Latent-Space Metrics for Complex-Valued VAE for Radar OOD Detection under Clutter

Yadang Alexis Rouzoumka[1,2] Eugénie Terreaux[1] Christèle Morisseau[1] Jean-Philippe Ovarlez[1,2] Chengfang Ren[2]

[1]DEMR, ONERA, Universite Paris-Saclay, France [2]SONDRA, CentraleSupelec, Universite Paris-Saclay, France

Context and Motivations

Classical adaptive radar detectors such as Kelly, AMF, and ANMF rely on Gaussian or SIRV clutter assumptions. Their performance can drop when the background combines non-Gaussian clutter and additive thermal noise.

Which score is the most reliable for radar OOD detection: reconstruction, latent geometry, or a classical adaptive detector?

- Train a **complex-valued VAE** on clutter-only Doppler profiles.
- Compare **CVAE-MSE**, **Mahalanobis**, **empirical KLD**, and **ANMF-FP**.
- Evaluate at fixed $P_{fa} = 10^{-2}$ on synthetic and real radar data.

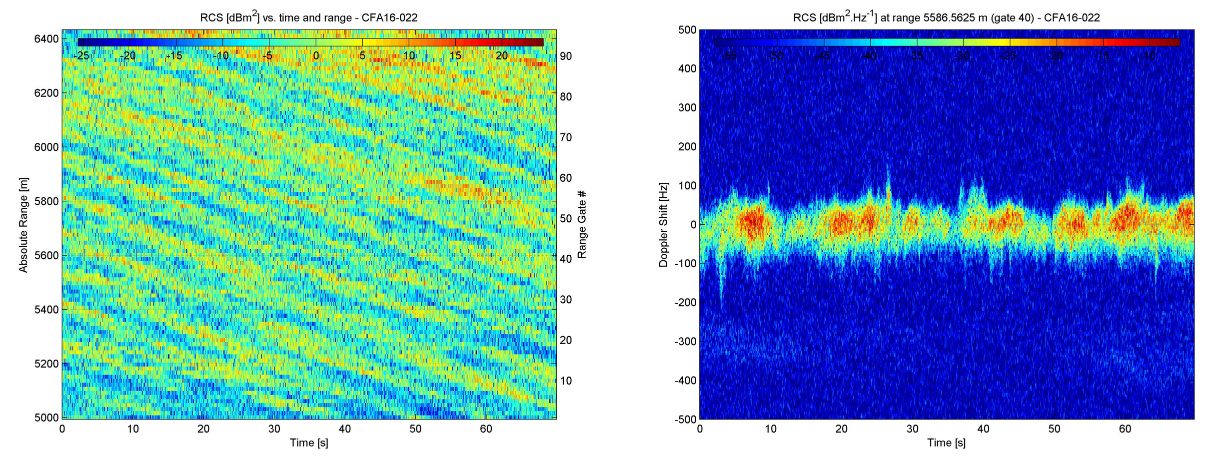
Main contributions

1. Benchmark **complex latent-space scores** against a strong classical detector.
2. Compare detectors on **cGN + AWGN**, **cCGN + AWGN**, and real **CSIR** sea clutter.
3. Separate the **mean over cells** from the difficult **Doppler-0** cell.
4. Derive **practical guidance** on when to favor reconstruction, latent geometry, or adaptive matched filtering.

Data and protocol

- **Target model:** $\alpha = \sqrt{\frac{\text{SNR}}{m}}\, e^{2i\pi\phi}$ with $m = 16$ Doppler bins.
- **Clutter covariance:** Toeplitz of vector $[1\ \rho\ \cdots \rho^{m-1}]$ with correlation $\rho = 0.5$.
- **Texture / secondaries:** Gamma distribution with unitary shape and scale, $K = 2m = 32$ secondary samples.
- **Training:** 15,000 clutter samples, 50 epochs, Adam optimizer, learning rate 10^{-3}.
- **Latent size:** 12 for CVAE-MSE, 32 for Mahalanobis and KLD.

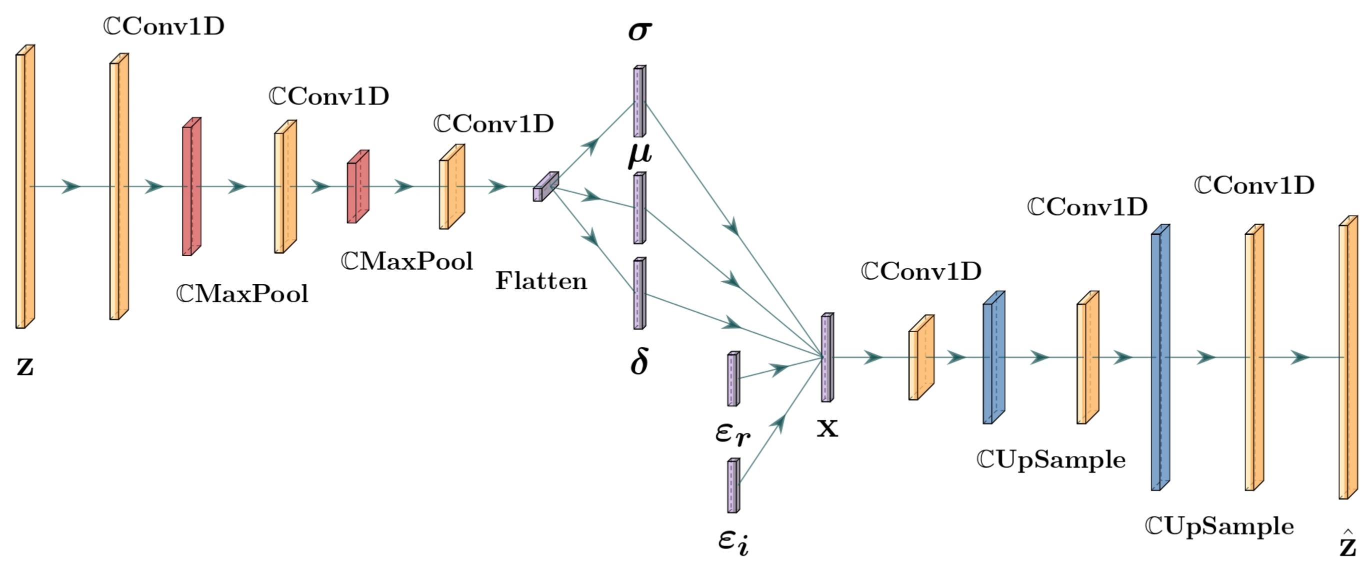
Real CSIR example



Range - slow time map Doppler - pulse spectrogram

These real sea-clutter measurements highlight the structured background that stresses both generative and classical detectors.

Complex-valued VAE



The encoder outputs a complex mean $\boldsymbol{\mu} \in \mathbb{C}^q$, a variance $\boldsymbol{\sigma}^{\odot 2} \in \mathbb{R}_+^q$, and a pseudo-variance $\boldsymbol{\delta} \in \mathbb{C}^q$, enabling phase-aware latent modeling.

Compared scores

Reconstruction score

$$S_{\text{MSE}}(\mathbf{x}) = \|\mathbf{x} - \hat{\mathbf{x}}\|^2$$

Detects samples that the clutter-trained CVAE fails to reconstruct.

Hermitian Mahalanobis

$$S_{\text{Maha}}(\mathbf{z}) = (\mathbf{z} - \hat{\boldsymbol{\mu}}_{\text{ref}})^H \hat{\boldsymbol{\Sigma}}_{\text{ref}}^{-1} (\mathbf{z} - \hat{\boldsymbol{\mu}}_{\text{ref}})$$

Uses second-order complex latent geometry estimated from clutter embeddings.

Empirical complex KLD

$$S_{\text{KL}}(\mathbf{x}) = D_{\text{KL}}\Big(q_\phi(\mathbf{z}|\mathbf{x}) \,\|\, \mathcal{CN}(\hat{\mu}_0, \hat{\boldsymbol{\Sigma}}_0, \hat{\boldsymbol{\Delta}}_0)\Big)$$

Measures how much the test posterior deviates from an empirical clutter law in augmented complex space.

Classical baseline

ANMF-FP with Tyler covariance estimate

Decision pipeline

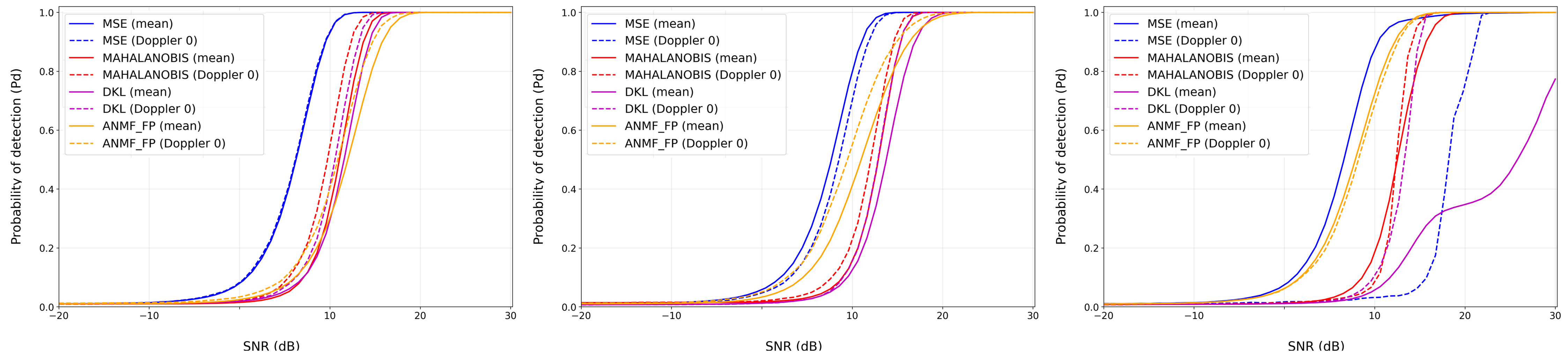
$\mathbf{x}_{\text{test}}$ → reconstruct → compute score → threshold at P_{fa}

Thresholds are calibrated on an independent clutter validation set to enforce the same false-alarm rate for all detectors.

Interpretation guide

CVAE-MSE	Best when the training clutter closely matches the test environment.
Mahalanobis	Most informative latent score; captures covariance geometry and mitigates reconstruction failures.
KLD	Sensitive but brittle in practice for this radar setting.
ANMF-FP	Strong baseline on structured real clutter, especially for difficult cells.

Main result: probability of detection versus SNR



cGN + AWGN cCGN + AWGN CSIR data

Solid lines show the mean over cells excluding Doppler 0. Dashed lines correspond to the Doppler-0 cell.

What the curves show

Detector	cGN+AWGN	cCGN+AWGN	CSIR mean	Doppler 0
CVAE-MSE	best	best	best	degrades
Mahalanobis	2nd	3rd	3rd	robust
KLD	weak	weak	weak	weak
ANMF-FP	3rd	2nd	2nd	best

- **Synthetic Gaussian:** CVAE-MSE reaches a target P_d at the lowest SNR. Mahalanobis is the best latent alternative.
- **Synthetic heavy-tail:** CVAE-MSE remains first, while ANMF-FP becomes the strongest non-reconstruction option.
- **Real CSIR mean:** CVAE-MSE stays slightly ahead, followed by ANMF-FP and Mahalanobis.
- **Doppler 0:** CVAE-MSE deteriorates sharply; Mahalanobis is more robust; ANMF-FP is the strongest overall.

Practical guidance

1. Use **CVAE-MSE** when the operational clutter is close to the clutter seen during training.
2. Prefer **ANMF-FP** for realistic structured clutter and especially for the difficult **Doppler-0** cell.
3. Add **Mahalanobis** when you need a latent-space score that is more robust than pure reconstruction.
4. Do not rely on **KLD** alone in this setting.

Conclusion

Complex-valued latent scores are useful, but **not all latent criteria are equally effective**. The overall picture is simple:

CVAE-MSE wins on most of Doppler bins, ANMF-FP wins on Doppler-0, and Mahalanobis is the best latent score for Doppler-0 robustness.

Selected References

1. D. E. Tyler, "A Distribution-Free M-Estimator of Multivariate Scatter," *Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.
2. D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," in *Proc. ICLR*, 2014.
3. C. Trabelsi, O. Bilaniuk, Y. Zhang, A. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, "Deep Complex Networks," in *Proc. ICLR*, 2018.
4. K. Lee, K. Lee, H. Lee, and J. Shin, "A Simple Unified Framework for Detecting Out-of-Distribution Samples and Adversarial Attacks," in *Proc. NeurIPS*, 2018.
5. Y. A. Rouzoumka, E. Terreaux, C. Morisseau, J.-P. Ovarlez, and C. Ren, "Complex-Valued Variational Autoencoders for Radar Detection in Joint Compound Gaussian Clutter and Thermal Noise," *Proc. EUSIPCO*, 2025.