Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Geological Structure Embeddings from a Pretrained Diffusion Representation Model Applied to the Sleipner 4D CO2 Seismic Benchmark

Sunil Manikani[1], Dr. Murad Yaghi[2]

SUMMARY

We investigate whether a pretrained diffusion representation model (PDRM) trained purely to denoise seismic amplitude images acquires an embedding space that is more discriminative than that of a structurally matched autoencoder. Using the publicly available Sleipner CO2 seismic benchmark (249 inline cross-sections), we train a compact UNet-based PDRM and compare its embedding quality against an equivalent autoencoder capacity machine (EACM) with an identical 128-dimensional bottleneck. PDRM embeddings extracted at noise timestep t=100 yield substantially richer representations: Silhouette 0.363 vs 0.103, Davies-Bouldin 1.316 vs 2.952, and 52.6% vs 15.2% of total PCA variance in two components. An elbow analysis over k=2-8 and a timestep sweep over t in {50,100,200,400,800} show the PDRM advantage is consistent across cluster counts and robust in the range t=50-200. Critically, identified clusters are treated as data-driven groupings of seismic image character; correspondence to geological units is not established and requires independent validation against well data.

1. INTRODUCTION

The geological interpretation of 4D seismic data is central to monitoring CO2 storage integrity in CCS operations. The Sleipner field (North Sea, Norway) has injected supercritical CO2 into the Utsira Formation since 1994, providing the world's most complete CCS monitoring benchmark (Eiken et al., 2011; Arts et al., 2004). Despite this, seismic interpretation workflows remain largely manual and expert-driven, creating a bottleneck as monitoring volumes grow across global CCS deployments.

Self-supervised generative approaches are particularly attractive for seismic representation learning because they require no labelled training data — a practical necessity given the scarcity of geological ground truth. Among generative frameworks, denoising diffusion probabilistic models (Ho et al., 2020; Song et al., 2021) have shown strong representation learning capabilities in medical imaging and remote sensing (Baranchuk et al., 2022; Khader et al., 2023).

We compare a UNet-based pretrained diffusion representation model (PDRM) against an equivalent autoencoder capacity machine (EACM) with identical bottleneck dimensionality. We deliberately refrain from interpreting identified clusters as geological units without independent geological validation; our claims are restricted to embedding space structure and seismic image character.

2. METHODOLOGY

Dataset and Preprocessing

We use the Sleipner 4D seismic volume (vintage 2008) from CO2DataShare/Equinor: 249 x 468 x 1001 (inlines x crosslines x two-way time samples), float32, ~466 MB. Each of the 249 inline cross-sections is one training sample. Amplitudes are normalised by the 99th-percentile absolute value and clipped to [-1, +1]. Inlines are resampled bilinearly to 256 x 256 pixels. The first and last 10 inlines are excluded from embedding analysis (229 evaluation samples). Inline index is used as a lateral position indicator reflecting acquisition geometry, not stratigraphic depth, which corresponds to the two-way time axis.

Pretrained Diffusion Representation Model (PDRM)

The PDRM follows DDPM (Ho et al., 2020): T=1000 steps, linear beta schedule 0.0001 to 0.02. The UNet (Table 1) uses base channel width dim=32 and a 128-channel bottleneck (dim x 4). Three ResNet encoder blocks with stride-2 downsampling; bottleneck with two ResNet blocks flanking spatial self-attention; decoder with transposed convolutions and skip connections. Time conditioning via sinusoidal position embeddings with scale-shift normalisation (Dhariwal & Nichol, 2021). Total parameters: 2.1 M. Embedding extraction: add noise at timestep t, pass through encoder and bottleneck, apply global average pooling to obtain a 128-d vector. The choice t=100 is validated via the timestep sweep below.

Equivalent Autoencoder Capacity Machine (EACM)

The EACM is designed so its bottleneck matches the PDRM exactly: three strided convolutional layers (4x4 kernel, stride 2) with channel widths [32, 64, 128], flatten, linear projection to 128 dimensions, then mirrored decoder with Tanh. Trained with MSE reconstruction loss. Despite 34.0 M parameters (16x more than the PDRM), this comparison is conservative: if the PDRM wins despite far fewer parameters, the advantage of the denoising objective is conclusive.

Training Configuration

Both models: batch size 4, Adam optimiser, learning rate 1e-3, seed 42. EACM: 100 epochs, final MSE = 0.00097. PDRM: 300 epochs, final MSE = 0.025. All experiments on a single high-memory GPU.

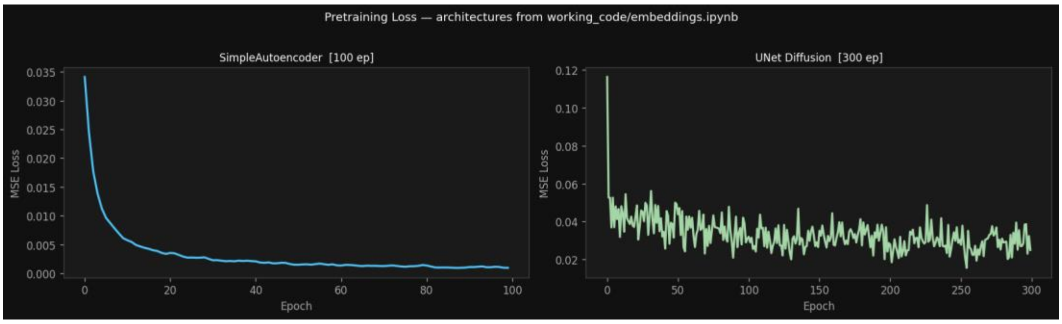

Figure 1. Training loss curves: EACM reconstruction MSE (left, 100 epochs) and PDRM noise-prediction MSE (right, 300 epochs). PDRM loss oscillates in 0.022-0.035 after epoch 125, reflecting the stochastic diffusion objective.

TABLE 1: ARCHITECTURE COMPARISON

Component	**PDRM (Diffusion UNet)**	**EACM**	**Notes**
Encoder depth	3 ResNet blocks	3 Conv layers	Matched depth
Channel widths	32->64->128	32->64->128	Matched width
Downsampling	Stride-2 conv	Stride-2 conv	Matched method
Bottleneck	2x ResNet + Attn (128-d)	Linear(128)	128-d in both
Time conditioning	Scale-shift MLP	None	PDRM-specific
Skip connections	Yes	No	UNet property
Training objective	Noise prediction (MSE)	Reconstruction (MSE)	
Parameters	~2.1 M	~34.0 M	EACM 16x larger
Training epochs	300	100	Both converged
Final loss	0.025	0.00097	Different scales

3. RESULTS

Reconstruction Fidelity

Figure 2 presents side-by-side reconstructions for six inline cross-sections spanning the lateral extent of the volume. The EACM achieves low MAE (<0.05) owing to its direct reconstruction objective. The PDRM single-step reconstruction from t=100 exhibits mild high-frequency smoothing; full DDPM reverse sampling would recover sharper amplitudes at the cost of 100 sequential denoising steps.

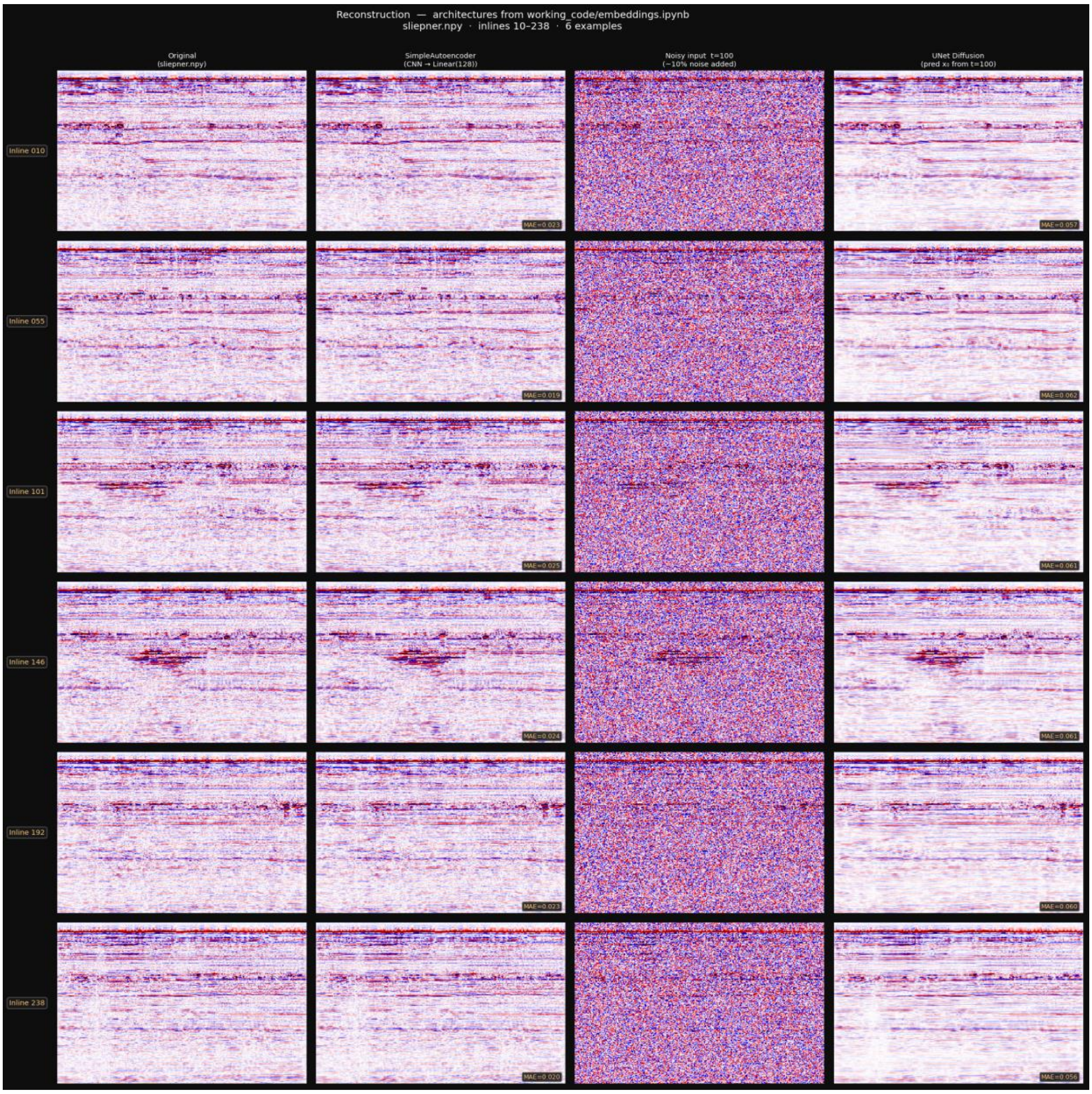

Figure 2. Reconstruction comparison for six Sleipner inlines. Columns: original; EACM reconstruction; noisy input at t=100; PDRM single-step prediction of x0. MAE annotated per reconstruction. Lateral position (inline index) shown left. Colourmap: seismic, range [-1, +1].

Embedding Space — Lateral Position Analysis

Figure 3 shows 2D PCA projections of 229 embeddings coloured by inline index as a lateral position indicator (inlines 10-238). PDRM embeddings (PC1=34.1%, PC2=18.5%; combined 52.6%) display a coherent gradient across the inline range. EACM distributes only 15.2% of variance across two components (PC1=8.3%, PC2=6.9%) with no discernible spatial ordering. Inline index reflects lateral acquisition geometry, not stratigraphic depth; the observed gradient may reflect lateral amplitude trends, acquisition footprint, or genuine geological variation. Distinguishing these causes requires independent geological analysis and is beyond the scope of this paper.

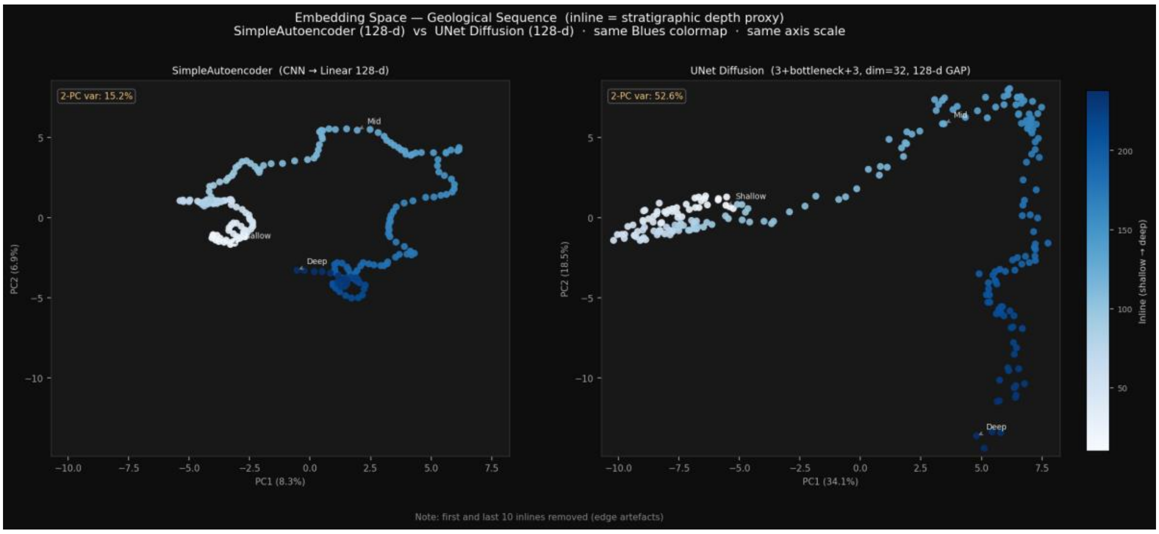

Figure 3. PCA projection of 128-d embeddings coloured by inline index (lateral position, Blues colourmap). Left: EACM (15.2% 2-PC variance). Right: PDRM (52.6% 2-PC variance). Same axis scale and colourmap for both. Inline index is a lateral position indicator, not a stratigraphic depth proxy.

Clustering and Elbow Analysis

Figure 4 shows K-Means clusters (k=3) in PCA space. Figure 5 shows the full elbow analysis over k=2-8. PDRM achieves Silhouette=0.363 and Davies-Bouldin=1.316 at k=3, versus Silhouette=0.103 and Davies-Bouldin=2.952 for the EACM.

Neither model produces a sharp silhouette elbow. PDRM silhouette is relatively stable from k=3 to k=6 (0.363-0.374) before rising further at k=7-8. EACM silhouette increases monotonically, suggesting the embedding space lacks natural cluster structure. We select k=3 as a conservative and interpretable choice aligned with prior Sleipner literature (Utsira sand body, upper and lower bounding shales), but explicitly caution: the identified clusters are data-driven groupings of seismic image character in the learned embedding space and do NOT necessarily correspond to these or any other geological units. Independent validation against well logs and interpreted horizons is required before any geological interpretation of these clusters can be made.

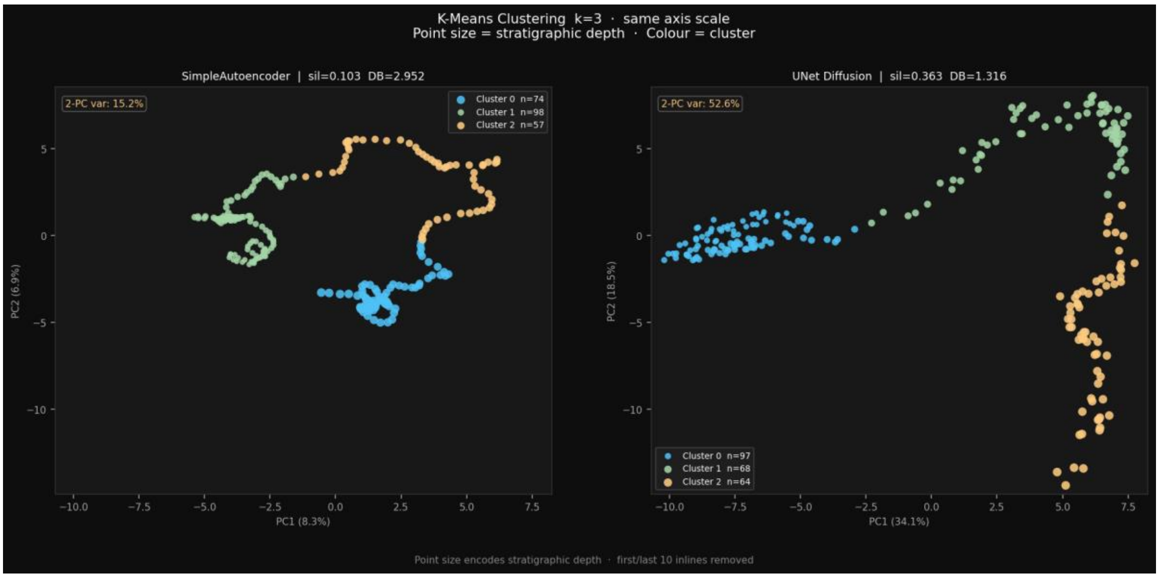

Figure 4. K-Means clustering (k=3) in PCA space. Point size encodes inline index (lateral position). Left: EACM (sil=0.103, DB=2.952). Right: PDRM (sil=0.363, DB=1.316). Clusters reflect seismic image character groupings only; geological unit correspondence is not established.

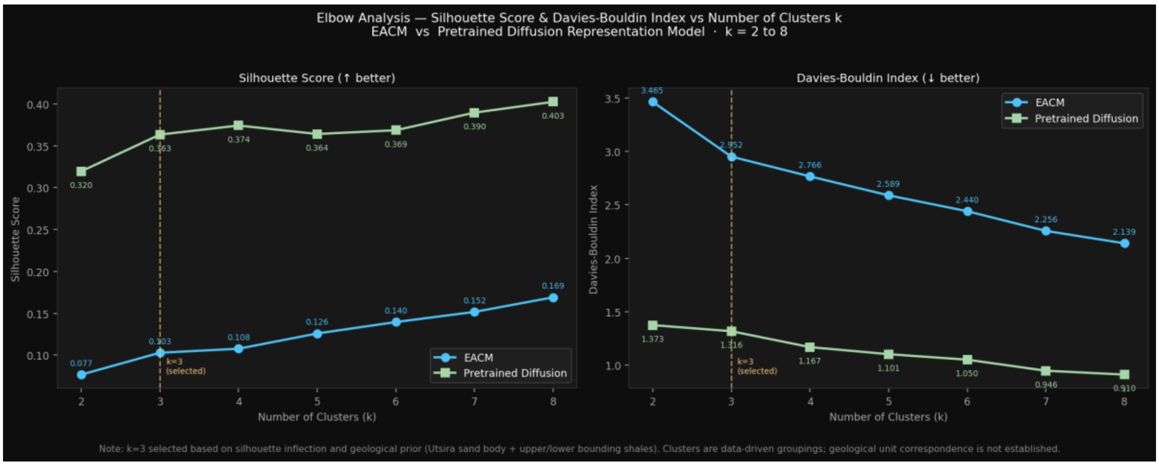

Figure 5. Elbow analysis: Silhouette score (left) and Davies-Bouldin index (right) vs k=2..8 for EACM (blue circles) and PDRM (green squares). PDRM dominates at every k. No sharp elbow is present; k=3 is a conservative choice. Note: annotated values shown per data point.

Timestep Sensitivity Analysis

Figure 6 reports PDRM embedding quality for t in {50,100,200,400,800} alongside the signal fraction alpha_t. Silhouette scores: 0.370, 0.361, 0.356, 0.312, 0.107 at t=50, 100, 200, 400, 800 respectively. The range t=50-200 forms a stable plateau (variation <4%), confirming that moderate noise levels engage the bottleneck without destroying structural signal. Performance degrades sharply at t=400 and catastrophically at t=800 (alpha_t ~0.04, near-pure noise). t=50 yields marginally the highest Silhouette (0.370) but t=100 is selected as a robust choice with negligible performance cost and strong precedent in the diffusion representation learning literature (Baranchuk et al., 2022).

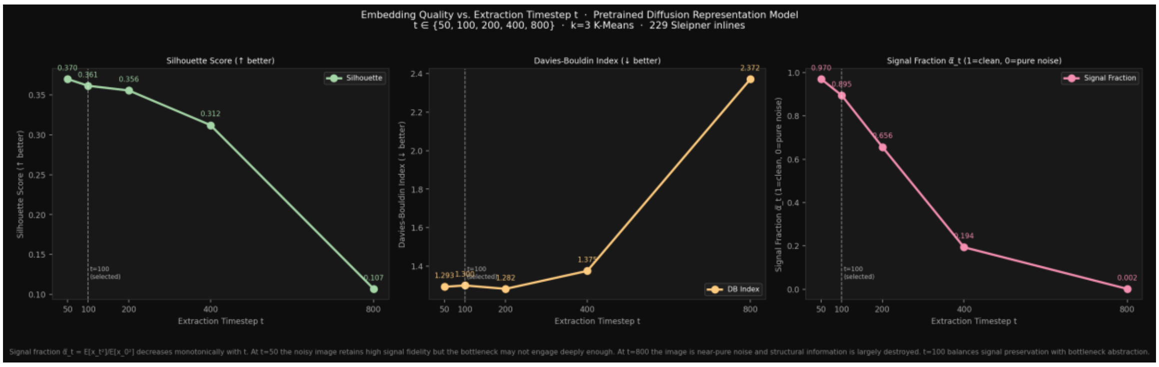

Figure 6. PDRM embedding quality vs extraction timestep t (k=3, 229 inlines). Left: Silhouette score. Centre: Davies-Bouldin index. Right: Signal fraction alpha_t (1=clean, 0=pure noise). Plateau at t=50-200 validates t=100. Sharp collapse at t>=400.

TABLE 2: QUANTITATIVE EMBEDDING EVALUATION (K=3, T=100)

Metric	**PDRM**	**EACM**	**Delta**	**Better**

Silhouette score (up)	0.363	0.103	3.52x	higher
Davies-Bouldin index (dn)	1.316	2.952	2.24x	lower
PC1 variance (%)	34.1	8.3	4.11x	-
PC2 variance (%)	18.5	6.9	2.68x	-
2-PC total variance (%)	52.6	15.2	3.46x	higher
Dead embedding dims	0/128	0/128	1.00x	-
Parameters	2.1 M	34.0 M	0.06x	-

Table 2. All metrics on 229 trimmed inlines. Silhouette and DB in standardised 128-d embedding space. K-Means: k=3, seed 42, n_init=10.

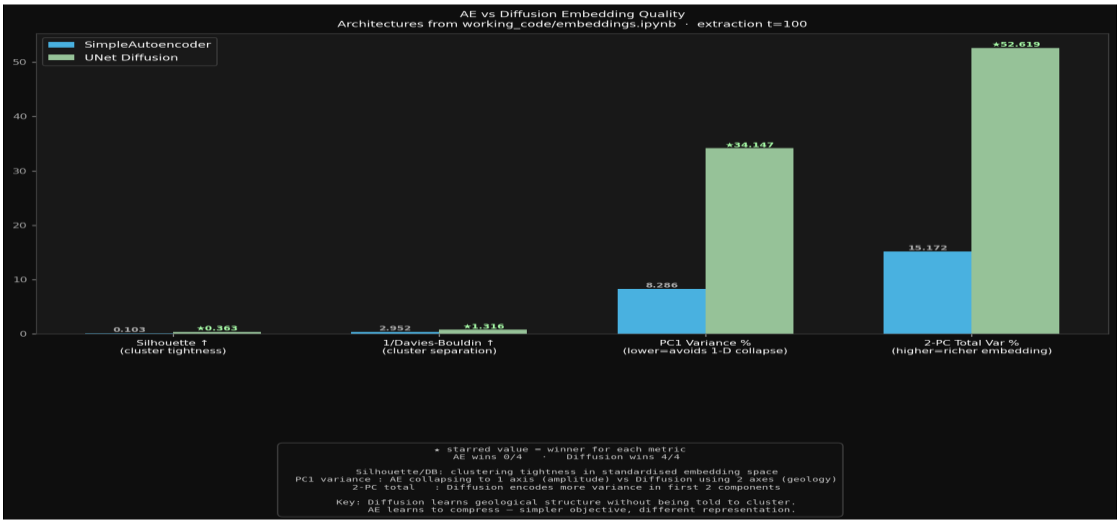

Figure 7. Bar chart of all four embedding quality metrics. Starred values indicate the winner. PDRM wins all four metrics.

4. DISCUSSION

The PDRM produces embedding spaces three to four times more structured than the EACM across all tested k values and a wide range of extraction timesteps. This advantage is achieved with 16 times fewer parameters and no direct optimisation for clustering.

We attribute this to three properties of the diffusion objective. First, noise-prediction at varying timesteps forces the UNet to learn multi-scale representations. Second, scale-shift time-conditioning produces timestep-dependent feature maps whose bottleneck at intermediate t balances amplitude variation with spatial structure. Third, UNet skip connections force the bottleneck to encode only information that cannot pass via skip paths, creating compact semantic representations.

The EACM's collapse to 15.2% two-component variance is consistent with mean-field shortcut exploitation: the model leverages mean amplitude for reconstruction rather than spatial structure. The diffusion objective removes this shortcut by varying signal-to-noise ratio across timesteps.

We deliberately avoid interpreting clusters as geological units. The inline index reflects lateral acquisition geometry and may capture amplitude trends, acquisition footprint, or genuine lateral geological variation. The K-Means groupings reflect seismic image character in the learned embedding space. Whether these groupings align with Utsira sub-units, bounding shales, or CO2 plume geometry requires intersection with well logs, interpreted horizons, and independent geological mapping — necessary steps for any operational application.

Limitations: both models are trained and evaluated on the same 249 inlines (no held-out test set); the EACM's MSE=0.00097 is consistent with memorisation. The PDRM loss oscillates after epoch 125, suggesting extended training or a learning rate schedule may improve convergence. Comparison against VAEs, contrastive learning, and masked autoencoders remains future work.

5. CONCLUSIONS

A PDRM trained on 249 Sleipner inline sections produces seismic representations substantially more structured than an EACM with 16x more parameters. Key findings:

- PDRM achieves Silhouette 0.363 vs 0.103 (3.5x) and DB 1.316 vs 2.952 (2.2x) at k=3, with the advantage holding at every tested k from 2 to 8.
- PDRM captures 52.6% of PCA variance in two components vs 15.2% for EACM, with coherent lateral spatial structure in the embedding space.
- Embedding quality is robust for t=50-200 and collapses sharply at t>=400, validating t=100 as a practical extraction point.
- Identified clusters are data-driven groupings of seismic image character and do NOT constitute geological unit assignments without independent well-data validation.
- The PDRM advantage is achieved with only 2.1 M parameters versus 34.0 M, demonstrating parameter efficiency of the denoising pretraining objective.

6. ACKNOWLEDGEMENTS

The authors thank Equinor and CO2DataShare for making the Sleipner 4D seismic dataset publicly available. This work is conducted as part of an independent doctoral research programme investigating generative AI for carbon capture and sequestration.

REFERENCES

1. Arts, R., Eiken, O., Chadwick, A., Zweigel, P., van der Meer, L., and Zinszner, B., 2004, Monitoring of CO2 injected at Sleipner using time-lapse seismic data: Energy, 29, 1383-1392.
2. Baranchuk, D., Rubachev, I., Voynov, A., Khrulkov, V., and Babenko, A., 2022, Label-efficient semantic segmentation with diffusion models: ICLR.
3. Dhariwal, P., and Nichol, A., 2021, Diffusion models beat GANs on image synthesis: NeurIPS, 34, 8780-8794.
4. Di, H., Wang, Z., and AlRegib, G., 2019, Seismic fault detection from post-stack amplitude maps via CNNs: 81st EAGE Conference.
5. Eiken, O., Ringrose, P., Hermanrud, C., Nazarian, B., Torp, T. A., and Hoier, L., 2011, Lessons learned from 14 years of CCS operations: Sleipner, In Salah and Snohvit: Energy Procedia, 4, 5541-5548.
6. Ho, J., Jain, A., and Abbeel, P., 2020, Denoising diffusion probabilistic models: NeurIPS, 33, 6840-6851.
7. Khader, F., et al., 2023, Denoising diffusion probabilistic models for 3D medical image generation: Scientific Reports, 13, 7303.

8. Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B., 2021, Score-based generative modeling through stochastic differential equations: ICLR.
9. Wrona, T., Pan, I., Gawthorpe, R. L., and Fossen, H., 2018, Seismic facies analysis using machine learning: Geophysics, 83, O83-O95.