Research Square

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

Hybrid Complex-Valued Deep Learning and Reference-Anchored ICA for Small-Target Detection in Sea Clutter: Detection and Calibration Analysis

Hamzeh Ghahramani

hamze.ghahramani@iau.ac.ir

Islamic Azad University, Bandar Lengeh Branch, Khamir Centre

Research Article

Keywords:

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11648448/v1

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

Additional Declarations: No competing interests reported.

Hybrid Complex-Valued Deep Learning and Reference-Anchored ICA for Small-Target Detection in Sea Clutter: Detection and Calibration Analysis

Hamzeh Ghahramani[1*]

[1*]Department of Electrical Engineering, Bandar Khamir Center, Bandar Lengeh Branch, Islamic Azad University, Hormozgan, Iran.

Corresponding author(s). E-mail(s): hamze.ghahramani@gmail.com; hamze.ghahramani@iau.ac.ir;

Abstract

We investigate small-target detection in nonstationary maritime sea clutter using a hybrid architecture that combines a complex-valued neural warm start, reference-anchored independent component analysis (ICA), and a separated-domain normalised matched-filter statistic. The original Swerling detection-power curves, low-false-alarm receiver-operating-characteristic curves, and structural ablations are preserved with their original protocols explicitly distinguished from a corrected 16-acquisition IPIX evaluation. The latter enforces target/guard exclusion and independent calibration/evaluation splits. Under a common retrospective false-alarm operating point, the learned warm start improves detection by 0.057 and 0.095 across the two held-out folds, while the full hybrid does not consistently outperform diagonally loaded ANMF or a two-tap notch. The study also quantifies the dependence of an autoregressive reference detector on reference purity: as target contrast in the cross-polarised reference increases, its detection probability falls sharply, whereas the hybrid statistic remains comparatively stable. False-alarm thresholds transferred between acquisitions do not reliably retain their nominal $\mathbf{10^{-3}}$ rate. The combined power, ROC, ablation, convergence-diagnostic, and corrected-calibration evidence thus defines the scope of the hybrid architecture and clarifies when its learned initialization is useful without claiming universal detector superiority or independently certified cross-acquisition constant-false-alarm performance.

Keywords: sea clutter · small-target detection · polarimetric radar · complex independent component analysis · false-alarm calibration · adaptive cancellation

1 Introduction

Detecting a small maritime target in sea clutter is a low-signal-to-clutter-ratio problem in which the clutter is non-Gaussian, correlated in range and Doppler, and non-stationary over seconds [1–3]. Classical detectors address it with a compound-Gaussian model and a threshold calibrated on clutter-only data [4–6]. When a second polarisation channel is available, the cross-polarised return is attractive as a reference: the target is usually weaker there, and the channel sees the same patch of sea surface. Predictors fitted on that channel, and notch filters that estimate clutter nulls from it, are therefore widely used as reference-assisted detector front ends [1, 7].

Making the reference useful also makes it an assumption. A predictor fitted on the cross-polarised channel suppresses whatever is coherent between the two channels, so if the target contributes to both, the predictor learns the target and the residual cancels it. The same assumption appears in space–time adaptive processing, where target contamination of the secondary (training) data is a first-order problem. What has not been established is whether the reported advantage of a specific polarimetric sea-clutter detector is an instance of that phenomenon rather than a property of its clutter suppression, and how much of a reported detector ranking survives when (i) the reference is modelled with the target contrast the data actually exhibit and (ii) all detectors are compared at one realised false-alarm rate.

This paper develops and investigates a hybrid detector: a complex-valued convolutional network that warm-starts an anchored complex independent component analysis (ICA) stage, followed by a normalised matched filter (NMF) statistic in the separated domain, which we refer to as the hybrid detector. We present its architecture and evaluate its detection curves, warm start, and reference-sensitive operating conditions. Using frozen trials, we test target contamination of the reference and report calibration failures with independently measured rates.

1.1 What the corrected evaluation establishes

C1 — Polarimetric evidence. A gain-free cross-polarised target-contrast proxy estimated on 16 IPIX cells has median -1.8 dB and range -12.5 to $+5.4$ dB. No one contrast describes all acquisitions, and the absence of a VH target is an assumption, not a measured fact.

C2 — Reference mechanism. As the reference becomes target-bearing, AR(8) deployment P_{d} drops from approximately 0.82 to 0.05; Hybrid, Notch and ANMF-DL remain within 0.05 of their original values. A same-channel one-sample-delay control reproduces self-cancellation without polarimetry. A temporally decorrelated but target-free reference also fails: the predictor requires a *simultaneous, target-free* reference.

C3 — Fair comparison. At the retrospective common $P_{\mathrm{fa}} = 5 \times 10^{-3}$, Hybrid does not exceed Notch (paired differences -0.089 and -0.041 across folds) or consistently beat ANMF-DL (-0.023 and $+0.070$). Its warm start improves its own pipeline by $+0.057$ and $+0.095$; this is not a general architectural-superiority claim.

C4 — Calibration transfer. Four CAL-only rules fail to control the nominal 10^{-3} across acquisitions (worst pooled rates 4.75–20.5 times nominal, and worst-acquisition rates 16–75 times nominal). Threshold placement can change P_{d} by 0.22. We distinguish deployed thresholds from retrospective matched-rate thresholds throughout.

Section 3 fixes the model and theory; Section 4 describes calibration and inference; Sections 5–7 give results, followed by limitations and conclusions.

2 Background

Low-grazing-angle sea clutter is commonly represented by compound-Gaussian models, with speckle modulated by slowly varying texture [1, 8, 9]. Normalised and adaptive matched-filter methods [4, 5, 10–13] and CFAR methods [14] depend on how their covariance and threshold reference data represent the test cell. Heterogeneity and target contamination may violate these assumptions.

Dual-polarised sea clutter has been studied in [7, 15, 16]. Cross-polarised observations can provide useful reference structure, but uncalibrated receiver gains prevent direct interpretation of the target amplitude ratio; the clutter-normalised ratio of ratios is the identifiable quantity here. Target polarisation may vary with target aspect, motion and sea state.

Complex ICA [17–21] provides a separation stage that can be warm-started by a learned complex network. Blind identifiability with stochastic sources must be distinguished from reference-anchored recovery of a deterministic target (Section 3.4). Finally, transfer of calibration across acquisitions requires stable upper tails, while distribution-free threshold guarantees also depend on exchangeability assumptions [22, 23]. Our experiments test these assumptions; they do not claim a new calibration solution.

3 Signal model, detectors and identifiability

3.1 Data and CPI representation

The data are 16 IPIX acquisitions: eight from the 1993 Dartmouth campaign and eight from the 1998 Grimsby campaign, X-band, vertical transmit, sampled at PRF = 1000 Hz. Each acquisition contributes a co-polarised and a cross-polarised complex range–time cube. A coherent processing interval (CPI) is $N_p = 512$ pulses (0.512 s) with a hop of 256 pulses, a Hann taper, and a 512-point Doppler transform, giving $F = 512$ Doppler bins. The detection channel is VV and the reference channel VH. Only a per-channel temporal mean is removed: there is no inter-channel gain calibration anywhere in the pipeline, which is why the target model below is parameterised by a contrast rather than by an amplitude ratio.

Two properties of the data are established in the companion audit and used throughout. First, the two channels are mutually incoherent in complex amplitude (median coherence 0.077) but share their power envelope (median log-power correlation 0.669) and satisfy reciprocity between HV and VH to within 0.6 dB, so the reference is a genuine cross-polarised observation of the same patch. Second, in the two Grimsby files the HH and VV cubes are identical to 0.1 dB, so the "VV" label is not established there; those files contribute detection trials but no polarimetric inference.

3.2 Coherent two-channel target model

A point target is one coherent tone at a single Doppler frequency, present in both channels with a fixed amplitude ratio and a relative phase:

$$x_{VV}[n] = c_{VV}[n] + \alpha s[n], \tag{1}$$

$$x_{VH}[n] = c_{VH}[n] + \gamma e^{j\varphi} \alpha s[n], \tag{2}$$

with $s[n] = \exp\{j(2\pi f_d n/\mathrm{PRF} + \varphi_0)\}$, $n = 0, \ldots, N_p - 1$. The VV amplitude α is set so that the VV signal-to-clutter ratio is exactly 0 dB, unchanged from the frozen campaign; f_d is drawn uniformly in the clutter-free Doppler band and φ_0 uniformly in $[0, 2\pi)$ per trial, with both redrawn identically at every condition.

The cross-polarised amplitude is written through the gain-free contrast

$$\begin{aligned} \Delta\mathrm{SCR}_{VH} &= 10 \log_{10} \frac{P_{VH}^{\mathrm{tgt}}/P_{VH}^{\mathrm{clut}}}{P_{VV}^{\mathrm{tgt}}/P_{VV}^{\mathrm{clut}}} \\ &= 20 \log_{10} \gamma + \kappa_a, \end{aligned} \tag{3}$$

where $\kappa_a = 10 \log_{10}(P_{VV}^{\mathrm{clut}}/P_{VH}^{\mathrm{clut}})$ is the measured clutter ratio of acquisition a. Inverting (3) gives $\gamma = 10^{(\Delta\mathrm{SCR}_{VH} - \kappa_a)/20}$, so a nominal $\Delta\mathrm{SCR}_{VH}$ is the same physical condition in every acquisition even though κ_a spans 19 dB across the set. The relative phase φ is drawn uniformly per trial, because the measured mean resultant length of the VV–VH phase difference at the target Doppler is 0.343 (range 0.115–0.831): a fixed phase is not supported by the data. The coherent $\varphi = 0$ case is reported separately as a bound, since it is the worst case for a predictor fitted on the reference. A total-energy control rescales $\alpha' = \alpha/\sqrt{1+\gamma^2}$ so that the two-channel target power is held fixed; it is reported in the supplement.

3.3 Detectors

AR(8)+NMF. An order-8 linear predictor is fitted to the reference CPI by the Yule–Walker equations (with a 10^{-6} relative ridge); the residual $r[n] = x_{VV}[n] - \sum_{i=1}^{8} a_i x_{VV}[n-i]$ uses the coefficients estimated on VH. The statistic is the normalised matched filter

$$t = \max_f \frac{|p_f^H \hat{C}^{-1} r|^2}{(p_f^H \hat{C}^{-1} p_f)(r^H \hat{C}^{-1} r)}, \tag{4}$$

where p_f is the Doppler steering vector of bin f and $\hat{C}$ is the separated-domain covariance estimated from clutter-only units of the same acquisition.

Two-tap canceller (notch). The weight $w = [1, -\hat{\rho}]^T/\|w\|$ uses the one-lag autocorrelation $\hat{\rho}$ of the co-polarised CPI, so that $w^H[x_{VV}, x_{VH}]^T$ subtracts a scaled copy of the reference from the detection channel; t is then (4) on the cancelled series. The weight is estimated on the cell under test, so the canceller has no training-data requirement and no reference assumption.

ANMF with diagonal loading (ANMF-DL). The sample covariance matrix is estimated

from four secondary range windows (1024 samples each) with diagonal loading $10^{-3}\,\mathrm{tr}/N_p$; the statistic is (4) with r replaced by the CPI itself. The detector uses the co-polarised channel only, so it is invariant to the reference and serves as a reference-independent control.

Hybrid (M1–M3). A complex-valued network (two convolutional layers, 64 and 128 channels, followed by a 128-unit LSTM) maps the two-channel CPI to a 2×2 complex mixing matrix W_0 (M1). An anchored complex ICA stage (M2) refines the separation with a steering constraint; the separated-domain covariance and the statistic (4) follow (M3). Two ablations isolate the components: *cold start* uses the same network output but initialises ICA randomly instead of from W_0, and *ICA-only* bypasses the network entirely. A second checkpoint trained without the physics-informed regulariser isolates that term. All four variants use identical trials, identical calibration and identical thresholds.

Figure 1 displays the corrected processing chain and the separation of training, calibration and test information.

3.4 Identifiability: what may and may not be claimed

The investigated design rests on three statements. Two of them are correct once scoped, and one is not; the corrected forms below are used throughout.

Proposition 1 (Blind identifiability, random sources)**.** *Let $u = (u_c, u_t)$ be independent with finite, non-zero variances, and suppose at most one of them is circular Gaussian. Then any non-singular B whose outputs are independent satisfies $BA = PD$ with P a permutation and D diagonal, so the mixing matrix is identifiable up to permutation and scaling.*

Proposition 2 (Anchored identifiability, deterministic target)**.** *Let $x = a_T s + c$ with s a deterministic target waveform, a_T a known unit-norm steering vector and c zero-mean clutter with full-rank covariance C_c. For unit-norm w in the target subspace, the anchored objective $J(w) = |w^H a_T|^2 - \xi\, w^H C_c w$ is maximised at $w = \pm a_T$, uniquely for ξ below the residual level. This is a beamforming statement and does not rely on non-Gaussianity.*

Remark 1. A circular-Gaussian target is not separable *by its own statistics*: a point mass or a Gaussian source contributes nothing to a non-Gaussianity contrast. With one Gaussian target and non-Gaussian clutter, the clutter column remains identifiable and, in the 2×2 case, the target column follows from the mixing geometry up to scale and phase. An identifiability collapse therefore requires *both* sources to be Gaussian, which is not the regime of these data; the Swerling-2 degradation reported for the original design is an empirical observation and is not attributed to a theorem here.

Lemma 1 (Scale invariance)**.** *For a transform ϕ independent of the texture τ and a statistic with $T(cx) = T(x)$ for all $c > 0$, the null law of $T(\phi(x))$ does not depend on the law of τ.*

Lemma 2 (Beta anchor)**.** *If in addition ϕ is a matched whitening, the steering vector is fixed and independent of the data, and the statistic is a single-bin NMF, then $T \sim \mathrm{Beta}(1, N_p - 1)$.*

Remark 2. The deployed statistic maximises (4) over $F = 512$ Doppler bins and its transform depends on the observed window through M1 and the anchor, so it satisfies neither the fixed transform of Lemma 1 nor the matched whitening of Lemma 2. Both lemmas are used as analytic anchors and sanity checks only; every threshold in this paper is empirical, computed on clutter-only calibration data, and its realised rate is measured rather than assumed.

4 Experimental protocol

4.1 Partitions and leakage control

The 16 acquisitions are split into two disjoint folds of eight, each fold providing eight calibration and eight test acquisitions. Target range gates are taken from the published annotations where they exist and from a detection statistic otherwise; the split was audited for target-cell leakage between calibration and test before any experiment ran, and no test acquisition contributes a calibration unit. Each fold supplies 8,000 clutter-only calibration CPIs and 4,000 clutter-only test CPIs, and 1,990 (fold 1) or 2,000 (fold 2) injected-target trials at 0 dB. Trials are shared across detectors and across conditions: a contrast is always computed on identical trial identities.

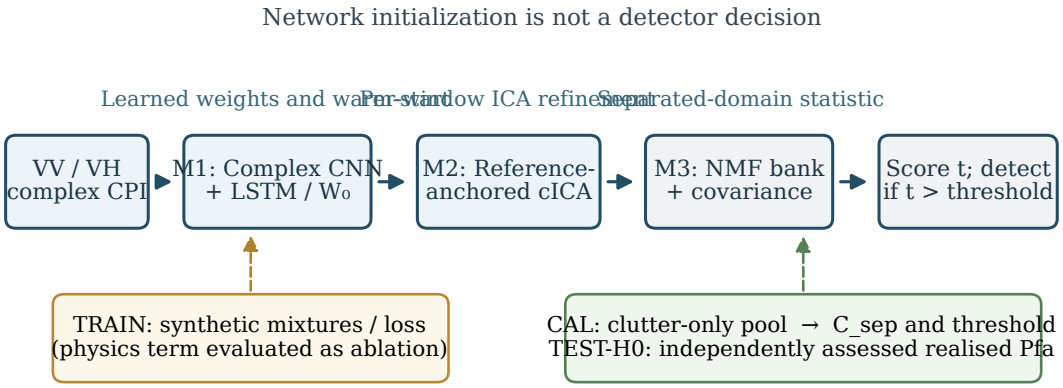

Fig. 1 Corrected architecture and data flow of the investigated hybrid detector. M1 supplies a complex-valued warm start, M2 performs reference-anchored ICA refinement, and M3 computes the separated-domain NMF statistic. The clutter-only CAL pool supplies covariance and threshold estimates; realised false-alarm rate is measured on independent TEST-H0 observations. The training regulariser is not part of inference, and its detection benefit was not confirmed in the corrected evaluation.

4.2 Two operating points, never mixed

Deployment. Each detector's threshold is the $1-10^{-3}$ quantile of its own clutter-only calibration pool, i.e. rank 8 of 8,000. This is the fielded rule, and it carries whatever calibration error the acquisition shift induces; the realised rate is measured on test clutter and is *not* 10^{-3} (Section 7).

Matched. Each detector's threshold is the $1-5\times10^{-3}$ quantile of its own *test* clutter scores, so the realised rate is 5×10^{-3} by construction. This operating point is retrospective, not deployable: it exists to compare separability free of calibration error. A matched 10^{-3} variant is reported only as a sensitivity, because with 4,000 test units the fourth largest score is itself an order statistic with substantial sampling noise.

The two points can disagree, and when they do the disagreement is informative rather than noise. For AR(8), selecting rank 7 instead of rank 8 of the same pool moves the threshold by 0.036 and P_d by 0.05; across ranks 4 to 12 the P_d span is 0.224. Every P_d in this paper is therefore reported with the rule that produced its threshold, and the pooled AR(8) value under the canonical rule, 0.8148 over 3,990 trials, is quoted only with that qualification.

4.3 Inference

Contrasts are paired differences of hit indicators on identical trials. Confidence intervals are 95 % acquisition-clustered bootstrap intervals [24] (20,000 resamples over the eight test acquisitions), so between-acquisition heterogeneity enters the interval directly. p-values are exact two-sided McNemar tests [25], Holm-corrected [26] within each family of 24 contrasts (four conditions × three detector pairs × two folds). An effect is called material only if its interval excludes zero *and* its sign is consistent across both folds; a p-value alone is not treated as evidence of practical relevance. The two instruments answer different questions and can disagree: the McNemar test treats trials as independent, whereas the clustered interval does not, so a contrast can be significant while its interval still covers zero (as happens for one cell of Online Resource 1, Table S4). Where they disagree we rely on the clustered interval and say so.

4.4 Original-protocol detection and warm-start diagnostics

The detector architecture was originally characterised by power-versus-SCR curves, low-false-alarm ROC curves, and structural/delay ablations (Fig. 2 and Fig. 3). These are the unchanged numerical plots from the original experimental protocol and are reproduced here to retain

the complete detector evidence. Crucially, they are *not* estimates from the later corrected 16-acquisition partitions: their original target/guard exclusion and calibration procedures were subsequently found to require correction. Consequently, the plots establish the behaviour of the implemented algorithm under the stated original conditions, not an independently verified across-acquisition 10^{-3} false-alarm guarantee or a corrected between-detector superiority claim. The independent corrected evaluation is reported in Sections 5–7.

Convergence behaviour. In the original fixed-point implementation, 47.5% of warm-start windows met the original stopping test and 52.5% reached the iteration cap due to sign-flip stagnation at a fixed point. The subsequently corrected signed-permutation-invariant stopping rule and stagnation break reported 0% cap hits and a median of one sweep with identical detection statistics. These are the recorded convergence-summary values; no unsupported iteration-history curve is inferred from them.

5 Results I: the reference decides what the statistic can see

5.1 Measured cross-polarised target contrast

Table 1 reports the contrast proxy for the four literature-annotated cells in which the target is visible in VV (peak-to-floor excess ≥ 3 dB). The estimator assumes the strongest Doppler peak at the annotated gate is the target, computes each channel's peak-to-floor excess δ_P, and forms $\eta = (\delta_{VH}-1)/(\delta_{VV}-1)$; the null is the same estimator applied to every eligible clutter cell of the same acquisition.

The interval is 18 dB wide and the between-acquisition spread dominates the within-acquisition precision by an order of magnitude. The data therefore support a *family* of target models, not a point: no single ratio may be called representative, and the frozen protocol's "target absent from VH" is the unattained lower edge of the measured interval. This is a negative result about the measurability of a polarimetric ratio, and we report the mechanism as a function of $\Delta\mathrm{SCR}_{VH}$ over the measured range instead of choosing the value that favours any detector.

Table 1 Measured cross-polarised target contrast, literature-annotated cells (η a power ratio, $\Delta\mathrm{SCR}_{VH} = 10\log_{10}\eta$; bootstrap percentile intervals of η, converted to dB as intervals). The proxy is not labelled ground truth.

Acquisition	δ_{VV} (dB)	η	η 95 % CI	$\Delta\mathrm{SCR}_{VH}$ (dB)
1993-11-07 141630	5.09	3.49	[3.42, 3.63]	+5.4
1993-11-08 213827	7.41	1.42	[0.77, 3.05]	+1.5
1993-11-08 220902	3.61	0.479	[0.40, 0.61]	−3.2
1993-11-07 145028	7.56	0.056	[0.040, 0.081]	−12.5

All 16 cells: median $\eta = 0.662$ (−1.8 dB), range −12.5 to +5.4 dB, 12 of 16 intervals excluding 0 dB.

5.2 The collapse, and three controls

Online Resource 1, Table S2 lists the matched-operating-point detection probabilities; Fig. 4 gives the deployment-threshold comparison across four conditions. AR(8) is monotone in $\Delta\mathrm{SCR}_{VH}$ and loses almost all of its detection ability once the target is in the reference: 0.886 → 0.052 (fold 1) and 0.985 → 0.101 (fold 2) at the matched rate, and 0.821 → 0.043 and 0.808 → 0.056 at the deployment threshold. The hybrid, the notch and ANMF-DL are flat to within 0.05 across the same range. At 0 dB the AR(8) loss is 0.83 and 0.88 at the matched rate (0.77 and 0.75 at the deployment threshold), with acquisition-clustered intervals excluding zero and all eight test acquisitions contributing in the same direction.

The mechanism is identified by the variants in Online Resource 1, Table S10, each with its own calibration built exactly as the frozen arm's. Fitting the predictor on the contaminated cross-channel and fitting it on the detection channel itself give the same result (0.052 versus 0.096 in fold 1; 0.062 versus 0.157 in fold 2): the collapse does not require polarimetry, only a predictor fitted on data that contain the target. Shifting the detection channel by a single sample — no polarimetry anywhere in the construction — reproduces it (0.097 and 0.154). A reference that is target-free but temporally decorrelated fails differently: VH taken 128 CPIs later yields 0.032 and 0.066, and the previous CPI of VH yields

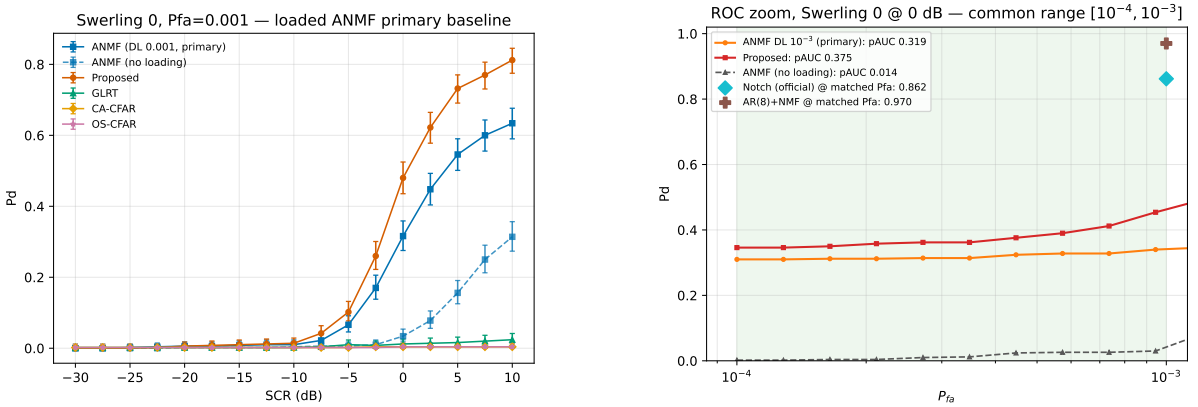

Fig. 2 Original-protocol detector diagnostics, preserved without modifying any ordinate: (a) Swerling-0 power P_D against SCR at design $P_{\text{fa}} = 10^{-3}$; (b) low-PFA ROC at 0 dB. These are the original 500-trial, originally calibrated results and must not be interpreted as independently verified corrected-PFA outcomes. The later 16-acquisition corrected comparisons are discussed separately.

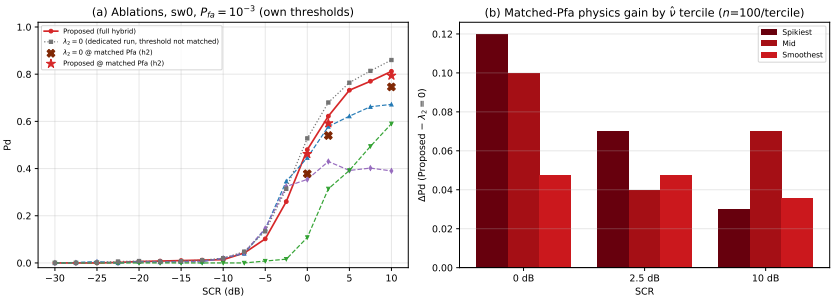

Fig. 3 Original-protocol architectural diagnostics: Full hybrid, ICA ablations and physics-term sensitivity across SCR. The original positive physics-term effect is not a confirmatory result under the corrected calibration and is therefore not claimed as a demonstrated advantage. Warm-start advantages from the corrected evaluation are shown in Fig. 5. Delay-domain warm/cold diagnostics are retained in Online Resource 1 (historical-protocol figures).

0.350 and 0.424. So the requirement is not merely "target-free" but "simultaneous and target-free", which is exactly the pair of properties the cross-polarised channel of the same CPI supplies under an assumption the physics does not guarantee.

6 Results II: what survives at a common operating point

Figure S4 shows the acquisition-clustered paired differences at the common retrospective rate across all four cross-channel conditions; Supplementary Tables S3–S4 give the numerical contrasts. AR(8) is the arm that collapses, so the hybrid's advantage over it is large and unambiguous (+0.768 and +0.650); that advantage is a statement about the reference assumption, not about detection skill. Against detectors that do not rely on the reference, the picture reverses. The two-tap notch is better than the hybrid in both folds (−0.089, −0.041), significantly so in fold 1 after Holm correction, and never worse by a margin that excludes zero. ANMF-DL is ahead in fold 1 by 0.023 with an interval covering zero and behind the hybrid in fold 2 by 0.070; the sign is not consistent across folds, so no superiority claim is made in either direction.

Figure 5 separates the learned components; full values and clustered intervals are retained in the archived Phase-7 ablation JSON. The warm start contributes +0.057 and +0.095 at the matched rate, positive in all eight acquisitions of both folds; at the deployment threshold the same contrast is +0.151 and +0.147, which is where the larger published figure came from. The network as a whole,

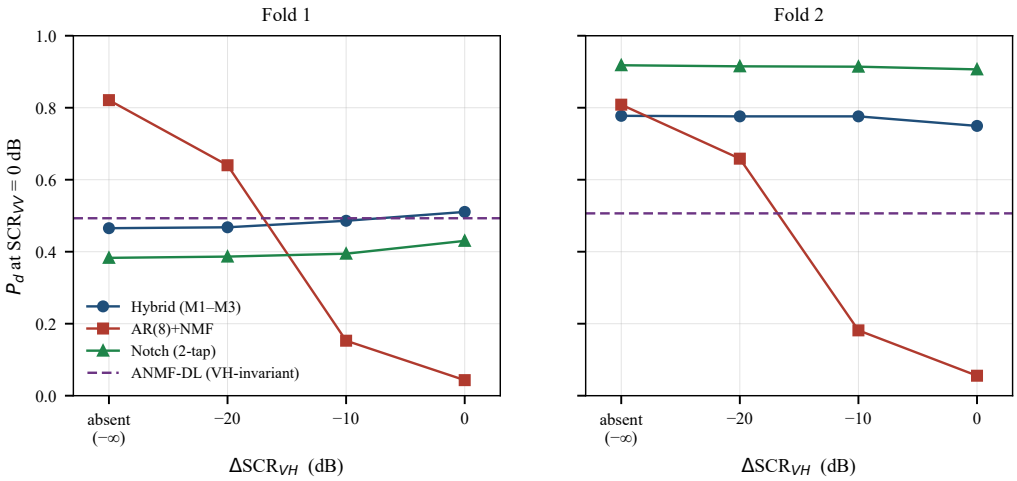

Fig. 4 Detection probability at the deployment threshold ($1 - 10^{-3}$ calibration quantile) against the cross-polarised target contrast, both folds, $SCR_{VV} = 0$ dB. AR(8) is monotone in ΔSCR_{VH}; the hybrid, the two-tap notch and the reference-invariant ANMF-DL are not. The operating point is the deployed one, whose realised false-alarm rate differs from nominal (Section 7); the full matched-rate values are in Online Resource 1, Table S2.

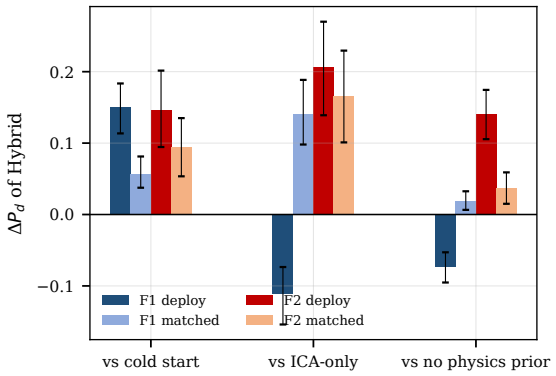

Fig. 5 Ablations of the Hybrid pipeline at the deployment threshold and common retrospective false-alarm rate, by fold. Positive warm-start differences remain in both folds; the effect attributed to the physics-informed loss changes sign at the deployment threshold and is not a confirmed contribution. Acquisition-clustered intervals are shown.

compared with the ICA-only pipeline, contributes +0.140 and +0.166. The physics-informed regulariser is a different matter: it gains +0.019 and +0.036 at the matched rate but loses −0.074 in fold 1 and gains +0.140 in fold 2 at the deployment threshold. The sign flips with the operating point, so the effect is classified as inconclusive and is not used as a contribution.

7 Results III: the threshold decides what the ranking can promise

Online Resource 1, Table S5 and Fig. 6 report four clutter-only calibration rules applied identically to four detectors: (A) the pooled quantile, which is the deployment rule; (B) the median of the eight per-acquisition quantiles; (C) a pooled quantile with a generalised-Pareto tail fit; and (D) a hierarchical rule that standardises each acquisition by its median and interquartile range, takes the global quantile of the standardised scores and maps it back per acquisition. Rules B–D are the natural remedies for acquisition shift; B is the one that a practitioner would try first.

Three features of the calibration audit are central. First, no rule realises the nominal rate; the best (D) is still 4.75 times nominal in the pooled worst case and 16 times in the worst acquisition, and D achieves this only by adapting the threshold to each acquisition's own calibration scores. Second, the stratified rule is the worst, which rules out the obvious remedy: stratifying by acquisition when each stratum contributes a few dozen effective tail samples increases variance faster than it removes bias. Third, the tail is not shared: the eight threshold-setting scores come from a median of two acquisitions, and in the worst case seven of the eight come from a single file, a 1998 'antstep'

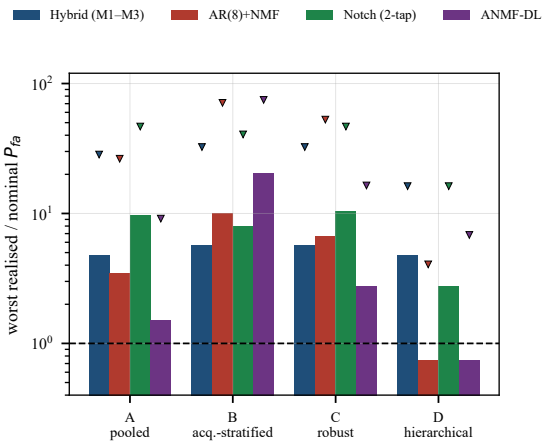

Fig. 6 Realised-to-nominal false-alarm ratio at the design rate 10^{-3}. Bars: worst pooled ratio per detector over both folds; markers: worst single acquisition. The dashed line is exact calibration. No rule reaches it, and the stratified rule — the intuitive remedy — is the worst of the four.

recording whose upper tail is far heavier than the rest.

The consequence for the detector comparison is direct. For AR(8) against the hybrid at VH-absent, the deployment contrast in fold 1 is −0.356 and the matched contrast is −0.019; for the hybrid against the ICA-only pipeline, the deployment contrast is −0.112 and the matched contrast is +0.140. In both cases the deployment number is a statement about two different realised false-alarm rates, not about the detectors. A pooled hybrid-minus-ANMF figure of +0.123 is likewise the average of a −0.028 loss and a +0.271 win obtained at rates of 0.00 and 4.75 times nominal, and we do not report it as a single result.
Additional Swerling-1/2 power curves, the delay and acquisition sensitivity maps, Gaussian controls, replication curves, and measured-target plots are provided with explicit historical-protocol labels in Online Resource 1 (Supplementary Figs. S5–S14).

8 Discussion

8.1 Reference validity, not detector superiority

The results support one organising idea: a reference-conditioned detector is only as good as the reference's validity, and validity has two components that are easy to conflate. The reference must be *simultaneous* — drawn from the same CPI, so that the clutter it predicts is the clutter being cancelled — and it must be *target-free*. The cross-polarised channel is the only channel that can supply both, and only under an assumption that the measured contrasts in Table 1 do not support: at a median of −1.8 dB the target is present in the reference at a contrast comparable to its contrast in the detection channel, and at the upper end of the measured range it is stronger there.

The practical implication is a test, not a redesign. Inject the target into the reference at the contrast the data exhibit, and add a delay-only control. If the detector's advantage disappears in both, the advantage was the reference assumption. The test costs one re-scoring pass over existing trials and needs no retraining.

8.2 What the learned component contributes

The warm start is real: +0.057 and +0.095, positive in all sixteen acquisition-fold cells, with intervals excluding zero. It is also smaller than the earlier reported +0.149, which is a deployment-threshold value inflated by a realised rate 4.75 times nominal in fold 2. And it does not make the learned detector better than a two-tap notch. The honest summary is that the learned initialisation improves *its own* pipeline by a reproducible margin, while the pipeline as a whole does not exceed a reference-free classical detector at a common operating point. The physics-informed regulariser does not survive the operating-point check at all.

8.3 Calibration as the binding constraint

At 10^{-3} with 8,000 calibration units and 16 acquisitions, the threshold is an order statistic whose rank uncertainty is worth up to 0.22 in P_{d} — larger than every architectural effect measured here. This is not a defect of any particular detector; it is the precision limit of the experiment. Two consequences follow for reporting. A nominal rate should always be accompanied by the realised rate per acquisition, because the pooled rate can look acceptable while one acquisition runs at 46–75 times nominal. And a ranking obtained at a deployment threshold should be accompanied by

the matched-rate ranking, because the two can differ in sign.

We do not claim that no calibration method can work. We claim that with 16 acquisitions and this tail behaviour, none of the four rules examined certifies 10^{-3}, that 10^{-4} is not resolvable at all with 4,000 test units (0.4 expected exceedances), and that acquisition-level uncertainty must therefore be reported rather than assumed away.

8.4 Limitations

The polarimetric contrast is a proxy: it assumes the strongest Doppler peak at the annotated gate is the target, it cannot separate target from clutter within one CPI, and it uses a clutter cell as its null. It bounds a family; it does not measure a target's polarimetric radar cross-section. There is no inter-channel gain calibration, so the amplitude ratio itself is not identifiable and is never reported. Two of the 16 acquisitions have an ambiguous polarisation label. The study covers one dataset, one target model (a deterministic tone with uniform relative phase), and two folds of eight acquisitions; the 16-acquisition scope is the reason the calibration conclusion is a characterisation rather than a guarantee. The total-energy control and the coherent-phase bound are reported in the supplement and do not change the ordering, but they are sensitivity analyses, not independent experiments.

9 Conclusion

A cross-polarised reference makes an adaptive detector work by cancelling what the two channels share — including the target, once the target is in the reference. On 16 IPIX acquisitions the measured cross-polarised target contrast is far from zero (median -1.8 dB, 12 of 16 intervals excluding zero), and under that contrast an autoregressive reference-anchored detector falls from 0.89 to 0.05 at a common rate (0.82 to 0.04 at its deployed threshold) while three other detectors are unchanged. A one-sample-delay control with no polarimetry reproduces the collapse, and a target-free but decorrelated reference fails too, so the requirement is a simultaneous *and* target-free reference. At a common operating point the learned detector's advantage over the classical reference-conditioned baseline is an artefact of the reference assumption; against a two-tap notch it has none; its warm start is worth 0.06–0.10. No clutter-only calibration rule realises 10^{-3} across acquisitions, and adjacent threshold order statistics are worth up to 0.22 in detection probability. Reported rankings of these detectors at low false-alarm rates are therefore dominated by threshold placement, and should be reported at both the deployment and a matched operating point, with the retrospective one labelled.

Declarations

Author contributions. Hamzeh Ghahramani is the sole author and is accountable for all aspects of the work. He designed the study, reviewed and approved the research code, specified and supervised the experiments, verified the numerical outputs, performed the scientific interpretation, and wrote and approved the final manuscript.

Use of AI tools. GLM 5.3 Flash assisted with experiment design, code, computational experiments and manuscript drafting; DeepSeek V4.1 Flash assisted with code, experimental-output and numerical-verification audits; ChatGPT (OpenAI) assisted with research planning, revision organisation, scientific discussion and final manuscript review. The sole human author carefully reviewed and verified all AI-assisted code, outputs and text, assumes full responsibility for the work, and does not regard AI assistance as independent validation.

Funding. This research received no external funding or financial support.

Competing interests. The author declares no competing interests.

Data availability. Third-party IPIX recordings are not redistributed. Trained model checkpoints and compact results (not the full 175 MB score archive) are at doi:10.5281/zenodo.24005920 (CC BY 4.0); original and corrected protocols are separately documented.

Code availability. Author-developed code: GitHub (MIT); archived version v1.0.0: doi:10.5281/zenodo.24005344.

Ethics approval. Not applicable (previously recorded radar data; no human or animal participants).

References

[1] Ward KD, Tough RJA, Watts S (2013) Sea Clutter: Scattering, the K-Distribution and Radar Performance, 2nd edn. IET, London

[2] Greco M, Bordoni F, Gini F (2004) X-band sea-clutter nonstationarity: influence of long waves. IEEE Journal of Oceanic Engineering 29:269–283

[3] Rosenberg L, Watts S (2013) High grazing angle sea clutter. In: Radar Conference (RadarCon), IEEE, pp 1–6

[4] Conte E, Lops M, Ricci G (1995) Asymptotically optimum radar detection in compound-Gaussian clutter. IEEE Transactions on Aerospace and Electronic Systems 31:617–625

[5] Gini F (1997) Sub-optimum coherent radar detection in a mixture of K-distributed and Gaussian clutter. IEE Proceedings – Radar, Sonar and Navigation 144:39–48

[6] Conte E, De Maio A, Ricci G (2002) CFAR detection of distributed targets in non-Gaussian disturbance. IEEE Transactions on Aerospace and Electronic Systems 38:612–621

[7] Farina A, Gini F, Greco MV, Verrazzani L (1997) High resolution sea clutter data: statistical analysis of recorded live data. IEE Proceedings – Radar, Sonar and Navigation 144:121–130

[8] Ward KD (1981) Compound representation of high resolution sea clutter. Electronics Letters 17:561–563

[9] Jakeman E, Pusey PN (1976) A model for non-Rayleigh sea echo. IEEE Transactions on Antennas and Propagation 24:806–814

[10] Kelly EJ (1986) An adaptive detection algorithm. IEEE Transactions on Aerospace and Electronic Systems 22:115–127

[11] Robey FC, Fuhrmann DR, Kelly EJ, Nitzberg R (1992) A CFAR adaptive matched filter detector. IEEE Transactions on Aerospace and Electronic Systems 28:208–216

[12] Kay SM (1998) Fundamentals of Statistical Signal Processing, Vol. II: Detection Theory. Prentice Hall, Upper Saddle River

[13] Van Trees HL (2001) Detection, Estimation, and Modulation Theory, Part I. Wiley, New York

[14] Rohling H (1983) Radar CFAR thresholding in clutter and multiple target situations. IEEE Transactions on Aerospace and Electronic Systems 19:608–621

[15] Greco M, Gini F, Rangaswamy M (2006) Statistical analysis of measured polarimetric clutter data at different range resolutions. IEE Proceedings – Radar, Sonar and Navigation 153:473–481

[16] Carretero-Moya J, Gismero-Menoyo J, Blanco-del-Campo A, Asensio-López A (2010) Statistical analysis of a high-resolution sea-clutter database. IEEE Transactions on Geoscience and Remote Sensing 48:2024–2037

[17] Comon P (1994) Independent component analysis, a new concept? Signal Processing 36:287–314

[18] Cardoso JF (1998) Blind signal separation: statistical principles. Proceedings of the IEEE 86:2009–2025

[19] Adali T, Li H, Novey M, Cardoso JF (2008) Complex ICA using nonlinear functions. IEEE Transactions on Signal Processing 56:4536–4544

[20] Hyvärinen A (1999) Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks 10:626–634

[21] Novey M, Adali T (2008) On extending the complex FastICA algorithm to noncircular sources. IEEE Transactions on Signal Processing 56:2148–2154

[22] Vovk V, Gammerman A, Shafer G (2005) Algorithmic Learning in a Random World. Springer, New York

[23] Quiñonero-Candela J, Sugiyama M, Schwaighofer A, Lawrence ND (eds) (2009) Dataset Shift in Machine Learning. MIT Press, Cambridge

[24] Efron B, Tibshirani RJ (1993) An Introduction to the Bootstrap. Chapman & Hall, New York

[25] McNemar Q (1947) Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 12:153–157

[26] Holm S (1979) A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6:65–70

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SIVPSUPPLEMENTARYLATEXSOURCE.zip