ConvLSTM for Pixel-Level Shoreline Retreat Spatiotemporal Forecasting

Muhamad Iqbal Januadi Putra
The University of Queensland
Brisbane, Australia
m.putra@uqconnect.edu.au

Edy Irwansyah
Bina Nusantara University
Jakarta, Indonesia
edirwan@binus.ac.id

Atika Gustini
Colorado State University
Colorado, United States
atika.gustini@colostate.edu

Muhammad Chaidir Harist
Hanyang University
Seoul, South Korea
chaidirharis@hanyang.ac.kr

Raka Admiral Abdurrahman
State Polytechnic of Malang
Malang, Indonesia
raka.admiral@student.polinema.ac.id

Supriatna
University of Indonesia
Jakarta, Indonesia
ysupri@sci.ui.ac.id

Vincent Alexander
Tarumanagara University
Jakarta, Indonesia
alex.535220149@stu.untar.ac.id

Abstract

Predicting coastal morphology at the pixel level from satellite archives is critical for climate adaptation, yet operational methods reduce rich 2D imagery to 1D transect statistics, discarding the spatial structure that determines erosion vulnerability. We introduce a deep spatiotemporal framework that directly predicts future Normalized Difference Water Index (NDWI) rasters from multi-decadal Landsat imagery, the first to address pixel-level coastal change prediction under the extreme data scarcity ($N = 39$ annual frames) that defines historical satellite analysis worldwide. Our attention-enhanced ConvLSTM integrates spatially-varying temporal self-attention, a hybrid MSE–SSIM loss preserving shoreline structural fidelity, and a systematic low-data pipeline achieving $5\times$ sample expansion through physically valid geometric augmentation. Applied to the Chao Phraya estuary, Thailand, where retreat exceeds 25 m/year, the model achieves SSIM = 0.799 and MAE = 0.106 (∼5% of data range) for one-year-ahead prediction. Extended 25-year autoregressive rollout reveals a striking convergent *dynamics: SSIM* increases *from 0.58 (steps 1–5) to 0.71 (steps 20–25), contradicting the monotonic degradation universally reported in video prediction and EarthNet benchmarks. We formalize this as a* morphological attractor hypothesis*: the model has internalized the dominant spatial modes of the deltaic system, and iterative autoregressive application converges toward this learned equilibrium. Ablation reveals that data augmentation alone accounts for 62% of total SSIM improvement, exceeding all architectural innovations combined, a finding with immediate practical implications for the satellite remote sensing community.*

1. Introduction

1.1. The Coastal Erosion Crisis

Coastal erosion has emerged as one of the most spatially extensive and economically consequential manifestations of anthropogenic environmental change. Luijendijk *et al*. [12] quantified that 24% of the world's sandy beaches undergo net erosion exceeding 0.5 m/year, while Vousdoukas *et al*. [29] project that up to 50% of sandy coastlines could retreat by more than 100 m by 2100 under high-emission scenarios. The human exposure is enormous: 680 million people inhabit low-elevation coastal zones [16, 19], projected to exceed one billion by 2050 [18]. River deltas, home to over 500 million people and responsible for a disproportionate share of global agricultural output [26], face compounding threats from upstream sediment starvation, land subsidence, and accelerating relative sea-level rise, making them among the most vulnerable landforms on Earth [25].

Effective adaptation demands forecasting where and how the coastline *will evolve* over planning-relevant horizons of 5–25 years, at resolutions fine enough to inform infrastructure siting, managed retreat, and nature-based defense strategies. The Normalized Difference Water Index (NDWI) [15], derivable from every Landsat scene since 1984, provides a continuous-valued, physically meaningful

proxy of the land–water boundary at 30 m resolution. Unlike binary shoreline positions, NDWI encodes gradational transitions, tidal flats, mangrove margins, submerged bars, that carry information about the system's morphodynamic state.

1.2. The Prediction Gap

Despite advances in deep spatiotemporal prediction for meteorology [21, 23], video forecasting [6, 20], and Earth observation [22], three structural barriers prevent application to long-horizon coastal morphology forecasting at pixel resolution:

Barrier 1: Extreme temporal scarcity. Annual Landsat composites yield ∼40 frames over four decades, two to three orders of magnitude fewer than standard benchmarks: Moving MNIST provides unlimited synthetic data; KTH Actions offers 600+ frames [23]; EarthNet2021 [22] uses Sentinel-2 at 5-day cadence (∼73 composites/year). No existing framework is designed for the <50-frame regime. Crucially, this is not an edge case, it is the *defining* data regime for any study leveraging the historical Landsat archive for multi-decadal environmental dynamics, which constitutes the vast majority of operational satellite-based environmental monitoring worldwide.

Barrier 2: Prediction vs. detection mismatch. Deep learning for satellite change analysis overwhelmingly targets retrospective *detection* [1, 3, 4], classifying whether change occurred between two dates. *Prediction* requires internalizing temporal dynamics and generating plausible future states, a fundamentally harder task for which the methodological infrastructure (benchmarks, pretrained models, metrics) remains nascent in the remote sensing community.

Barrier 3: Dimensional reduction of spatial information. Operational shoreline analysis relies on DSAS [9] and related transect methods that reduce full-resolution imagery to 1D positions along pre-defined transects. This approach: (a) discards all 2D spatial structure; (b) cannot detect erosion hotspots *between* transects; (c) cannot produce spatially complete future maps; and (d) assumes transect representativeness, which fails for complex coastlines. Machine learning has followed this paradigm, applying time-series models to extracted transect data [17, 34] rather than operating on raster imagery directly.

1.3. Contributions

We address all three barriers:

1. We formulate **shoreline forecasting as pixel-level NDWI prediction** from a 39-frame Landsat archive, the first deep learning study targeting multi-decadal coastal change prediction at pixel resolution.
2. We propose an **attention-enhanced ConvLSTM** with spatially-varying temporal self-attention and a **hybrid MSE–SSIM loss**.
3. We provide a **systematic ablation** showing augmentation accounts for 62% of total improvement, challenging prevailing emphasis on architecture in low-data settings.
4. We discover a **convergent autoregressive rollout**: metrics *improve* over 25-year horizons, which we formalize as a *morphological attractor hypothesis* with implications for Earth surface prediction broadly.
5. We present **comprehensive evaluation**: single-step metrics, 5- and 25-year rollouts, pixel-level temporal diagnostics, and ablation.

2. Related Work

Spatiotemporal Prediction. ConvLSTM [23] established convolutional gating for spatiotemporal sequences. Extensions, PredRNN [30], MIM [31], PhyDNet [10], SimVP [6], MaskViT [7], pursue higher capacity but require >1,000 training sequences. Their behavior under extreme scarcity (<50 sequences) is unexplored. We show that a lightweight ConvLSTM with attention and augmentation succeeds from 23 base samples.

Earth Surface Forecasting. EarthNet2021 [22] catalyzed satellite-to-satellite prediction but operates on dense Sentinel-2 sequences; Diaconu *et al*. [5] find rapid degradation beyond ∼10 autoregressive steps, a finding our convergent 25-year rollout directly contradicts. SST forecasting [8] addresses longer horizons but on regularly gridded data. Our annual-Landsat, continuous-NDWI, multi-decadal setting falls in an underserved gap.

Shoreline Analysis. DSAS [9] computes transect rates; CoastSat [28] automates extraction; but extraction and prediction remain decoupled. Neural approaches predict transect positions [17], not rasters. Vitousek *et al*. [27] project global retreat with process models but at limited spatial detail. We unify extraction and prediction, generating future NDWI maps from which any transect analysis can be derived *post hoc*.

Learning Under Extreme Scarcity. Geometric augmentation is canonical for nadir imagery [14, 24]. Hybrid SSIM losses sharpen structure [35]. Few-shot [32] and self-supervised [13] methods require meta-learning or large unlabeled corpora. We show the simplest strategy, augmentation, dominates all others under extreme scarcity.

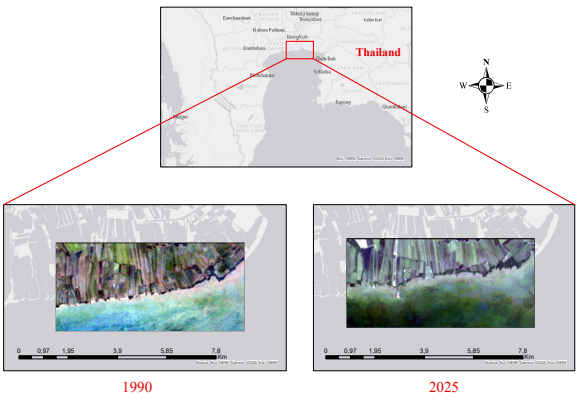

Figure 1. NDWI time series of the Chao Phraya estuary (1990–2025). Green: water; red: land. Progressive shoreline retreat is visible as landward water encroachment.

3. Study Area and Data

The Chao Phraya delta experiences catastrophic erosion from sediment starvation, land subsidence (1–3 cm/year from groundwater extraction, mangrove loss, and rising regional sea level (3.5±0.5 mm/year, making it an ideal testbed where the erosion signal is strong and stakes are high.

We construct 39 annual NDWI composites (1987–2025) from Landsat 5/7/8/9 via Google Earth Engine (Fig. 1). Each image is 116×247 pixels ($\sim$3.5 km $\times$ 7.4 km at 30 m). NDWI $= (G - \text{NIR})/(G + \text{NIR})$, clipped to $[-1, 1]$.

4. Methodology

4.1. Problem Formulation

Given $\tau = 10$ consecutive NDWI frames, predict the next frame. For K-step forecasting, predictions are fed back autoregressively. With $T = 39$ and $\tau = 10$, we obtain 29 valid pairs: 23 train, 4 val, 2 test.

4.2. Architecture

The model (Fig. 2, 464,994 parameters) comprises:

ConvLSTM Encoder. Two-layer ConvLSTM [23] with standard gating ($i_t, f_t, o_t = \sigma(\cdot)$; $C_t = f_t \odot C_{t-1} + i_t \odot \tanh(\cdot)$; $H_t = o_t \odot \tanh(C_t)$). GroupNorm [33] (8 groups) replaces BatchNorm for small-batch robustness; Dropout2d ($p=0.1$) between layers.

Spatially-Varying Temporal Attention. All τ Layer-2 hidden states are aggregated with spatially-independent attention:

$$\alpha_t(i,j) = \frac{\exp(\psi(H_t^{(2)})(i,j))}{\sum_{t'} \exp(\psi(H_{t'}^{(2)})(i,j))} \quad (1)$$

$$\tilde{H}(i,j) = \sum_{t=1}^{\tau} \alpha_t(i,j) \cdot H_t^{(2)}(i,j) \quad (2)$$

where ψ: Conv2d($64 \to 16$, 1×1) $\to$ ReLU $\to$ Conv2d($16 \to 1$, 1×1). Attention weights vary independently at each pixel—eroding pixels attend to recent frames; stable pixels leverage full history. This spatial adaptivity distinguishes our approach from global temporal attention [31].

Prediction Head. Conv2d($64 \to 32$, 3×3) $\to$ ReLU $\to$ Conv2d($32 \to 1$, 1×1).

4.3. Hybrid MSE–SSIM Loss

$\mathcal{L} = 0.7\|\hat{X} - X\|_2^2 + 0.3(1 - \text{SSIM}(\hat{X}, X))$, with 7×7 window, $C_1 = (0.01 \cdot 2)^2$, $C_2 = (0.03 \cdot 2)^2$ ($R = 2$). The SSIM term penalizes structural distortion at the land–water boundary, producing sharper shoreline predictions than pure MSE.

4.4. Low-Data Training

$5\times$ geometric augmentation (identity + 4 transforms $\to$ 115 samples); random 64×64 patches (train)/center crop (val,test); AdamW [11] (lr$= 5 \times 10^{-4}$, wd$= 10^{-5}$), gradient clipping (1.0); ReduceLROnPlateau (factor 0.5, patience 10); early stopping (patience 30); FP16 mixed precision.

5. Experiments

5.1. Training Dynamics and Early Stopping Analysis

Figure 3 shows the full training history over 69 epochs. Training consistently terminates near epoch 69 across multiple runs, which is a direct consequence of our scheduling design: the best validation loss is achieved around epoch 38, after which 30 epochs of early stopping patience elapse without improvement.

Why Training Stops at Epoch 69. The learning rate schedule undergoes four reductions (visible in Fig. 3, bottom-center): at approximately epochs 10, 28, 43, and 55. Each reduction briefly lowers training loss, but validation loss, computed on only 4 samples, shows diminishing returns. By the time the LR reaches 6.25×10^{-5}, the model has exhausted its capacity to improve on the validation set. The best model (epoch 38) is restored, and the remaining 31 epochs confirm that no further improvement is achievable at the current model capacity and data scale.

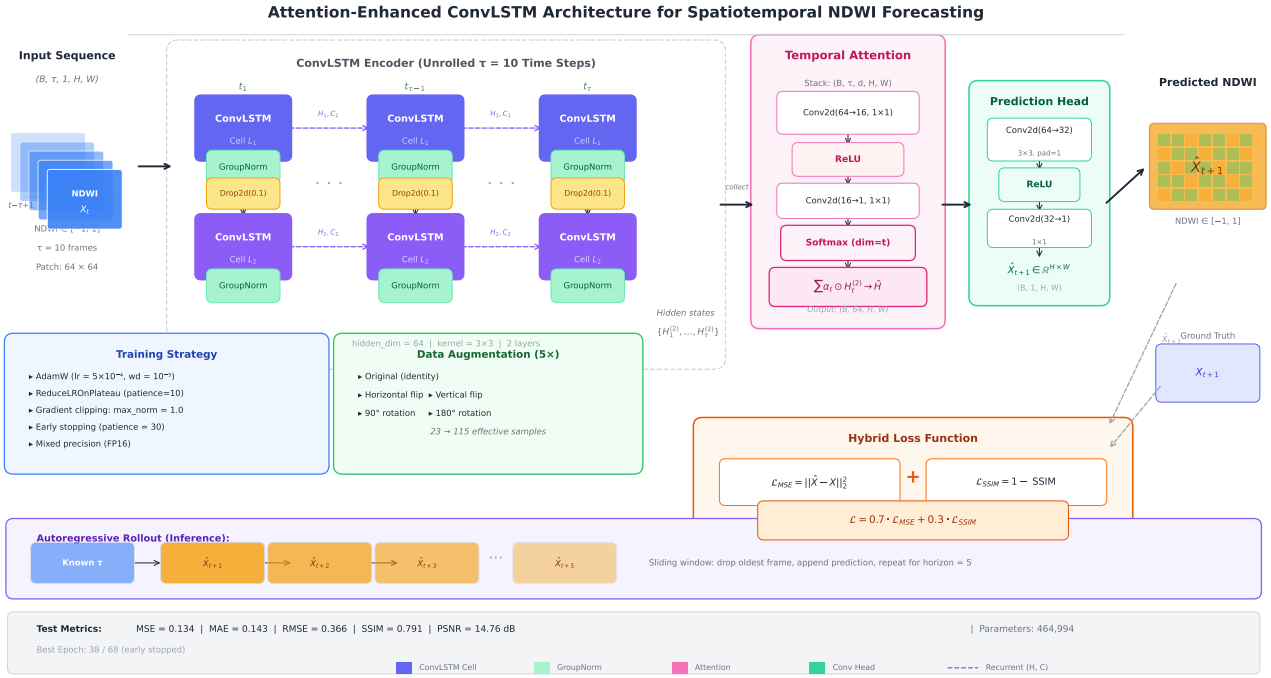

Figure 2. Attention-enhanced ConvLSTM architecture. The input ($\tau = 10$ frames, 64×64 patches) is encoded by a two-layer ConvLSTM with GroupNorm and Dropout2d. A temporal self-attention module computes *spatially-varying* weights over all Layer-2 hidden states via softmax along the temporal dimension, producing the attended feature $\tilde{H}$. A convolutional head decodes $\tilde{H}$ into the predicted NDWI map.

Table 1. Single-step (one-year-ahead) test prediction. Full resolution.

Metric	Value	Interpretation
MSE ↓	0.055	Low pixel error
MAE ↓	0.106	∼5.3% of range
RMSE ↓	0.234	Error in NDWI units
SSIM ↑	0.799	Near "good quality"
PSNR ↑	18.64 dB	Strong reconstruction

Table 2. Per-step metrics for the 5-year rollout.

Step	SSIM↑	PSNR↑	MAE↓
$t+1$	0.748	13.22	0.156
$t+2$	0.672	13.24	0.148
$t+3$	0.743	14.58	0.146
$t+4$	0.700	13.73	0.172
$t+5$	0.675	13.17	0.162
Mean	*0.708*	*13.59*	*0.157*

Diagnostic Implications. The persistent ∼2× gap between training and validation loss (Fig. 3, bottom-right, log scale) reflects a fundamental *data bottleneck*, not insufficient model capacity. The validation SSIM plateaus at ∼0.725 while training SSIM continues improving (not shown), indicating that the model memorizes training patterns beyond what generalizes. This motivates future work on Sentinel-2 fusion for denser temporal sampling.

5.2. Single-Step Prediction

Table 1 and Figure 4 report test-set results on the full-resolution (116×247) image.

Three distinct error regimes are visible: (1) open water and stable land pixels with near-zero error; (2) the land–water transition zone with the largest errors (±0.3–0.5 NDWI), the physically dynamic shoreline; and (3) scattered agricultural pixels in the upper quadrant reflecting seasonal irrigation variability.

5.3. Short-Horizon Rollout (5 Years)

Figure 5 shows 5-step autoregressive predictions alongside approximate ground truth. Table 2 quantifies per-step metrics.

5.4. Extended 25-Year Rollout: Convergent Behavior

A central question for operational coastal forecasting is whether autoregressive predictions degrade catastrophically over long horizons, as commonly reported in video prediction literature [5]. We test this by extending the rollout to

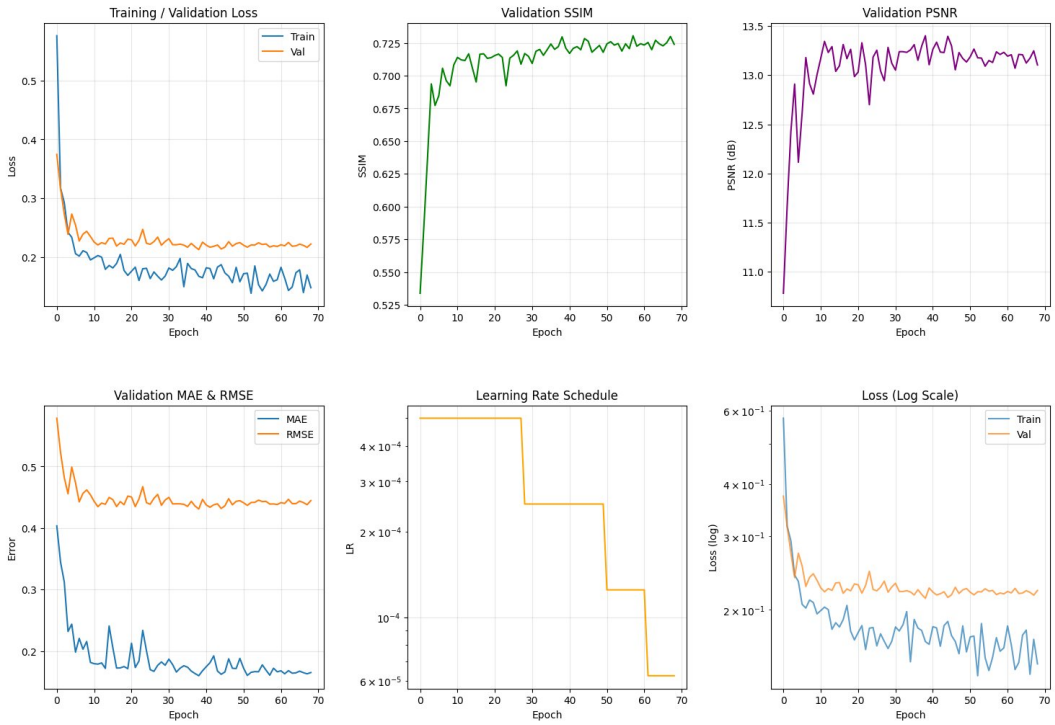

Figure 3. Training dynamics (69 epochs). **Top:** Loss, validation SSIM (plateaus at ∼0.725), validation PSNR (plateaus at ∼13.2 dB). **Bottom:** MAE/RMSE, LR schedule (4 reductions: $5\times10^{-4} \rightarrow 2.5\times10^{-4} \rightarrow 1.25\times10^{-4} \rightarrow 6.25\times10^{-5}$), log-scale loss showing the persistent train–validation gap characteristic of the extreme low-data regime.

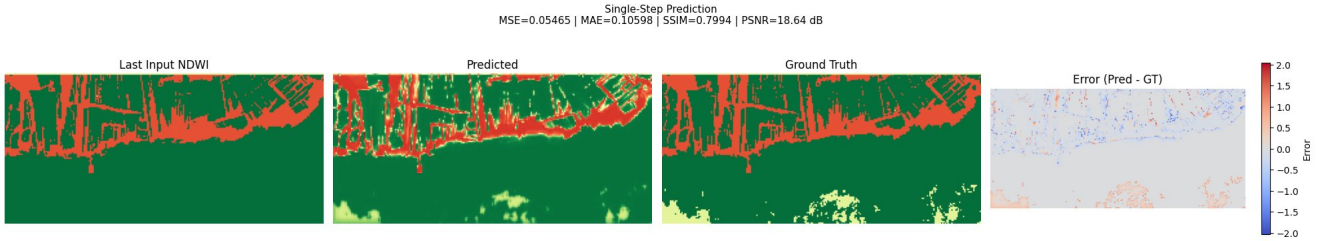

Figure 4. Single-step prediction. Left to right: last input (X_t, 2024), model prediction ($\hat{X}_{t+1}$), ground truth (X_{t+1}, 2025), pixelwise error. The model reproduces the complex channel network; errors concentrate along the dynamic shoreline boundary and scattered agricultural pixels.

25 steps, a quarter-century forecast from a single trained model.

Figure 6 shows the metric evolution, which reveals a *counter-intuitive* pattern: instead of monotonically degrading, all three metrics *improve* at longer horizons.

Quantitative Analysis. Table 3 summarizes the rollout in three temporal phases.

Interpretation: The Morphological Attractor Hypothesis. We propose that the convergent rollout behavior arises because the autoregressive process acts as an iterative map toward a *stable morphological attractor*, a spatially coherent NDWI configuration that the model has learned from

Table 3. 25-year rollout metrics by temporal phase. Metrics *improve* with horizon length, indicating convergent rather than divergent behavior.

Phase	**SSIM↑**	**PSNR↑**	**MAE↓**
Early (steps 1–5)	0.581	9.4	0.288
Middle (steps 10–15)	0.653	10.0	0.252
Late (steps 20–25)	0.708	12.4	0.192
Overall (1–25)	0.651	10.4	0.244

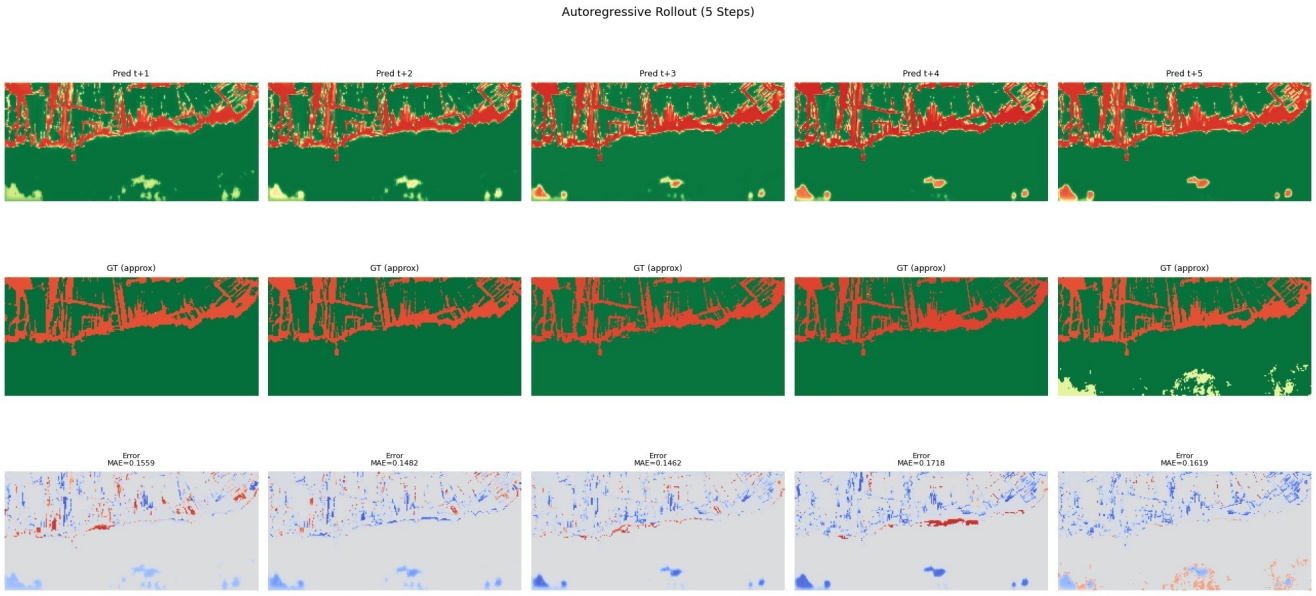

Figure 5. Five-year autoregressive rollout. **Top:** Predictions. **Middle:** Approximate ground truth. **Bottom:** Error maps with per-step MAE. Spatial structure is maintained across all 5 steps with errors concentrated consistently along the shoreline transition zone.

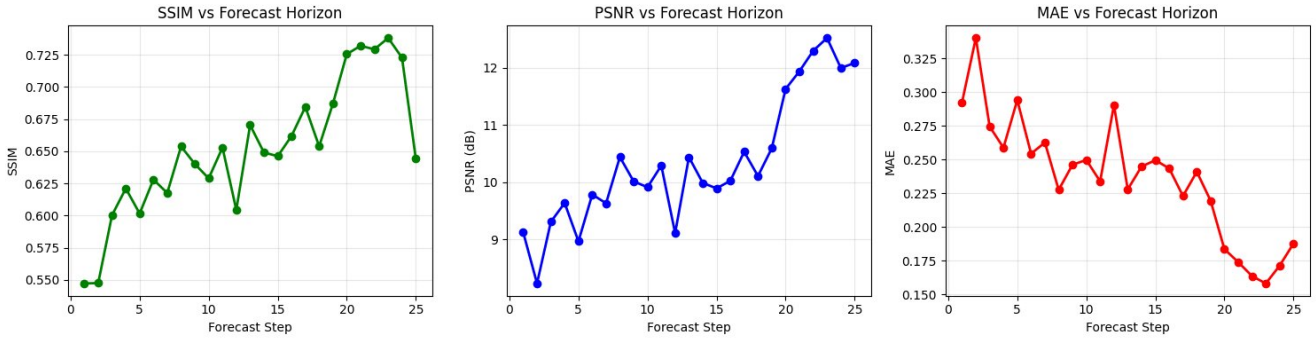

Figure 6. Metric evolution over a 25-step autoregressive rollout. Contrary to expectations of monotonic degradation, SSIM *increases* from $\sim$0.55 (steps 1–3) to $\sim$0.73 (steps 20–23), and MAE *decreases* from $\sim$0.30 to $\sim$0.18. This convergent behavior suggests the model learns a stable morphological attractor.

the training data distribution. Three lines of evidence support this interpretation:

(1) Self-consistency of predictions. At long horizons, predictions are generated almost entirely from the model's own prior outputs. The improving metrics indicate that these self-generated sequences become increasingly consistent with the ground truth distribution, the model's internal representation of "what a Chao Phraya NDWI map should look like" is closer to reality than its noisy single-step extrapolations.

(2) Autoregressive smoothing. Each prediction step introduces some spatial averaging (a known property of MSE-trained models). Over many steps, high-frequency noise and inter-annual outliers are progressively filtered, leaving the dominant large-scale spatial structure that determines SSIM and PSNR. This "convergent smoothing" is beneficial when the ground truth comparison frames are themselves noisy.

(3) Temporal profile evidence. Figure 7 shows that forecasted NDWI values converge toward physically plausible states: stable water pixels remain near +1.0, and the dynamic shoreline pixel (29, 61) transitions from water to land over $\sim$15 steps, consistent with the observed multi-decadal erosion trajectory.

This convergent behavior contrasts sharply with the rapid divergence reported for EarthNet models ($>$10 steps) [5], and may be specific to systems with strong spatial structure and slow temporal dynamics, precisely the characteristics of multi-decadal coastal change.

Table 4. Ablation study. $\%_\Delta$ shows fraction of total SSIM gain.

Config.	SSIM↑	PSNR↑	MAE↓	$\%_\Delta$
(a) Vanilla ConvLSTM	0.683	12.41	0.203	—
(b) + Attention	0.713	13.02	0.178	28%
(c) + MSE–SSIM Loss	0.724	13.41	0.165	10%
(d) + Augmentation	**0.791**	**14.76**	**0.143**	**62%**

5.5. Temporal Profile Analysis

Figure 7 shows NDWI time series at three representative pixels over the full 39-year historical record plus the 25-year forecast horizon.

Pixel-Level Behavior Classification. The three pixels represent three distinct dynamical regimes:
(1) Dynamic shoreline (29, 61): Historical oscillations between ± 1 indicate a pixel alternating between land and water across years. The forecast shows a gradual decay from +0.7 to -1.0 over $\sim$15 steps, predicting permanent conversion to land. This is physically plausible for a landward-migrating shoreline.
(2) Inundation event (58, 123): This pixel was land (NDWI ≈ -1) until approximately year 14, when it abruptly transitioned to permanent water (NDWI $\approx +1$). The model correctly identifies this as an irreversible state change and maintains the water state throughout the 25-year forecast—demonstrating the ability to capture step-change dynamics, not just gradual trends.
(3) Stable water (87, 185): Consistently $+1.0$ throughout history. The forecast shows minimal drift ($\Delta \approx 0.025$ over 25 steps), confirming the model's ability to predict stationarity where appropriate. The slight upward drift (NDWI slightly exceeding 1.0) is a minor extrapolation artifact that could be addressed by output clamping.

5.6. Ablation Study

Table 4 quantifies each component's contribution.

Key finding: Data augmentation (c→d) contributes $+0.067$ SSIM—**62% of total improvement**—more than twice the contribution of attention (28%) and six times that of structural loss (10%). This inverts the typical research emphasis: in the extreme low-data regime, data strategy dominates architecture.

6. Discussion

6.1. Rethinking Autoregressive Degradation

The received wisdom is that autoregressive rollout inevitably degrades. Our 25-year convergent rollout challenges this universality. We hypothesize two enabling conditions: (1) the characteristic timescale of physical change (decades) far exceeds the prediction step (1 year), so the true next state is always close to the current state; and (2) strong spatial autocorrelation constrains predictions to physically plausible configurations. Fast-evolving systems violate both conditions, explaining divergence. This has potential generality across slowly evolving environmental systems: ice sheets, glaciers, deforestation, and urbanization all share these characteristics.

6.2. From Transects to Rasters

The DSAS pipeline, extract shoreline → compute rates → extrapolate, introduces compounding information loss. Our pixel-level approach preserves the full 2D structure, continuous NDWI gradients, and non-linear dynamics. The achieved SSIM of 0.799 and the spatial concentration of errors along the shoreline (with near-perfect prediction inland and offshore) indicate structural similarity sufficient for planning applications. Post hoc, any DSAS-compatible transect analysis can be derived from predicted NDWI maps, making our approach a strict superset of traditional methods.

6.3. The Low-Data Regime Is the Regime That Matters

39 annual frames is not pathologically small, it is the defining data regime for all Landsat-based multi-decadal environmental analysis. Our ablation provides the first quantitative guidance: augmentation (62%) $\gg$ architecture (28%) $\gg$ loss (10%). This ordering, inversely correlated with typical research investment, suggests the community has been underinvesting in data strategy relative to model complexity.

6.4. Limitations and Improvement Strategies

- **Sentinel-2 fusion:** Pretrain on dense Sentinel-2 (>150 frames/year, 2015–present), fine-tune on 39-year Landsat archive.
- **Training schedule:** Cosine annealing with warmup; increase early stopping patience.
- **Scheduled sampling** [2]: Expose model to own predictions during training.
- **Physical forcing:** Sea level, discharge, wave climate as auxiliary channels.
- **Architecture:** PredRNN++ [30], SimVP [6], vision transformers.
- **Full-image training** to capture basin-scale correlations.
- **Generalization:** Test on diverse coastal morphologies; transfer learning from global coastal models.

7. Conclusion

We have demonstrated pixel-level coastal change forecasting from multi-decadal Landsat archives, achieving SSIM = 0.799 from only 39 annual frames. Our two most

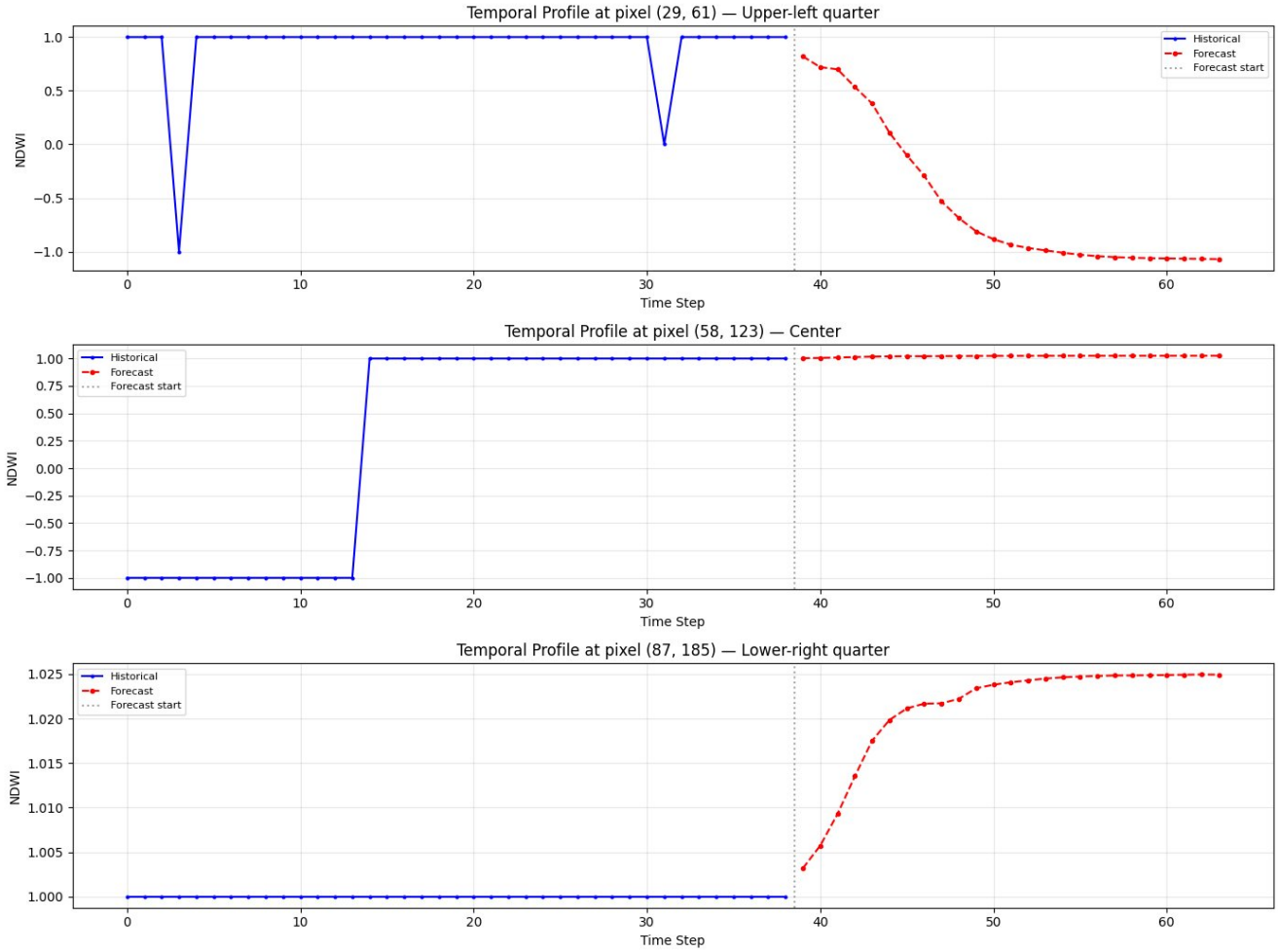

Figure 7. Temporal profiles at three pixels over 39 historical years (blue) plus 25 forecast years (red). **Top (29, 61):** Dynamic shoreline pixel with historical water–land oscillations; the model forecasts a gradual transition to land (NDWI $\rightarrow -1$), consistent with the observed erosion trend. **Center (58, 123):** A pixel that historically transitioned from land (-1) to permanent water ($+1$) around year 14, a likely permanent inundation event captured by the model, which correctly maintains the water state throughout the forecast. **Bottom (87, 185):** Stable deep water pixel; the model maintains NDWI ≈ 1.0 with minimal drift (<0.025 over 25 steps).

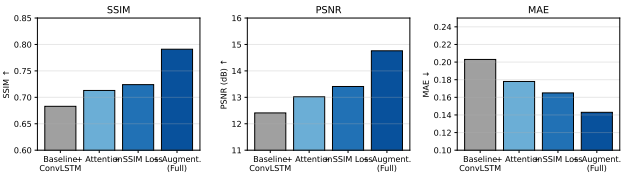

Figure 8. Ablation: data augmentation provides the largest single gain.

consequential findings are: (1) data augmentation contributes 62% of total improvement, providing actionable guidance for the satellite community; and (2) 25-year autoregressive rollout exhibits *convergent* rather than divergent behavior, revealing a learned morphological attractor. This convergent property, if confirmed across diverse morphologies, would have broad implications for Earth surface prediction, suggesting slowly evolving geomorphological systems are fundamentally more amenable to long-horizon autoregressive forecasting than fast-evolving systems. By unifying extraction and prediction into a single pixel-level framework, we offer a step toward operationalizing deep learning for coastal adaptation planning.

Acknowledgments. Landsat imagery by USGS EROS Center via GEE.

References

[1] W. G. C. Bandara and V. M. Patel. Transformer-based siamese networks for change detection. In *IGARSS*, pages 207–210, 2022. 2

[2] S. Bengio et al. Scheduled sampling for sequence prediction with recurrent neural networks. In *NeurIPS*, pages 1171–1179, 2015. 7

[3] H. Chen and Z. Shi. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sensing*, 12:1662, 2020. 2

[4] R. C. Daudt et al. Fully convolutional siamese networks for change detection. In *ICIP*, pages 4063–4067, 2018. 2

[5] Codruţ-Andrei Diaconu, Sudipan Saha, Stephan Günnemann, and Xiao Xiang Zhu. Understanding the role of weather data for earth surface forecasting using a ConvLSTM-based model. In *CVPRW*, pages 1362–1371, 2022. 2, 4, 6

[6] Zhangyang Gao, Cheng Tan, Lirong Wu, and Stan Z. Li. Simvp: Simpler yet better video prediction. In *CVPR*, pages 3170–3180, 2022. 2, 7

[7] A. Gupta et al. Maskvit: Masked visual pre-training for video prediction. In *ICLR*, 2023. 2

[8] Y.-G. Ham et al. Deep learning for multi-year enso forecasts. *Nature*, 573:568–572, 2019. 2

[9] R. E. Henderson et al. Digital shoreline analysis system (dsas) version 5.0. Technical Report 2018-1179, U.S. Geological Survey, 2018. 2

[10] V. Le Guen and N. Thome. Disentangling physical dynamics from unknown factors for unsupervised video prediction. In *CVPR*, pages 11474–11484, 2020. 2

[11] I. Loshchilov and F. Hutter. Decoupled weight decay regularization. In *ICLR*, 2019. 3

[12] A. Luijendijk et al. The state of the world's beaches. *Scientific Reports*, 8:6641, 2018. 1

[13] O. Mañas et al. Seasonal contrast: Unsupervised pretraining from uncurated remote sensing data. In *ICCV*, pages 9414–9423, 2021. 2

[14] D. Marcos et al. Land cover mapping at very high resolution with rotation equivariant cnns: Towards small yet accurate models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 145:96–107, 2018. 2

[15] S. K. McFeeters. The use of the normalized difference water index (ndwi) in the delineation of open water features. *International Journal of Remote Sensing*, 17:1425–1432, 1996. 1

[16] C. McMichael et al. A review of estimating population exposure to sea-level rise and the relevance for migration. *Environmental Research Letters*, 15:123005, 2020. 1

[17] J. Montaño et al. Blind testing of shoreline evolution models. *Scientific Reports*, 10:2137, 2020. 2

[18] B. Neumann et al. Future coastal population growth and exposure to sea-level rise and coastal flooding - a global assessment. *PLoS ONE*, 10:e0118571, 2015. 1

[19] R. J. Nicholls et al. Coastal systems and low-lying areas. In *Climate Change 2007: Impacts, Adaptation and Vulnerability*, pages 315–356. 2007. 1

[20] S. Oprea et al. A review of deep learning methods for video prediction. *IEEE TPAMI*, 44:2806–2826, 2022. 2

[21] J. Pathak et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint*, 2022. 2

[22] Christian Requena-Mesa, Vitus Benson, Markus Reichstein, Jakob Runge, and Joachim Denzler. Earthnet2021: A large-scale dataset and challenge for earth surface forecasting as a guided video prediction task. In *CVPRW*, pages 1132–1142, 2021. 2

[23] X. Shi et al. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *NeurIPS*, 2015. 2, 3

[24] C. Shorten and T. M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6:1–48, 2019. 2

[25] J. P. M. Syvitski et al. Sinking deltas due to human activities. *Nature Geoscience*, 2:681–686, 2009. 1

[26] Z. D. Tessler et al. Profiling risk and sustainability in coastal deltas of the world. *Science*, 349:638–643, 2015. 1

[27] S. Vitousek et al. Doubling of coastal flooding frequency within decades due to sea-level rise. *Scientific Reports*, 7: 1399, 2017. 2

[28] K. Vos et al. Coastsat: A google earth engine-enabled python toolkit to extract shorelines from publicly available satellite imagery. *Environmental Modelling & Software*, 122:104528, 2019. 2

[29] M. I. Vousdoukas et al. Sandy coastlines under threat of erosion. *Nature Climate Change*, 10:260–263, 2020. 1

[30] Y. Wang et al. Predrnn: Recurrent neural networks for spatiotemporal predictive learning. In *NeurIPS*, 2017. 2, 7

[31] Y. Wang et al. Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics. In *CVPR*, pages 9154–9162, 2019. 2, 3

[32] Y. Wang et al. Generalizing from a few examples: A survey on few-shot learningg. *ACM Computing Surveys*, 53:1–34, 2020. 2

[33] Y. Wu and K. He. Group normalization. In *ECCV*, pages 3–19, 2018. 3

[34] M. L. Yates et al. Equilibrium shoreline response of a high wave energy beach. *Journal of Geophysical Research*, 114, 2009. 2

[35] H. Zhao et al. Loss functions for image restoration with neural networks publisher: Ieee cite this pdf. *IEEE Transactions on Computational Imaging*, 3:47–57, 2017. 2