



Synthesizing Hyperspectral Data using Generative Models to Train Spectral Unmixing Methods for low-cost Crop Residue Cover Mapping



UNIVERSITY OF
TORONTO

Ege Artan^{1,3}, Shuo Chen^{2,3}, Andrew Peng^{2,3}, Sammuell Aldrich Karya^{2,3}, Kyaw Thiha^{1,3}

¹ University of Toronto Scarborough, Toronto, Canada, ² University of Toronto Faculty of Arts and Science, Toronto, Canada,

³ University of Toronto Aerospace Team Space Systems Division, Toronto, Canada.

Motivation & The Data Gap

- FINCH (UTAT): low-cost CubeSat mapping crop residue cover (CRC) over 900–1700 nm SWIR1 [1].
- SWIR1 lacks index-defining features $\Rightarrow$ complex data-driven unmixers.
- Output: three abundances — green veg. (GV), non-photosynthetic veg. (NPV, i.e. residue), soil.
- Gap: such unmixers need dense manifolds, yet CRC ground-truth is sparse, heterogeneous, and costly [2].
- Idea: decouple interpolation from mapping via a generator, injecting spectral priors U cannot use.

The Decoupling Framework

Two stages: a generator G (abundances $\rightarrow$ spectra) and an unmixer U trained on $k \gg n$ synthetic spectra. Testing U on the synthetic manifold approximates the population risk [6], giving the decoupling gap (prob. $1 - \delta$) [5]:

$$\mathcal{R}(U) \leq \hat{\mathcal{R}}_k(U) + 2\mu \text{Rad}_k(U) + M\sqrt{\frac{\ln(1/\delta)}{2k}} + \epsilon_{\text{gen}} + \epsilon_{\text{app}}(U) \quad (1)$$

As $k \gg n$, the complexity and confidence terms become non-limiting, so the bound is dominated by empirical risk, generator bias, and unmixer approximation error: $\hat{\mathcal{R}}_k(U) + \epsilon_{\text{gen}} + \epsilon_{\text{app}}(U)$. Thus, increasing k helps only when G is low-bias; otherwise U learns the wrong manifold more densely.

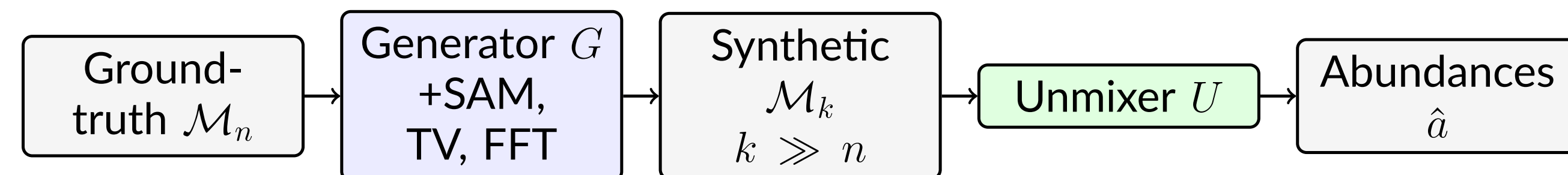


Figure 1. Two-stage decoupled pipeline; priors enter only at G .

Why Decoupling Helps

- Manifold: priors on G shrink the empty space $\mathbb{R}^{210} \setminus \mathcal{M}_n$.
- Lipschitz: a denser manifold lowers U 's constant $K \Rightarrow$ less oscillation, better OOD stability.
- DPI: priors add no information, but make it accessible to U .

Unmixing Models

- MLP: MLP-S ($\sim 16k$), MLP-M ($\sim 61k$) — per-pixel.
- CNN: 1D conv + residual blocks (M/L/XL).
- FNO: global Fourier convolution for long-range bands (M/L/XL) [3].

Accuracy: mean (total) R^2 over GV/NPV/soil.

Generative Models

GD-Streamline — Gaussian Diffusion (DDIM)

1D Conditional U-Net, Conformer layers (from Matcha-TTS [4]); abundance conditioning; classifier-free guidance. Loss = L2 + SAM + FFT-L1 + TV.

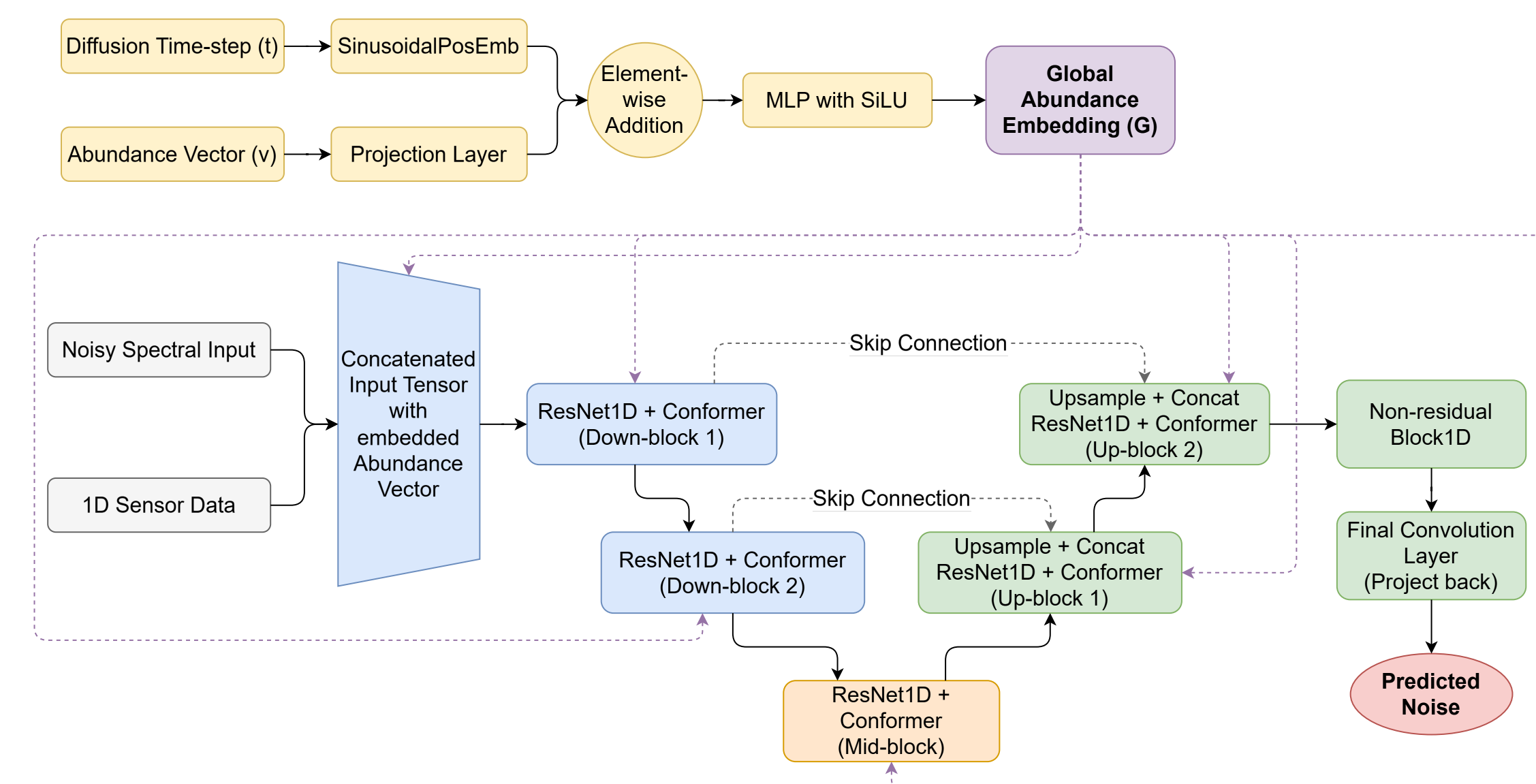


Figure 2. GD-Streamline (1D Conditional U-Net, Conformer layers).

Dual-Path TCVAE — Transformer Conditional VAE

6 encoder transformer blocks $\rightarrow$ Gaussian latent. Dual-path decoder: global path = conditions, local path = latents, fused with learnable weights $\Rightarrow$ removes over-smoothing and mode collapse.

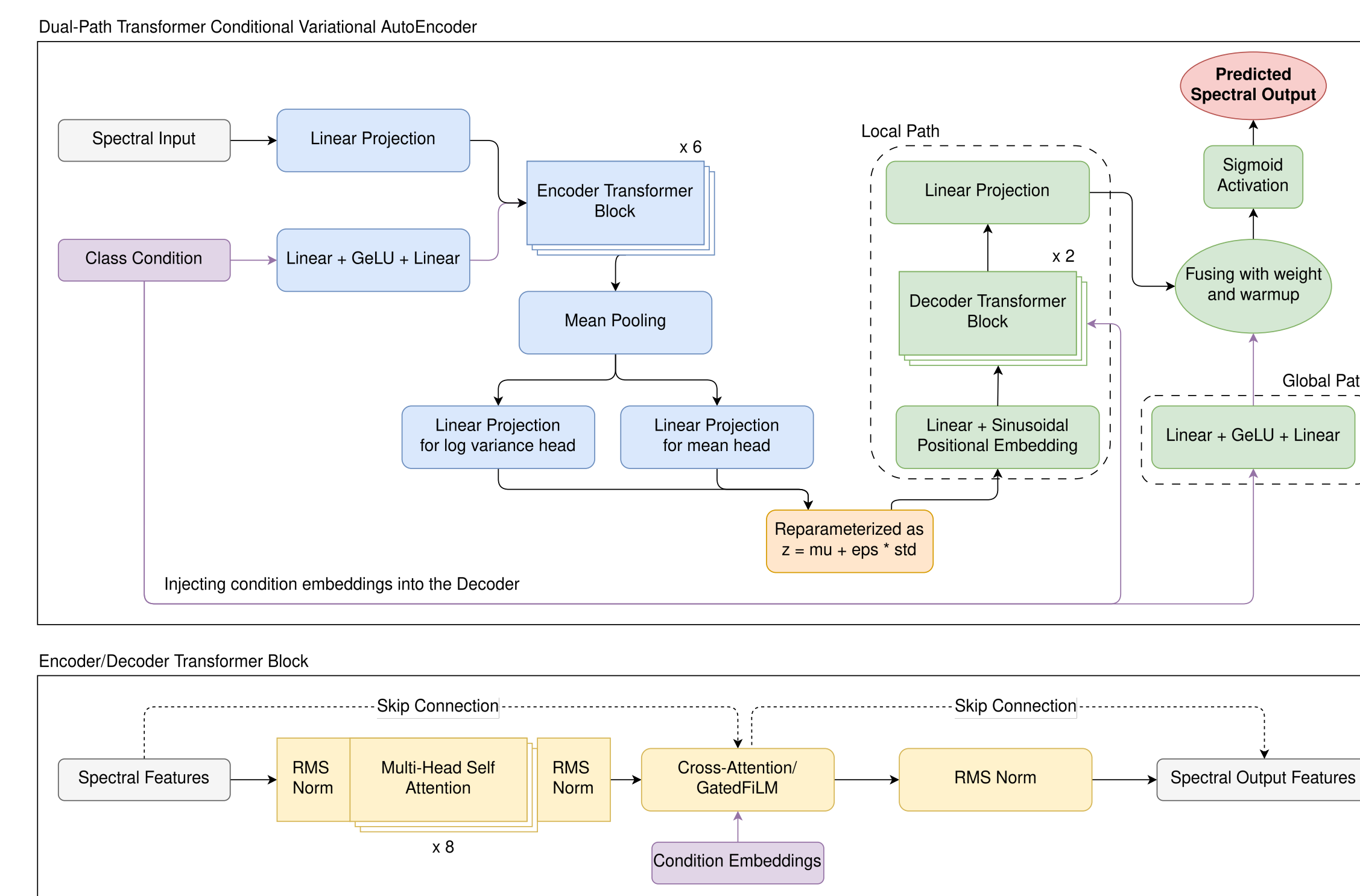


Figure 3. Architecture of the Dual-Path TCVAE.

Evaluation: Manifold Metrics

A 1-nearest-neighbour adjacency (Euclidean and spectral-angle) between generated and real spectra gives quadrant counts γ_{ij} :

$$\delta := \frac{\gamma_{11} + \gamma_{22}}{\gamma_{12} + \gamma_{21}}, \quad \epsilon := \gamma_{11} - \gamma_{22}, \quad \zeta := \gamma_{12} - \gamma_{21} \quad (2)$$

- δ : domain gap (≈ 1 ideal, $\gg 1$ collapse).
- ϵ : clustering (mode collapse); $\zeta > 0$ collapse, $\zeta < 0$ hallucination.

Generator	Distance	δ	ζ	ϵ
GD-Streamline	Euclidean	4.068	−10	−11
GD-Streamline	Spec. Angle	4.068	−6	−7
TCVAE	Euclidean	5.509	−50	−49
TCVAE	Spec. Angle	14.00	−24	−23

Table 1. Nearest-neighbour metrics (400–2490 nm). Large δ = domain collapse; $\zeta < 0$ = hallucination.

Results

Ground-truth baseline (400–2490 nm)

FNO-L/XL best ($R^2 \approx 0.91$); MLP-M 0.89; CNN-M 0.86. $\Rightarrow$ probe decoupling with FNO-XL (high capacity) and MLP-M (low capacity).

Effect of decoupling on unmixing accuracy

Generator	Unmixer	k	R^2 -T	R^2 -NPV
GD-Streamline	FNO	27k	−1.672	−1.968
GD-Streamline	MLP	27k	−1.558	−1.884
TCVAE	FNO	3k	−2.072	−4.531
TCVAE	FNO	27k	−1.065	−2.216
TCVAE	MLP	3k	0.434	0.065
TCVAE	MLP	27k	0.307	−0.140

Table 2. Decoupling accuracy (R^2 total & NPV) by generator, unmixer, and size k . Only TCVAE $\rightarrow$ MLP is positive; full $3 \times 2 \times 2$ ablation in the paper.

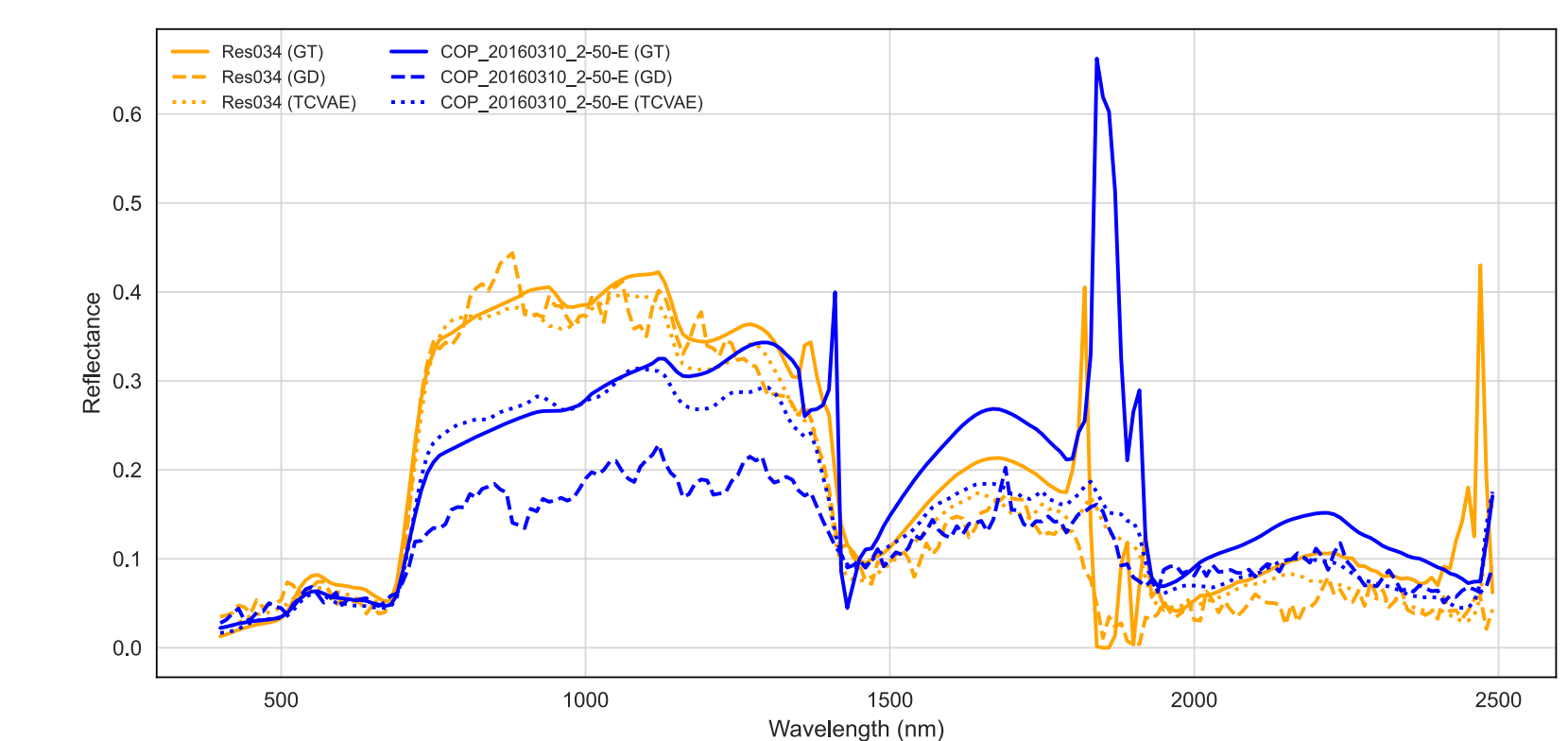


Figure 4. Ground-truth vs. generated spectra (GD-Streamline and TCVAE).

Discussion & Conclusion

- Current priors (TV, SAM, FFT) are insufficient: both generators collapse and hallucinate.
- GD-Streamline: best heuristic manifold, but noisy; priors never reach U (all $R^2 < 0$).
- TCVAE: worse metrics yet the only positive run (TCVAE $\rightarrow$ MLP, $R^2 \approx 0.43$) $\Rightarrow$ localized collapse; the KL term helps.
- NPV is the hardest endmember; scaling k helps only TCVAE $\rightarrow$ FNO (high-capacity U).
- Take-away: decoupling is sound but needs a low-bias, prior-rich G .
- Next: NPV-aware mappings; end-to-end $G+U$ tuning; flow-matching generators.

References

- Cafisch, R. E. (1998). Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, 7.
- Dennison, P. E. *et al.* (2019). Fractional cover simulated VSWIR dataset Version 2, original 10nm spectra. *EcoSIS*.
- Fraji, A. *et al.* (2025). Double Descent as a Lens for Sample Efficiency in Autoregressive vs. Discrete Diffusion Models. *arXiv*.

Note: Reference list abridged. Scan QR code for the full bibliography.



Full Paper