomaly direction. □

A.7 Finite-sample gap control (Thm. 3.3)

A whitened gap survives sampling only if it exceeds concentration; as contamination approaches same-record, $\hat{\delta}$ stays below every fixed $\tau > \bar{\eta}_M$ with the stated pointwise probability—even if the plotted direction looks unchanged.

Restatement (Thm. 3.3). Under Assumption 3.1 and (2) with $\varepsilon \in [0, 1)$, assume $C_\varepsilon \succeq \mu I$ for a constant $\mu > 0$ (e.g. whenever $C \succeq \mu I$), fix failure level $t \geq 1$, and write $\eta := \max(\eta_\Sigma(T, t), \eta_C(T_c, t))$ and η_M as in Lem. A.2. Assume the sample sizes are large enough that $\eta \leq \mu/4$. If $\delta_\star \geq 4\eta_M$, then with probability at least $1 - 2e^{-t}$: the empirical gap satisfies $|\hat{\delta} - \delta_\star| \leq 2\eta_M$, and hence *(i)* $\hat{\delta} := \lambda_1(\hat{M}) - \max(1, \lambda_2(\hat{M})) \geq \frac{1}{2}\delta_\star$; *(ii)* if $u_\star$ and $\hat{u}$ are top eigenvectors of $M_\star$ and $\hat{M}$, then $\sin\angle(\hat{u}, u_\star) \leq 2\eta_M/\delta_\star$; *(iii)* for $\tau \in [\eta_M, \frac{1}{2}\delta_\star]$, Alg. 1's gap guard passes by (i). More generally, a fixed gap guard passes when $\delta_\star \geq \tau + 2\eta_M$ and rejects when $\delta_\star < \tau - 2\eta_M$ on the same event. The angle in (ii) is a whitened-coordinate statement and is not transferred unchanged through the nonorthogonal whitening maps. The excess $g := \lambda_1(\hat{M}) - 1$ obeys $g \geq \hat{\delta}$, with equality when $\lambda_2(\hat{M}) \leq 1$. Conversely, assume $C_\varepsilon \succeq \mu I$ uniformly in $\varepsilon \in [0, 1)$ and $\bar{\eta}_M := \sup_{\varepsilon<1} \eta_M(\varepsilon) < \infty$. At fixed (T, T_c), $\delta_\star(\varepsilon) \to 0$ as $\varepsilon \to 1$. For every fixed CAP threshold $\tau > \bar{\eta}_M$ and each sufficiently large $\varepsilon < 1$, $g < \tau$ and therefore $\hat{\delta} \leq g < \tau$ with probability at least $1 - 2e^{-t}$. On that event, if a residual spike remains, Alg. 1 abstains with `gap_fail`; if the spike has fallen below τ_{spike}, Alg. 1 returns `no_detectable_residual`.

Proof. Apply Lem. A.1 under Assump. 3.1 (independent analysis and calibration samples) and set $\eta := \max(\eta_\Sigma(T, t), \eta_C(T_c, t))$. The hypothesis $\eta \leq \mu/4$ is a deterministic sample-size condition on these concentration radii. On the event of probability at least $1 - 2e^{-t}$ where both covariance bounds hold, Lem. A.2 supplies $\|\hat{M} - M_\star\|_{\text{op}} \leq \eta_M$.

Under (2), App. A.5 gives $M_\star = I + (1-\varepsilon)\beta\, w_\varepsilon w_\varepsilon^\top$ with $\lambda_1(M_\star) = 1+\delta_\star$, $\lambda_{j\geq 2}(M_\star) = 1$, and whitened top eigenvector $u_\star \propto C_\varepsilon^{-1/2} v_\star$. Weyl yields $|\lambda_j(\hat{M}) - \lambda_j(M_\star)| \leq \eta_M$ for all j, hence

$$\lambda_1(\hat{M}) \geq 1 + \delta_\star - \eta_M, \qquad \lambda_2(\hat{M}) \leq 1 + \eta_M.$$

Therefore

$$\hat{\delta} = \lambda_1(\hat{M}) - \max\big(1, \lambda_2(\hat{M})\big) \geq (1+\delta_\star - \eta_M) - (1+\eta_M) = \delta_\star - 2\eta_M.$$

Conversely, $\lambda_1(\hat{M}) \leq 1 + \delta_\star + \eta_M$ and $\max(1, \lambda_2(\hat{M})) \geq 1$, so $\hat{\delta} \leq \delta_\star + \eta_M$. Together these inequalities give $|\hat{\delta} - \delta_\star| \leq 2\eta_M$. The hypothesis $\delta_\star \geq 4\eta_M$ gives $\hat{\delta} \geq \frac{1}{2}\delta_\star$, which is (i).

For (ii), apply Lem. A.5 to $A_0 = M_\star$ and $A = \hat{M}$ with gap $\delta_0 = \delta_\star$ and perturbation $\eta_M \leq \delta_\star/4 < \delta_\star/2$, obtaining $\sin\angle(\hat{u}, u_\star) \leq 2\eta_M/\delta_\star$.

For (iii), Alg. 1 thresholds $\hat{\delta}$. Part (i) gives $\hat{\delta} \geq \frac{1}{2}\delta_\star \geq \tau$ whenever $\tau \leq \frac{1}{2}\delta_\star$ and $\delta_\star \geq 4\eta_M$. The two-sided bound also gives the general pass condition $\delta_\star \geq \tau + 2\eta_M$ and rejection condition $\delta_\star < \tau - 2\eta_M$. The lower constraint $\tau \geq \eta_M$ keeps the threshold above the concentration scale. The leading generalised eigenvector of $(\hat{\Sigma}, \hat{C})$ is the image of $\hat{u}$ under $\hat{C}^{-1/2}$, but a nonorthogonal map need not preserve angles; part (ii) therefore controls only the whitened eigenvectors. Alg. 1 returns $\mathrm{top}(\hat{R})$, not that generalised eigenvector. The excess $g = \lambda_1(\hat{M}) - 1$ satisfies $g \geq \hat{\delta}$, with equality whenever $\lambda_2(\hat{M}) \leq 1$. By Weyl, $g \geq \delta_\star - \eta_M$, so $\delta_\star \geq 2\eta_M$ already yields $g \geq \delta_\star/2$; that fact is about g, not about Alg. 1's guard. The angle bound of (ii) needs $\eta_M < \delta_\star/2$ for Lem. A.5.

Conversely, assume $C_\varepsilon \succeq \mu I$ for all $\varepsilon \in [0,1)$ and define the finite uniform bound $\bar{\eta}_M := \sup_{\varepsilon<1} \eta_M(\varepsilon)$. The lower bound is automatic if $C \succeq \mu I$, because $C_\varepsilon = C + \varepsilon\beta v_\star v_\star^\top \succeq C$. By Thm. 3.1, $\delta_\star(\varepsilon) \to 0$. On the high-probability event of Lem. A.1, $\lambda_1(\hat{M}) - 1 \leq \delta_\star + \eta_M$. For every fixed $\tau > \bar{\eta}_M$ one eventually has $\delta_\star + \eta_M \leq \delta_\star + \bar{\eta}_M < \tau$, so $g < \tau$ and therefore $\hat{\delta} \leq g < \tau$ with probability at least $1 - 2e^{-t}$. On that event Alg. 1 does not return `certified_direction`: if $\|\hat{R}\|_{\mathrm{op}} \geq \tau_{\mathrm{spike}}$ it abstains with `gap_fail`, and if $\|\hat{R}\|_{\mathrm{op}} < \tau_{\mathrm{spike}}$ it returns `no_detectable_residual`. □

Corollary A.1 (Finite-sample contamination band)**.** *Let $A := \beta Q$, fix $\tau_{\mathrm{gap}} > 2\bar{\eta}_M$, and define $e(a) := (A-a)/(A(1+a))$ for $a \in (0, A)$. If $\tau_{\mathrm{gap}} + 2\bar{\eta}_M < A$, then for each fixed $\varepsilon \leq e(\tau_{\mathrm{gap}} + 2\bar{\eta}_M)$ the gap guard passes with probability at least $1 - 2e^{-t}$. For each fixed $\varepsilon > e(\tau_{\mathrm{gap}} - 2\bar{\eta}_M)$, it rejects with the same probability. The interval between these crossings is the finite-sample ambiguity band.*

Proof of Cor. A.1. The two-sided gap bound in Thm. 3.3 yields the stated pass and rejection conditions after replacing η_M by its uniform upper bound $\bar{\eta}_M$. By Thm. 3.1, $\delta_\star(\varepsilon) = A(1-\varepsilon)/(1+A\varepsilon)$ is strictly decreasing, with $\delta_\star(\varepsilon) = a$ at $\varepsilon = e(a)$. Substitution yields the two crossings; the probability statement is pointwise in each fixed ε. □

Corollary A.2 (Observable Gaussian radii)**.** *If both samples are Gaussian, let $d_N = N$ for a known mean and $d_N = N - 1$ for an estimated mean with unbiased covariance, and set $q_N = (\sqrt{n} + \sqrt{2(t + \log 4)})/\sqrt{d_N}$ and $a_N = 2q_N + q_N^2$. When $q_T, q_{T_c} < 1$, both relative covariance bounds hold jointly with probability at least $1 - e^{-t}$ using (a_T, a_{T_c}); when also $a_T, a_{T_c} < 1$, Lem. A.3 with explicit whitening constant $c_2 = 2$ gives computable theorem-mode radii without unknown concentration constants.*

Proof of Cor. A.2. Apply Lem. A.4, then Lem. A.3 with the explicit $c_2 = 2$ from Lem. A.2. Every resulting radius and threshold depends only on observed quantities and prespecified inputs. □

A.8 Quiet residual spike (Thm. 4.1)

With quiet held-out C, a residual spike is identifiable above noise, but that spike is $\mathrm{top}(\Sigma - C)$, not residual EOF1 of Σ (Table 1, third row).

Restatement (Thm. 4.1). Assume $\varepsilon = 0$, $C_\star = C$, $\Sigma = C + \beta v_\star v_\star^\top$ with $\beta > 0$, and Assumption 3.1. Fix $t \geq 1$ and write $\eta := \max(\eta_\Sigma(T,t), \eta_C(T_c,t))$. If $\beta > 4\eta$, then with probability at least $1 - 2e^{-t}$: $\big|\|\hat{R}\|_{\mathrm{op}} - \beta\big| \leq 2\eta$ and the leading eigenvector $\hat{v}$ of $\hat{R}$ satisfies $\sin\angle(\hat{v}, v_\star) \leq 4\eta/\beta$, by Lem. A.5 on a perturbation of size 2η under quiet held-out C.

Proof. On the event of Lem. A.1, $\|\hat{\Sigma} - \Sigma\|_{\mathrm{op}} \le \eta_\Sigma \le \eta$ and $\|\hat{C} - C\|_{\mathrm{op}} \le \eta_C \le \eta$. Write $R := \Sigma - C = \beta v_\star v_\star^\top$ and $A := \hat{\Sigma} - \hat{C}$, so

$$A - R = (\hat{\Sigma} - \Sigma) - (\hat{C} - C)$$

and $\|A - R\|_{\mathrm{op}} \le \eta_\Sigma + \eta_C \le 2\eta$. By Weyl, $|\lambda_1(A) - \beta| \le 2\eta$. Under $\beta > 4\eta$ one has $\lambda_1(A) \ge \beta - 2\eta > 2\eta > 0$. The matrix $\hat{R} := \mathrm{PSD}(A)$ therefore shares the leading eigenpair of A. Zeroing nonpositive eigenvalues does not alter an eigenvector with a strictly positive eigenvalue, and $\lambda_1(A) > 0$ is the unique candidate for $\|\hat{R}\|_{\mathrm{op}}$. Hence $\|\hat{R}\|_{\mathrm{op}} = \lambda_1(A)$ and $\big|\|\hat{R}\|_{\mathrm{op}} - \beta\big| \le 2\eta$.

For the angle, $R = \beta v_\star v_\star^\top$ has leading gap β. The perturbation satisfies $\|A - R\|_{\mathrm{op}} \le 2\eta < \beta/2$ because $\beta > 4\eta$. Lem. A.5 applied to the pair (R, A) therefore yields $\sin\angle(\hat{v}, v_\star) \le 2 \cdot (2\eta)/\beta = 4\eta/\beta$ for any unit leading eigenvector $\hat{v}$ of A, equivalently of $\hat{R}$. □

Proposition A.1 (Residual strength and background-normalised separation)**.** *Under quiet $C \succ 0$ and $\Sigma = C + \beta v_\star v_\star^\top$, the residual matrix has leading eigengap β and direction $v_\star$, whereas $M = I + \beta(C^{-1/2}v_\star)(C^{-1/2}v_\star)^\top$ has leading gap $\delta_\star = \beta v_\star^\top C^{-1} v_\star$ and whitened direction proportional to $C^{-1/2}v_\star$. If $\mu I \preceq C \preceq LI$, then $\beta/L \le \delta_\star \le \beta/\mu$. Without both spectral bounds, neither margin controls the other: for $C_L = I + (L-1)v_\star v_\star^\top$, β stays fixed while $\delta_\star = \beta/L \to 0$; for $\beta_k \downarrow 0$ and $C_k = I + (\beta_k - 1)v_\star v_\star^\top$, $\delta_\star = 1$ while the residual spike vanishes. Thus a failed whitened-gap guard does not logically establish instability of the residual direction returned by CAP. Conversely, a large whitened gap need not imply an absolutely resolved residual. The conjunction is therefore logically minimal for a report that asserts both properties, although it is not a claim of uniformly optimal detection.*

Proof. Both rank-one eigendecompositions are immediate from the displayed matrices. The Loewner inequalities $L^{-1}I \preceq C^{-1} \preceq \mu^{-1}I$ give the bounds after taking the quadratic form at unit $v_\star$; the two diagonal rank-one constructions give the stated limits. □

Remark A.1 (Threshold interval). On the event of Thm. 4.1, $\|\hat{R}\|_{\mathrm{op}} \ge \beta - 2\eta$. Any *oracle* threshold $\tau_{\mathrm{spike}} \in [2\eta, \beta - 2\eta]$ therefore accepts the residual spike whenever that interval is nonempty, which holds for $\beta \ge 4\eta$. The angle in Thm. 4.1 needs the strict inequality $\beta > 4\eta$, so that the perturbation 2η lies strictly below half the gap and Lem. A.5 applies. The midpoint $\beta/2$ is a population reference when $\beta \ge 4\eta$. Reported runs use the declared numerical cutoffs of App. B, not the unknown β.

A.9 EOF1–CAP DISAGREEMENT (PROP. 4.1)

Under quiet C, certifying $\mathrm{top}(\Sigma - C)$ is not the same as reporting residual EOF1. Under high alignment, u_{var} can track the background while the residual spike points elsewhere.

Restatement (Prop. 4.1). In the spiked-background family of §2 with quiet C, there is an open set of (α, β, γ) on which $\Delta_{\mathrm{CAP}}(\Sigma; C) > 0$: u_{var} can nearly align with the background axis v_c while $\Sigma - C$ still recovers $v_\star$. A concrete witness is $(\alpha, \beta, \gamma) = (0.8, 1, 8)$ in $n = 2$. Under quiet C, CAP's theorem-backed path therefore targets $\mathrm{top}(\Sigma - C)$, not residual EOF1 of Σ.

Proof. Take $n = 2$, $v_c = e_1$, $v_\perp = e_2$, and $v_\star = \alpha v_c + s v_\perp$ with $s = \sqrt{1 - \alpha^2}$. Quiet $C = I + \gamma v_c v_c^\top$ and $\Sigma = C + \beta v_\star v_\star^\top$ give the exact residual $R = \Sigma - C = \beta v_\star v_\star^\top$, so $\mathrm{top}(\Sigma - C) = v_\star$ (up to sign) whenever $\beta > 0$. The SNR direction is $u_{\mathrm{snr}}(C) \propto C^{-1} v_\star$. Since $C^{-1} = I - \frac{\gamma}{1+\gamma} v_c v_c^\top$,

$$C^{-1} v_\star = v_\star - \frac{\gamma\alpha}{1+\gamma} v_c = \frac{\alpha}{1+\gamma} v_c + s v_\perp,$$

hence u_{snr} has a nontrivial $v_\perp$ component whenever $s > 0$.

Now fix the witness $(\alpha, \beta, \gamma) = (0.8, 1, 8)$. In the $\{v_c, v_\perp\}$ basis one has

$$\Sigma = \begin{pmatrix} 1 + \gamma + \beta\alpha^2 & \beta\alpha s \\ \beta\alpha s & 1 + \beta s^2 \end{pmatrix} = \begin{pmatrix} 9.64 & 0.48 \\ 0.48 & 1.36 \end{pmatrix}$$

(here $s = 0.6$). The larger eigenvalue of this matrix is simple, and its unit eigenvector u_{var} satisfies $|\langle u_{\mathrm{var}}, v_c \rangle| > 0.998$. With $u_{\mathrm{snr}}(C) \propto (\alpha/(1+\gamma), s)$ one obtains

$$\Delta_{\mathrm{CAP}}(\Sigma; C) = 1 - \big(u_{\mathrm{var}}^\top v_\star\big)^2 \approx 0.306, \qquad \Delta_{\mathrm{SNR}}(\Sigma; C) = 1 - \big(u_{\mathrm{var}}^\top u_{\mathrm{snr}}(C)\big)^2 \approx 0.959,$$

while $|\langle u_{\mathrm{snr}}, v_\star\rangle| > 0.71$. Thus u_{var} tracks v_c while both u_{snr} and $\mathrm{top}(\Sigma - C)$ retain a nontrivial innovation component.

The maps $(\alpha, \beta, \gamma) \mapsto \Sigma(\alpha, \beta, \gamma)$, $(\alpha, \beta, \gamma) \mapsto u_{\mathrm{var}}(\Sigma)$, and $(\alpha, \beta, \gamma) \mapsto u_{\mathrm{snr}}(C)$ are continuous on the open set $\{0 < \alpha < 1, \beta > 0, \gamma > 0\}$. The leading eigenvalues remain simple in a neighbourhood of the witness because the witness has positive gap $\lambda_1(\Sigma) - \lambda_2(\Sigma)$ and eigenvalues are continuous by Weyl. Hence both disagreements in (4) are continuous there, and $\Delta_{\mathrm{CAP}}(\Sigma; C)$ remains strictly positive on a nonempty open neighbourhood of $(0.8, 1, 8)$. On that neighbourhood, certifying $\mathrm{top}(\Sigma - C)$ is not the same as reporting residual EOF1 of Σ. □

A.10 CAP DECISION BEHAVIOUR (COR. 5.1)

We prove CAP's decision behaviour under declared side information. On the Σ-alone path CAP abstains; under quiet held-out C it returns $\mathrm{top}(\hat{R})$—not u_{var}—only when spike and gap both clear.

Restatement (Cor. 5.1). (i) $m =$ `none` implies deterministic abstention. *(ii)* Matching provenance identifiers or exact numerical equality $\hat{C} = \hat{\Sigma}$ on an exact same-record input imply deterministic abstention. *(iii)* Under Assumption 3.1, $C \succeq \mu I$, $\eta \leq \mu/4$, quiet C, $\beta > 4\eta$, and $\delta_\star \geq 4\eta_M$, thresholds in the oracle intervals $\tau_{\mathrm{spike}} \in [2\eta, \beta - 2\eta]$ and $\tau_{\mathrm{gap}} \in [\eta_M, \frac{1}{2}\delta_\star]$ yield guard acceptance as $\mathrm{top}(\hat{R})$ on the joint concentration event, provided $\hat{C} \neq \hat{\Sigma}$ and no earlier prespecified conditioning guard fires; that event has probability at least $1 - 2e^{-t}$. Declared mode labels the result `guard_pass_direction` and theorem mode labels it `certified_direction`; neither status assigns physical anomaly semantics. *(iv)* Separately, as $\varepsilon \to 1$ under uniform $C_\varepsilon \succeq \mu I$ and finite $\bar{\eta}_M := \sup_{\varepsilon<1} \eta_M(\varepsilon)$, for every fixed $\tau_{\mathrm{gap}} > \bar{\eta}_M$ and each sufficiently large $\varepsilon < 1$, $\hat{\delta} < \tau_{\mathrm{gap}}$ with probability at least $1 - 2e^{-t}$. *(v)* Under the quiet null $\beta = 0$ and Assumption 3.1, every $\tau_{\mathrm{spike}} > 2\eta$ prevents either positive status with probability at least $1 - 2e^{-t}$. *(vi)* Under the quiet rank-one model and Assumption 3.1, for theorem-mode multiplier $m_0 > 6$ and valid observable radii $(\bar{\eta}, \bar{\eta}_M)$ from Lem. A.3, acceptance implies $\beta > 4\bar{\eta}$ and $\delta_\star > 4\bar{\eta}_M$ and carries the displayed data-dependent angle bounds of the main statement; if $\beta > (m_0 + 2)\bar{\eta}$ and $\delta_\star > (m_0 + 2)\bar{\eta}_M$, both guards pass.

Proof. (i) With $m =$ `none`, Alg. 1 returns `abstain` by construction. By Props. 3.1–3.2, Σ alone does not identify a unique decomposition.

(ii) If provenance identifiers match or $\hat{C} = \hat{\Sigma}$, Alg. 1 returns `abstain` with flag `same_record`. Thm. 3.2 shows that $\hat{\delta} = 0$ and $g = 0$ on this input (so every $\tau > 0$ rejects), and that any rule that never abstains fails on some decomposition in Prop. 3.1.

(iii) Take $\varepsilon = 0$ so $C_\star = C$. Assume Assump. 3.1, $C \succeq \mu I$, $\eta \leq \mu/4$, $\hat{C} \neq \hat{\Sigma}$, $\beta > 4\eta$, and $\delta_\star \geq 4\eta_M$. On the event of Lems. A.1–A.2, which has probability at least $1 - 2e^{-t}$, the following bounds hold. Thm. 4.1 gives $\|\hat{R}\|_{\mathrm{op}} \geq \beta - 2\eta \geq \tau_{\mathrm{spike}}$ for any oracle $\tau_{\mathrm{spike}} \in [2\eta, \beta - 2\eta]$, so the residual clears the spike guard. Thm. 3.3(i) gives $\hat{\delta} \geq \frac{1}{2}\delta_\star \geq \tau_{\mathrm{gap}}$ for any oracle $\tau_{\mathrm{gap}} \in [\eta_M, \frac{1}{2}\delta_\star]$. Alg. 1 therefore accepts direction $\mathrm{top}(\hat{R})$, returning `guard_pass_direction` in declared mode or `certified_direction` in theorem mode. Thm. 4.1 supplies $\sin\angle(\mathrm{top}(\hat{R}), v_\star) \leq 4\eta/\beta$. The Thm. 3.3 angle bound controls the whitened SNR direction, not the quiet-path vector returned by Alg. 1. Prop. 4.1 shows that this returned residual direction need not equal u_{var}.

(iv) Under (2) with $\varepsilon \to 1$ at fixed (T, T_c), $C_\varepsilon \succeq \mu I$ uniformly, and finite $\bar{\eta}_M$, Thm. 3.1 implies $\delta_\star \to 0$. The converse in Thm. 3.3 then yields, for any fixed $\tau_{\mathrm{gap}} > \bar{\eta}_M$ and each sufficiently large $\varepsilon < 1$, $\hat{\delta} \leq g < \tau_{\mathrm{gap}}$ with probability at least $1 - 2e^{-t}$. If $\|\hat{R}\|_{\mathrm{op}} \geq \tau_{\mathrm{spike}}$, Alg. 1 returns `abstain` with `gap_fail`; if $\|\hat{R}\|_{\mathrm{op}} < \tau_{\mathrm{spike}}$, it returns `no_detectable_residual`. In either case it returns neither positive status, matching the same-record row of Table 1.

(v) Under $\beta = 0$, the quiet population residual is $\Sigma - C = 0$. On the event of Lem. A.1, $\|\hat{\Sigma} - \hat{C}\|_{\mathrm{op}} \leq \eta_\Sigma + \eta_C \leq 2\eta$, hence $\|\mathrm{PSD}(\hat{\Sigma} - \hat{C})\|_{\mathrm{op}} \leq 2\eta < \tau_{\mathrm{spike}}$. If a provenance, equality, or conditioning guard fires, CAP abstains; otherwise it returns `no_detectable_residual` before the gap guard and never returns a certified direction.

(vi) On the event of Lem. A.3, write $s := \|\hat{R}\|_{\mathrm{op}}$ and $d := \hat{\delta}$. By Weyl, $|s - \beta| \le 2\bar{\eta}$ and $|d - \delta_\star| \le 2\bar{\eta}_M$. If theorem mode accepts, then $s \ge m_0\bar{\eta}$ and $d \ge m_0\bar{\eta}_M$; because $m_0 > 6$,

$$\beta \ge s - 2\bar{\eta} > 4\bar{\eta}, \qquad \delta_\star \ge d - 2\bar{\eta}_M > 4\bar{\eta}_M.$$

Substitute these observable lower bounds into Thms. 4.1 and 3.3(ii) to obtain the reported angle radii. Conversely, if $\beta > (m_0 + 2)\bar{\eta}$ and $\delta_\star > (m_0 + 2)\bar{\eta}_M$, then $s \ge \beta - 2\bar{\eta} > m_0\bar{\eta}$ and $d \ge \delta_\star - 2\bar{\eta}_M > m_0\bar{\eta}_M$, so both guards pass. The provenance and conditioning guards can only replace this statistical decision by abstention.

We treat the numerical cutoffs of App. B as declared inputs; the real-panel experiments do not verify oracle-interval membership. □

B CAP DECISION SPECIFICATION

The main decision cases follow Alg. 1; optional physical-v_c and fingerprint analyses remain exploratory and cannot produce either positive main-path conclusion. The abstention reasons are no side information, same-record reuse, ill-conditioned calibration, or an unresolved gap after the residual spike clears its threshold. The exploratory physical-v_c analysis also records EOF1–CAP disagreement and aligned collapse.

B.1 EXCHANGEABLE-CONTROL CALIBRATION

Proposition B.1 (Exchangeable-control calibration)**.** *After the representation and score maps are fixed independently of evaluation outcomes, let $(G_i, S_i)_{i=0}^m$ be exchangeable quiet-null gap and spike scores. For $H \in \{G, S\}$ define the conservative upper-tail rank*

$$p_H = \frac{1 + \sum_{i=1}^m \mathbf{1}\{H_i \ge H_0\}}{m + 1}.$$

Then $\Pr\{\max(p_G, p_S) \le \alpha\} \le \alpha$ *without assuming guard independence.*

Proof. Exchangeability makes the non-randomised upper-tail rank p_G super-uniform; using "$\ge$" preserves validity under ties. The joint return event satisfies $\{\max(p_G, p_S) \le \alpha\} \subseteq \{p_G \le \alpha\}$, hence its probability is at most α. The same argument could use p_S and requires no relationship between the two scores. If $\alpha < 1/(m+1)$, no return is attainable. The statement is marginal for one new record: simultaneous control over many future records needs an additional multiplicity rule. It also fails under distribution shift between development controls and the new null, and it supplies neither a direction-coverage guarantee nor evidence that the declared reference is scientifically quiet. □

Declared thresholds. We run the reported experiments with declared numerical thresholds, not theorem-mode radii. The primary synthetic run freezes $(\tau_{\mathrm{gap}}, \tau_{\mathrm{spike}}) = (0.290, 0.966)$ from an independent null pilot. INSTANCE freezes $(2.733, 5.030)$ from development-station noise, and NOAA/DCLDE freezes $(7.603, 7.305)$ from development-date quiet files. Etna freezes $(5.202, 6.593)$ from development-array quiet blocks, and TROPOMI freezes $(8.788, 115.899)$ from development-region off-source controls. The seismic case study uses $\tau_{\mathrm{gap}} = 0.380$ from event 1 quiet blocks and $\tau_{\mathrm{spike}} = 0.05\,\lambda_{\max}(\hat{C})$; exploratory SST uses $(0.05, 0.05\,\lambda_{\max}(\hat{C}))$. The dependent synthetic interpolation path uses $(0.08, 0.1)$, and the synthetic side-info grid uses $(0.05, 0.05)$. We do not claim these cutoffs equal the dimensioned covariance and whitened perturbation radii of theorem mode, or that any declared pair lies in the oracle intervals of Cor. 5.1. For INSTANCE we also report Prop. B.1's rank-calibrated rule at $\alpha = .05$ using the 381 development-station controls; its $1/382$ p-value resolution is adequate, but the held-out station split tests rather than guarantees exchangeability. Abstention under $m =$ `none` and exact same-record holds for every positive gap threshold. For partial contamination with $\varepsilon \to 1^-$, the pointwise high-probability non-acceptance result requires a fixed $\tau_{\mathrm{gap}} > \bar{\eta}_M$; smaller positive thresholds are not covered. Under quiet held-out C, $\hat{C} = \hat{\Sigma}$ forces `abstain` with `same_record`, and `gap_fail` likewise abstains (Alg. 1).

Theorem-aligned mode. The theorem-aligned analysis requires independent analysis and calibration samples as in Assumption 3.1; distinct provenance identifiers alone do not establish independence. With verified independence and zero PSD flooring, the general sub-Gaussian case takes finite

declared (K, μ, t, T, T_c, c_1) and uses $c_2 = 2$; the Gaussian case instead uses Cor. A.2 and needs no unknown concentration constant. When the Gaussian sample mean is estimated, the covariance must use the unbiased $1/(N-1)$ normalisation. CAP abstains if $a_\Sigma \geq 1$, $a_C \geq 1$, $\bar{\eta} > \mu/4$, or $\hat{C}$ fails the positive-definiteness or prespecified condition-number guard; the finite-eigenvalue check alone does not impose a finite maximum condition number. Otherwise, for a declared $m_0 > 6$, it sets

$$\tau_{\text{spike}} = m_0 \bar{\eta}, \qquad \tau_{\text{gap}} = m_0 \bar{\eta}_M.$$

An accepted result reports spike and gap margins, the lower bounds $\underline{\beta} = \|\hat{R}\|_{\text{op}} - 2\bar{\eta}$ and $\underline{\delta} = \hat{\delta} - 2\bar{\eta}_M$, the residual-angle radius $4\bar{\eta}/\underline{\beta}$, the whitened-angle radius $2\bar{\eta}_M/\underline{\delta}$, and failure probability at most $2e^{-t}$. Every certificate states the model, independence condition, normalisation, sample sizes, (μ, t, m_0), both thresholds and margins, population lower bounds, angle radii, and failure probability. The distributional assumptions, independence, and lower spectral bound must be justified independently of the residual; misspecification voids the theorem guarantee. The zero-floor requirement is essential: positive eigenvalue flooring can inflate the residual spike and is therefore unavailable in theorem mode.

Table 4: **Operational boundary relative to nearby covariance methods.** CAP reuses standard spectral objects but changes the permitted report: provenance and two resolved margins are required before returning a conditional direction claim.

Method	Object	Ordinary output	Boundary CAP adds
EOF / PCA	$\text{top}(\Sigma)$	leading variance direction	Σ alone is non-identifying
Contrastive PCA ($\alpha = 1$) (Abid et al., 2018)	$\text{top}(\Sigma - C)$	enriched contrast direction	provenance, spike/gap gates, abstention
Contrastive subspace test (Hawke et al., 2024)	test $V_X \subseteq V_Y$; estimate dimension	foreground-specific subspace evidence	audits the reference; resolves one leading contrast direction
Generalised covariance ratio / fingerprinting	leading (Σ, C) generalised eigenvector	ranked direction / prescribed-pattern evidence	ratio supplies a gap guard; CAP returns contrast top
Two-sample covariance testing (Cai et al., 2013)	test $\Sigma = C$	global rejection or support	rank-one direction plus observable margins
Selective prediction / reject option (Chow, 1970; Geifman & El-Yaniv, 2017)	labelled predictor and selection rule	risk–coverage tradeoff	abstention for an unlabelled spectral estimand
Spectral-projector inference (Koltchinskii & Lounici, 2017; Naumov et al., 2019)	eigenspace of a posited covariance	uncertainty set around that eigenspace	audits reference provenance and contrast margins first
Spiked PCA / spectral denoising (Gavish & Donoho, 2014; Zhang & Mondelli, 2024)	spike over a posited noise model	detection / direction	challenges the split and records reference scope

C Experimental studies

INSTANCE provides the primary out-of-station event/noise study; LOCATA supplies independently measured physical directions; NOAA and Etna cover low-signal regimes; and TROPOMI provides a geographically held-out replication. Additional seismic, SST, and dependent-interpolation analyses are sensitivity studies rather than confirmatory evidence.

INSTANCE primary benchmark. We use the official version-3 sample archive, whose 10,000 earthquake and 1,000 noise waveforms are three-component, 120-second, 100-Hz records with accompanying catalog metadata (Michelini et al., 2021). Of 341 network/station/channel identifiers shared by event and noise records, a prespecified outcome-blind rule defines 196 development and 145 test identifiers. No event outcome enters threshold selection. We filter each trace at 1–20 Hz. We use the first 30 seconds of a noise record for the mean, component scales, and $\hat{C}$, and its final

Table 5: **Negative controls and transfer limits** under frozen guards and prespecified physical units.

Question	Perturbation / unit	Observed result	Interpretation
Source pseudoreplication	one prespecified trace per earthquake	299 distinct sources	trace-count inflation
Same-record reuse	set $\hat{C} = \hat{\Sigma}$ for each source	$299/299$ abstain	forced binary label
Nuisance transport	baseline cutoff at $T/n = 3$	null pass 0.928	universal cutoff
Window transport	half-record guard on 2-s pairs	pass $22/29$, $6/8$	unit-free threshold
Timing destruction	100 circular-lag surrogates/event	all pass	coherence claim

disjoint 30 seconds for a null $\hat{\Sigma}$. For an event, the 30-second analysis window begins five seconds before the catalog P arrival and uses an independent same-key noise record for $\hat{C}$. We compute both matrices as second moments around the reference-noise mean and retain all three coordinates. We select one test record per source ID by a fixed deterministic ordering, yielding 299 distinct earthquakes, so the positive denominator is physical sources rather than correlated station traces. We freeze $(\tau_{\mathrm{gap}}, \tau_{\mathrm{spike}}) = (2.7325, 5.0298)$ at the marginal 0.95 development-noise quantiles. Development/test noise pass $16/381$ and $13/264$; events pass $240/299$. We report both unit-level Wilson intervals and a 20,000-resample percentile bootstrap over station/channel clusters (196 development, 145 test-noise, and 78 event stations); the latter gives 95% intervals $[.024, .063]$, $[.023, .080]$, and $[.717, .878]$. Magnitude-stratified counts for $(-\infty, 1.5), [1.5, 2), [2, 2.5), [2.5, 3), [3, \infty)$ are $(76, 48, 113, 42, 20)$ with pass rates $(.579, .833, .876, .905, .950)$. These fixed-threshold results measure out-of-station selectivity and do not invoke the Gaussian theorem assumptions.

Table 6: **Paired INSTANCE guard ablation under frozen development-noise cutoffs.** Entries are direction-return counts and rates on the identical held-out records; no threshold is refit. Forced-output methods have no mechanism for expressing no detectable residual or abstention, so this table compares return behaviour rather than directional accuracy.

Return rule	Required frozen guard(s)	Test noise	Test earthquakes
Forced EOF1 / residual top	none	$264/264$ (1.000)	$299/299$ (1.000)
Gap only	$\hat{\delta} \geq \tau_{\mathrm{gap}}$	$14/264$ $(.053)$	$250/299$ $(.836)$
Spike only	$\|\hat{R}\|_{\mathrm{op}} \geq \tau_{\mathrm{spike}}$	$15/264$ $(.057)$	$247/299$ $(.826)$
CAP	both guards	$13/264$ $(.049)$	$240/299$ $(.803)$

The paired outcomes show that neither guard is redundant. Among earthquakes, ten clear the gap but not the spike, and seven clear the spike but not the gap. On held-out noise, the joint rule removes one and two returns relative to the corresponding single-guard rules on the identical frozen records.

Matched-return comparison and real-background direction recovery. For a like-for-like scalar comparison, we re-freeze the gap-only and spike-only cutoffs on development noise so each returns exactly CAP's $16/381$ records; no test outcome enters either cutoff. Transported test noise/event counts are $13/264$ and $250/299$ for gap-only, $12/264$ and $243/299$ for spike-only, and $13/264$ and $240/299$ for CAP. The gap-only return set contains CAP's set on these test records and adds ten earthquakes but no noise; CAP deliberately gives up those returns because their absolute spike does not clear its separately frozen strength requirement. It is therefore a more selective decision rule, not a uniformly most powerful detector. We add three established scalar competitors under the identical matched-return protocol: Roy's excess largest generalised root (Roy, 1953), an equal-weight Gaussian log-determinant divergence proportional to the equal-size covariance likelihood-ratio score (Box, 1949), and the Frobenius distance of the whitened covariance spectrum from one. Their transported test noise/event counts are $13/264/254/299$, $7/264/248/299$, and $12/264/250/299$, respectively. Thus the first two outperform CAP's $13/264/240/299$ label discrimination on INSTANCE; these are empirically frozen gates, not uses of their parametric null laws, and each returns the same contrast direction when selected. As an external spectral-stability baseline, we independently resample the 30 nonoverlapping one-second blocks of each reference and analysis window 128 times (Künsch, 1989). Its score is the 0.10 quantile of absolute cosine between the resampled and full-sample contrast directions; the development cutoff is chosen to return the same $16/381$ null controls as CAP. Without test refitting it returns $11/264$ noise and $41/299$ earthquakes, showing that a stable contrast alone is not equivalent to CAP's two resolved margins. We treat this as a reproducible stability baseline, not a bootstrap confidence set with proven coverage under the waveform dependence.

We next retain each held-out noise trace's real 30-second reference and evaluation windows and add an independent zero-mean rank-one series $\sqrt{\beta} z v^{\top}$ to the standardised evaluation window. For

every record, we reproducibly generate one unit direction v and one unit-variance Gaussian series z, both fixed across $\beta/\tau_{\text{spike}} \in \{.05, .1, .25, .5, 1, 2, 4\}$; the 145 held-out station/channel identifiers remain the uncertainty clusters. This semi-synthetic intervention supplies direction ground truth without replacing the observed background covariance. At strengths at most .25, CAP's return rate stays .049 and its accurate-return rate is at most .019; at $1\times$ these become .492 and .470, and at $2\times$ they become 1.000 and .981 (station-cluster 95% interval $[.964, .996]$). At $1\times$, Roy, log-det, and whitened-Frobenius accurate-return rates are .337, .027, and .140; at $2\times$ they are .981, .042, and .951. The log-det gate's strong real-event separation therefore does not translate to the controlled rank-one direction target, while Roy catches CAP at $2\times$. At $4\times$, the accurate-return rate is .992. An unguarded contrast direction can already look stable below the frozen thresholds (median cosine .890 at $.05\times$), precisely the regime in which CAP declines to convert visual stability into a stronger report.

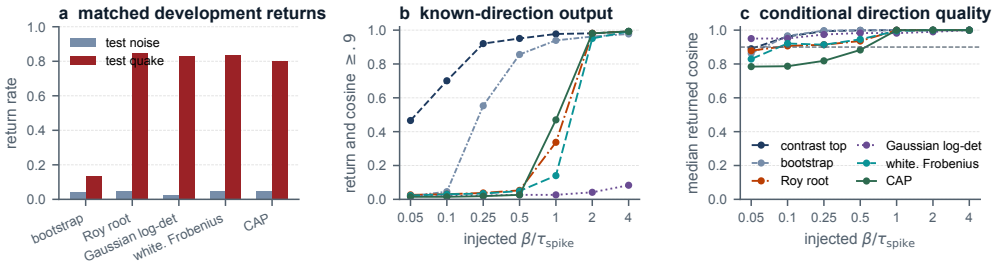

Figure 7: **Matched-return and known-direction comparisons on INSTANCE. (a)** Bootstrap, Roy, Gaussian log-det, whitened-Frobenius, and CAP cutoffs return the same $16/381$ development-noise records in count. **(b)** Accurate-return rate over 264 real-noise injections; forced methods always report, whereas selective methods may abstain. **(c)** Median cosine conditional on return separates the probability of a resolved direction from its conditional accuracy.

The repository provides the complete grid, station-cluster intervals, randomisation protocol, and record-level scores.

LOCATA acoustic direction replication. We use Task 1 of the final LOCATA release: a single static loudspeaker recorded at 48 kHz by the 15-microphone DICIT array in a real reverberant room (Evers et al., 2020). The corpus synchronizes the audio with an OptiTrack infrared-camera system that measures source and array position independently of the acoustic observations. The three official development recordings determine the representation and both guards; the separate 13-recording evaluation set remains untouched until these choices are fixed.

For each 2,048-sample Hann frame with 1,024-sample hop, six symmetric horizontal microphone pairs produce a GCC-PHAT spatial spectrum over $[-90°, 90°]$ in $5°$ steps using 300–3,000 Hz. The first half of the voice-inactive frames estimates the reference mean, scales, and covariance; the second half is a disjoint paired inactive control; and voice-active frames form the analysis sample. We add the same $0.05I$ ridge to both covariances and freeze $(\tau_{\text{gap}}, \tau_{\text{spike}}) = (22.830, 7.447)$ strictly above the largest of the three development-recording null scores. All three development sources pass with azimuth errors $1.3°$–$5.0°$, while all three inactive controls abstain. Optical positions remain unused until evaluation; their array-local azimuth follows the corpus rotation and translation convention.

On the 13 held-out evaluation recordings, active speech exceeds its paired inactive control for both the gap and spike in $13/13$ cases (two-sided exact sign $p = 0.000244$ for either statistic; Fig. 8a). The joint guard returns $11/13 = .846$ active recordings (Wilson 95% interval $[.578, .957]$) and $1/13 = .077$ inactive controls ($[.014, .333]$). Conditional on return, median absolute azimuth error against OptiTrack is $2.1°$; $10/13$ recordings both return and lie within $10°$, and $11/13$ both return and lie within $15°$. A development-median direction predictor has $39.2°$ median error on the same held-out angles. The two orange crosses in Fig. 8b show the unreported contrast-mode peaks for the recordings that fail the gap guard; CAP itself returns no direction there. On these identical records and unchanged cutoffs, gap-only returns $11/13$ active and $1/13$ inactive segments, spike-only returns $13/13$ and $3/13$, and CAP returns $11/13$ and $1/13$. The gap guard is therefore binding on LOCATA: relative to spike-only it rejects two accurate active peaks with insufficient background-normalised separation and two inactive returns, while adding the spike guard to gap-only changes

no decision. LOCATA supports physical accuracy conditional on joint return but does not independently establish spike-guard necessity; the paired INSTANCE ablation supplies that complementary evidence. As a post-hoc robustness check rather than part of the frozen protocol, multiplying both cutoffs by 0.75, 1.00, and 1.25 yields active/inactive return counts $12/3$, $11/1$, and $10/1$, with conditional median errors $1.8°$, $2.1°$, and $2.2°$. In a development-sensitivity analysis, each leave-one-recording-out fold refreezes both guards on the other two recordings: all three omitted sources return, no omitted inactive control returns, and evaluation counts range from 11–13 active and 1–2 inactive. The external stability baseline uses 128 independent four-frame block resamples and returns $13/13$ active but $7/13$ inactive recordings. Under the same development-maximum rule, Roy returns $7/13$ active and $1/13$ inactive, whitened Frobenius returns $6/13$ and $0/13$, and Gaussian log-det returns none; CAP has the highest active return count at Roy's matched inactive count, $11/13$ versus $7/13$ (Table 3).

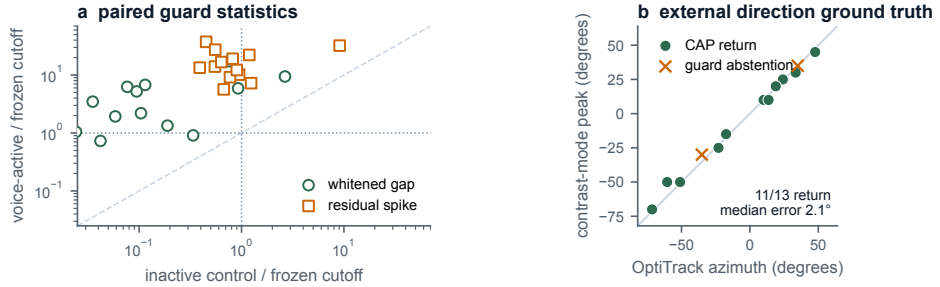

Figure 8: **Held-out real acoustic direction recovery on LOCATA. (a)** Every voice-active recording exceeds its paired inactive control on both normalised guard statistics; dotted lines mark the frozen cutoffs and the diagonal marks equality. **(b)** Green points are returned CAP azimuths against synchronized OptiTrack truth; orange crosses are diagnostic contrast-mode peaks that CAP does not report because a guard fails.

Table 7: **Record-level LOCATA evaluation outcomes.** Directions and errors are reported only when both development-frozen guards pass; the inactive column applies the identical guards to the paired second inactive segment.

Record	OptiTrack azimuth	CAP azimuth	Error	Active	Inactive
1	$10.0°$	$10°$	$0.0°$	pass	abstain
2	$34.9°$	–	–	abstain	abstain
3	$-60.5°$	$-50°$	$10.5°$	pass	abstain
4	$18.9°$	$20°$	$1.1°$	pass	abstain
5	$-35.1°$	–	–	abstain	abstain
6	$-17.4°$	$-15°$	$2.4°$	pass	abstain
7	$33.4°$	$30°$	$3.4°$	pass	abstain
8	$-51.0°$	$-50°$	$1.0°$	pass	pass
9	$13.6°$	$10°$	$3.6°$	pass	abstain
10	$-22.9°$	$-25°$	$2.1°$	pass	abstain
11	$47.7°$	$45°$	$2.7°$	pass	abstain
12	$-71.5°$	$-70°$	$1.5°$	pass	abstain
13	$24.1°$	$25°$	$0.9°$	pass	abstain

These fixed-threshold LOCATA results do not constitute a theorem-mode certificate or a general acoustic-localization benchmark. The recording is the scientific unit; all 13 recordings share one room and array geometry, and frames are repeated observations within a unit. The repository provides the data-processing details, complete analysis specification, and record-level results.

NOAA/DCLDE low-SNR regime. The fixed panel contains the 15-minute file beginning at every whole hour from 28 March through 3 April 2009 (168 files), a rule independent of call counts; expert annotations distinguish definite from possible North Atlantic right-whale upcalls (DCLDE 2013, 2023). Ten log-power features cover 45–500 Hz in 2,048-sample Hann windows with 50% overlap. Positive frames overlap definite calls plus one second; quiet excludes every definite or possible interval plus five seconds. Each eligible file contributes 48 event, 48 reference-quiet, and 48 disjoint evaluation-quiet frames selected by a fixed outcome-blind ordering of frame identifiers. We freeze the marginal 0.95 guards on 34 quiet–quiet files from the first three dates; the final four dates contain 37 eligible files and are not refit. Held-out quiet passes $1/37$ (Wilson $[.005, .138]$) and anno-

tated calls pass $4/37$ ($[.043, .247]$). Weak labelled signals therefore often fail CAP's covariance/gap criterion; this result concerns residual geometry, not annotation validity.

Etna cross-array analysis. We use 13 public VOSSIA windows from the ENCR/ENEA arrays, filtered 0.5–10 Hz and divided into disjoint 30-second blocks (De Angelis et al., 2021). We freeze the marginal 0.95 guards on 30 ENCR quiet–quiet block pairs and transport them to 30 ENEA quiet pairs and eight quiet/active recording pairs. Development ENCR and held-out ENEA quiet pass $2/30$ and $1/30$ blocks. The July 18 Strombolian and July 27 major-explosion windows pass 0.80 and 0.90 of blocks; their within-window median direction cosines are 0.924 and 0.597. The remaining six post-event or residual-activity windows pass 0–0.40 of blocks, so only two of eight clear a majority criterion. Because blocks within a recording are dependent and activity labels are window-level, the recording window, rather than each block, is the scientific unit.

Table 8: **Real-data outcomes in low-signal regimes.** Rates are file-level for NOAA and block-level within recording windows for Etna; Etna blocks are dependent and are not treated as binomial trials.

Domain / regime	Evaluation role	Units	Pass	Rate
NOAA development quiet	freeze guards	34 files	1	.029
NOAA held-out-date quiet	negative transport	37 files	1	.027
NOAA definite calls	low-SNR positive	37 files	4	.108
Etna ENCR quiet	freeze guards	30 blocks	2	.067
Etna ENEA quiet	cross-array transport	30 blocks	1	.033
Etna Strombolian Jul. 18	strong activity	20 blocks	16	.800
Etna major explosion Jul. 27	strong activity	10 blocks	9	.900
Etna six weaker windows	low-activity regime	120 blocks	15	.125

Held-out TROPOMI point-source replication. We evaluate geographic generalization using 24 daily TROPOMI tropospheric-NO_2 crops (Copernicus Sentinel-5P (processed by ESA), 2018) around power-plant clusters from the public Beirle et al. point-source catalog (Beirle et al., 2023). The scientific unit is a geographic cluster, not a pixel patch or day. A prespecified outcome-blind split of the region identifiers assigns 12 regions to development and 12 to test. Within each region, a 9×9 patch around the highest-emission catalog plant is the source panel. We select two equally sized off-source patches only by distance from all catalog plants, using one for calibration and the other as a matched negative control. Using development regions only, we finalise the patch size and shrinkage, normalise from the calibration patch, and learn a four-dimensional spatial subspace from development calibration patches. We freeze $(\tau_{\mathrm{gap}}, \tau_{\mathrm{spike}}) = (8.788, 115.899)$ at the higher empirical 0.90 quantiles of the 12 development controls; we refit neither the representation nor either guard on test regions.

On the 12 held-out regions, plant-centred spike and gap exceed their paired off-source values in $11/12$ regions each (under a region-level paired sign model, two-sided exact $p = 0.00635$ for either metric; Fig. 9). Their medians are 508.1 versus 17.8 for the residual spike and 54.3 versus 2.0 for the whitened gap. The joint guard passes $9/12$ plant-centred panels (Wilson 95% interval $[.468, .911]$) and $4/12$ controls ($[.138, .609]$). The paired discordances are six source-only versus one control-only, so the binary difference is suggestive rather than conclusive (two-sided exact McNemar $p = .125$; Table 9). Thus the replication provides held-out evidence that CAP's two continuous guard statistics rank source panels above matched controls. The four transported control passes reveal substantial background heterogeneity, so these fixed thresholds require a larger and more representative development panel before they can support a general selectivity claim. This analysis uses fixed thresholds and does not invoke the Gaussian theorem-mode certificate. The exact p-values treat the 12 regions as exchangeable units; off-source patches need not be emission-free; days within a region are repeated observations rather than replication units; and clear-sky selection can induce dependence. The repository provides the site-level outcomes, analysis specification, fixed thresholds, and data provenance.

Additional seismic case studies. Event 1 is a magnitude-4.13 local/regional event recorded at CI.PASC, with paired 120-second, three-component quiet and event traces. Event 2 is a Tohoku 2011 surface-wave example recorded at IU.ANMO, with a 48-second quiet trace and a 36-second event trace. Because source provenance and extraction records are incomplete, these records serve as illustrative labelled examples rather than a reproducible benchmark. For each pair we quiet-standardise, band-pass 1–20 Hz at the declared 100 Hz sampling rate, remove the leading quiet

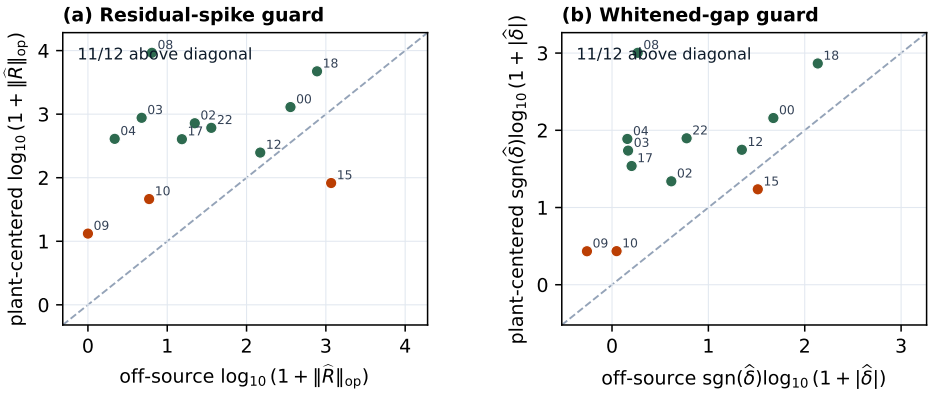

Figure 9: **Held-out TROPOMI point-source replication.** Each labelled point is one held-out geographic region; axes use signed $\log_{10}(1+|x|)$ where needed, and the dashed line is equality. Green points pass both frozen guards for the plant-centred panel; orange points do not. Plant-centred values exceed the matched off-source control in $11/12$ regions for both the residual-spike and whitened-gap guards, without treating days or pixels as independent units.

Table 9: **TROPOMI replication summary.** Continuous rows use a two-sided exact paired sign test; the binary row uses the exact McNemar test on six source-only and one control-only region. Development guard counts are descriptive because those controls froze the thresholds.

Held-out test quantity	Plant-centred	Off-source	Plant > off-source	Exact p	Development controls
Residual-spike median	508.1	17.8	11/12	.00635	12/12
Whitened-gap median	54.3	2.0	11/12	.00635	12/12
Both guards pass	9/12	4/12	6 vs. 1 discordant	.125	9/12 vs. 1/12

principal direction, and lag-embed (10 lags, stride 2). CAP uses all resulting 30 coordinates; the evaluation subspace is therefore not selected from analysis data. Same-record replaces $\hat{C}$ by $\hat{\Sigma}$. Event 1 yields $\cos \approx 0.108$ between the CAP direction and residual EOF1.

Seismic control design. We use 500 random partitions of disjoint 2-second blocks from event 1's quiet record to freeze the 0.95 gap quantile, then apply it to 500 partitions of event 2's different-station quiet record. Event 1 also uses 30 block bootstraps (block 40 embedded rows); its declared guard-pass rate is 0.53 under the frozen guard. We apply the same gap and relative-spike rules to event 2 without refitting, but re-estimate its quiet mean, scale, and ambient direction from its own quiet record. Table 11 varies only the gap-guard calibration quantile over $0.90, 0.95, 0.975, 0.99$ while retaining the same relative spike guard; 0.95 remains the primary analysis. Real-panel covariances subtract each panel's sample mean and divide by its row count; unlike the theorem estimators in Assumption 3.1, these are declared-mode descriptive covariances. Partitions reuse finite waveforms and overlap, so their frequencies describe resampling variability rather than independent trials; the single held-out record measures transport but not a population false-positive rate. We selected this domain and preprocessing during exploration and make no nominal-level or confirmatory p-value claim.

Window length and temporal phase. When we change evaluation from half-record aggregates to mutually exclusive pairs of 2-second quiet blocks, the frozen half-record guard passes $22/29$ event 1 pairs and $6/8$ event 2 pairs (Wilson 95% intervals $[0.58, 0.88]$ and $[0.41, 0.93]$). The cutoff therefore does not transport across window duration. Independently circularly shifting each lag coordinate of the labelled events preserves its marginal series but breaks cross-coordinate timing; all $100/100$ surrogates per event still pass because their second-moment shift remains. CAP is consequently a covariance-shift detector under a fixed calibration unit, not a waveform-coherence test; these perturbations therefore do not measure specificity.

CMEMS SST. North Atlantic panel (T=534, n=1529; lat/lon strides $[8, 4]$): reference = first half, analysis = second half, projected onto the leading 20-dimensional *reference-only* EOF sub-

space. Figs. 11(c) and 10(b)–(c) lift modes to the ambient grid. Uncertainty: n=12 quiet/analysis cuts (quiet fraction in $[0.30, 0.70]$). Shared contamination as in Table 13 / Fig. 12. Because this unlabelled temporal split has no independently validated stationary null, it measures ranking and contamination sensitivity rather than real-domain specificity.

Synthetic. The primary run uses $(n, \beta, \alpha, \gamma) = (6, 2.5, 0.7, 4)$ and independently samples known-zero-mean covariances $\hat{\Sigma}$ and $\hat{C}_\varepsilon$ with $T = T_c = 400$. A disjoint 2,000-replicate null pilot with $(n, \beta, \alpha, \gamma) = (6, 0, 0, 4)$ freezes both guards; 200 replicates per contamination point and five fresh 200-replicate null batches evaluate behaviour. Independent random streams generate the pilot, contamination grid, and five null batches. Because the pilot and evaluation use the same nuisance covariance, these runs do not establish uniform control across spectra, tails, or dimensions. Figure 3 varies these nuisance conditions directly: each cell uses a separate 600-null pilot and 250 fresh null plus 250 fresh alternative replicates. Nuisance-matched null pass rates are 0.004–0.020 across the seven cells. The low-sample cell ($n = 24, T = 72$) has power 0.076 and mean returned-direction cosine 0.729, while the near-tied rank-two cell has power 0.804 and cosine 0.820; the latter lies outside the rank-one guarantee. Reusing the baseline thresholds in the low-sample cell raises the fresh-null pass rate to 0.928, demonstrating that declared cutoffs do not transport uniformly. The secondary dependent path in Table 12 sets $\hat{C}_\varepsilon = (1-\varepsilon)\hat{C} + \varepsilon\hat{\Sigma}$ and therefore isolates the algebraic effect of covariance reuse. Side-info modes $\{\texttt{none}, \texttt{quiet-C}, \texttt{physical-}v_c\}$ use $n_{\mathrm{rep}} = 30$, $T = 800$. The physical-v_c row illustrates an exploratory analysis outside Cor. 5.1. Reported gaps, including Table 13, are the implemented statistic $\hat{\delta} = \lambda_1(\hat{M}) - \max(1, \lambda_2(\hat{M}))$ compared with τ_{gap}. Thresholds are the declared numerical cutoffs of App. B, not theorem-mode radii.

Explicit Gaussian theorem mode. We independently sample exact Wishart covariances under $n = 2$, $C = I$, $\beta = 2$, $t = 3$, and $m_0 = 6.01$, using 200 replicates at every (T, ε) pair. The observable Gaussian radii of Cor. A.2 require no fitted concentration constants. Observed certification boundaries expand with sample size and lie beyond the theorem's conservative sufficient boundary; across five null grids the positive rate is zero, and no accepted replicate violates its reported residual-angle radius (Fig. 4). Table 10 makes that conservatism explicit beyond $n = 2$. Using the population matrices only as plug-ins to the observable Gaussian radii, it reports the smallest equal $T = T_c$ clearing both sufficient inequalities at quiet $C = I$ and $\beta = 2$. The requirement is already 0.74 million samples in two dimensions and rises to 7.85 million at $n = 128$; these are planning diagnostics, not empirical acceptance guarantees.

Table 10: **Theorem-mode sample-size planning diagnostic.** Minimum equal $T = T_c$ clearing both quiet-model sufficient inequalities for $C = I$, $\beta = 2$, $t = 3$, and $m_0 = 6.01$ when population matrices are inserted into the observable radii.

n	minimum $T = T_c$	T/n	$\bar{\eta}_M$ at crossing
2	737,373	368,687	0.24969
8	1,290,976	161,372	0.24969
32	2,860,244	89,383	0.24969
128	7,847,025	61,305	0.24969

Table 11: **Frozen-guard sensitivity.** We refreeze the gap guard at quantile q of the same 500 event 1 calibration quiet partitions while keeping the relative spike guard fixed. The $q = 0.95$ row is the primary analysis; the other rows show sensitivity to this choice.

q	τ_{gap}	cal. quiet	held-out quiet	event 1 boot	event 1 / event 2
0.900	0.262	0.100	0.178	0.60	yes / yes
0.950	0.380	0.050	0.092	0.53	yes / yes
0.975	0.484	0.026	0.050	0.33	no / yes
0.990	0.578	0.010	0.022	0.27	no / yes

Population gap law. Table 14 evaluates Thm. 3.1 at population level: closed-form λ_ε matches the leading GEP of (Σ, C_ε) to machine precision. Table 12 separately reports the empirical guard $\hat{\delta}$, which can fail before the excess $\lambda_1(\hat{M}) - 1$ because $\lambda_2(\hat{M})$ can exceed one.

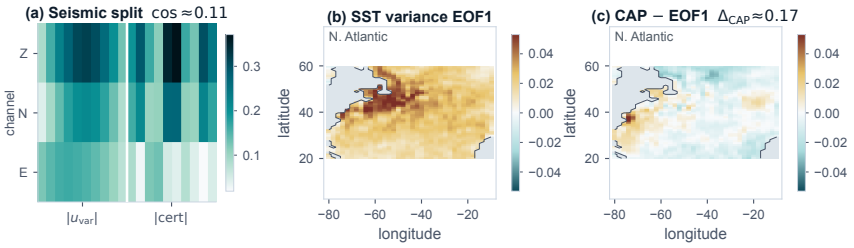

Figure 10: **A reference-adjusted residual direction need not be residual EOF1.** **(a)** Seismic event 1: $|\langle \hat{v}_\star, u_{\mathrm{var}} \rangle| \approx 0.108$. **(b)–(c)** Exploratory SST: $\cos \approx 0.91$, so $\Delta_{\mathrm{CAP}} \approx 0.17$; the distinct EOF1–SNR gap is $\Delta_{\mathrm{SNR}} \approx 0.83$. CAP follows reference C, not EOF1 by default.

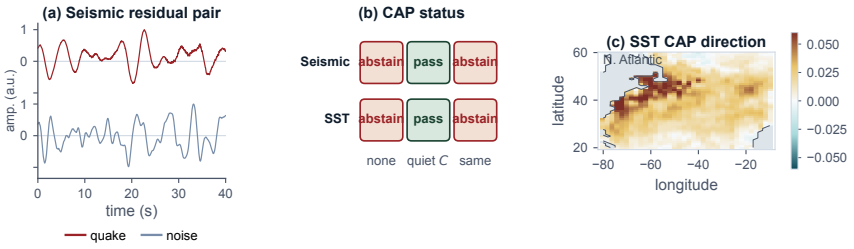

Figure 11: **Additional real-data case studies.** **(a)** Seismic quiet record versus labelled event. **(b)** None / reference C / same-record statuses under each panel's declared guards. **(c)** Exploratory North Atlantic SST half-split residual candidate, lifted from the reference-only subspace.

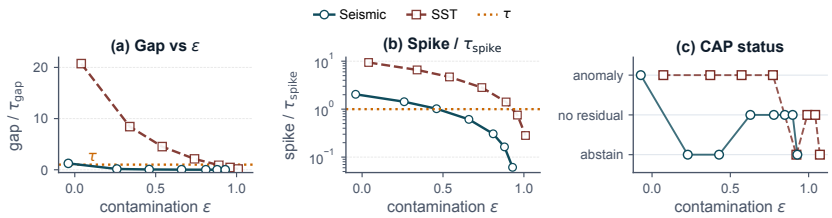

Figure 12: **Sensitivity to dependent interpolation.** **(a)–(b)** Gap and spike margins relative to each domain's guard. **(c)** CAP status as $\hat{C}_\varepsilon = (1-\varepsilon)\hat{C} + \varepsilon\hat{\Sigma}$ reuses the fixed analysis covariance. These paths isolate covariance reuse and do not estimate independent-sample coverage.

Table 12: Dependent synthetic interpolation (n_{rep}=40, T=T_c=400): empirical eigen-gap $\hat{\delta}$, mean±s.d., and empirical rates. Highlighted row = $\hat{\delta} < \tau_{\mathrm{gap}}$ on every replicate; at ε=0.95, abstain 0.85 via `gap_fail` and 0.15 return no detectable residual.

ε	$\hat{\delta}$ (mean±s.d.)	gap-fail	guard pass	abstain	λ_ε
0.00	1.408±0.349	0.00	**1.00**	0.00	2.52
0.30	0.617±0.093	0.00	**1.00**	0.00	1.73
0.60	0.241±0.028	0.00	**1.00**	0.00	1.32
0.85	0.072±0.011	0.75	0.25	0.75	1.10
0.95	0.022±0.003	**0.85**	**0.00**	**0.85**	1.03

Table 13: Dependent interpolation under fixed guards: $\varepsilon^\star$ is the first grid value at which a guard-passing direction is no longer returned (seismic event 1: 0.30; SST: 0.85); at $\varepsilon = 1$ both panels abstain under the same-record condition.

	Seismic event 1		SST	
ε	status	$\hat{\delta}$	status	$\hat{\delta}$
0.00	guard pass †	0.473	guard pass †	1.037
0.50	abstain ‡	0.022	guard pass †	0.226
0.70	no residual	0.008	guard pass †	0.106
0.92	no residual	0.001	no residual	0.022
1.00	abstain	—	abstain	—

† EOF1-misreport flag (seismic ε=0; every displayed guard-passing SST row); ‡ gap-fail (guard passing ends). "No residual" is threshold-relative.

Table 14: Population gap law from Thm. 3.1: closed-form λ_ε matches the leading generalised eigenvalue entrywise ($|\lambda_\varepsilon - \lambda_{\mathrm{GEP}}| < 10^{-12}$ throughout; cosine to the clean SNR direction $= 1$).

ε	λ_ε (theory)	λ_{GEP}
0.00	1.933	1.933
0.25	1.568	1.568
0.50	1.318	1.318
0.75	1.137	1.137
0.95	1.025	1.025

Summary. The appendix establishes when the residual direction is identified, when finite samples preserve that identification, and how the empirical conclusions change across sampling regimes and replication units. Throughout, second-order residual structure supports a direction only with identifying side information and a resolved gap; otherwise CAP abstains.