Compact Autoencoder Representations for Historical Retrieval of African Rainfall Anomalies

Abdulhakin Mohamud Ismail Abdirashid Omar Abdikadir Uggas

iamismail028@gmail.com rashiidmatan@gmail.com ugasbanin@gmail.com

Motivation

- Rainfall anomalies are important for understanding droughts, floods, and unusual climate conditions.
- African rainfall archives are large, seasonal, and difficult to compare directly.
- Raw rainfall-map comparison may retrieve months that are seasonally similar, but not truly similar in anomaly pattern.
- Compact representations can help retrieve meaningful dry and wet anomaly patterns more efficiently.

Method

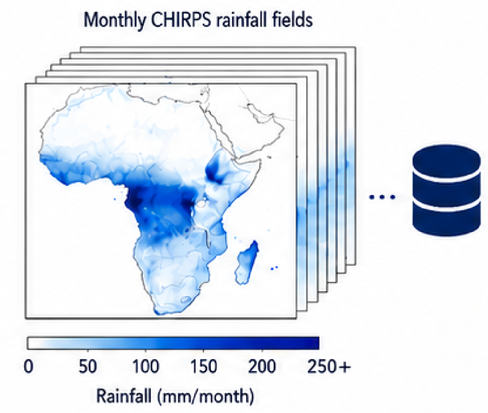
- Convert monthly CHIRPS Africa rainfall fields into standardized anomaly maps.
- Remove training-period monthly climatology to reduce seasonal effects.
- Compress each anomaly map into a 48-dimensional representation.
- Retrieve historical analogue months by comparing distances in the latent space.

Main Finding

- The linear autoencoder gives the best retrieval performance among the learned models.
- It preserves rainfall-anomaly similarity better than the convolutional autoencoder.
- The convolutional autoencoder reconstructs maps better, but gives weaker retrieval rankings.
- Reconstruction error alone is not enough to evaluate rainfall analogue retrieval.

Research Background

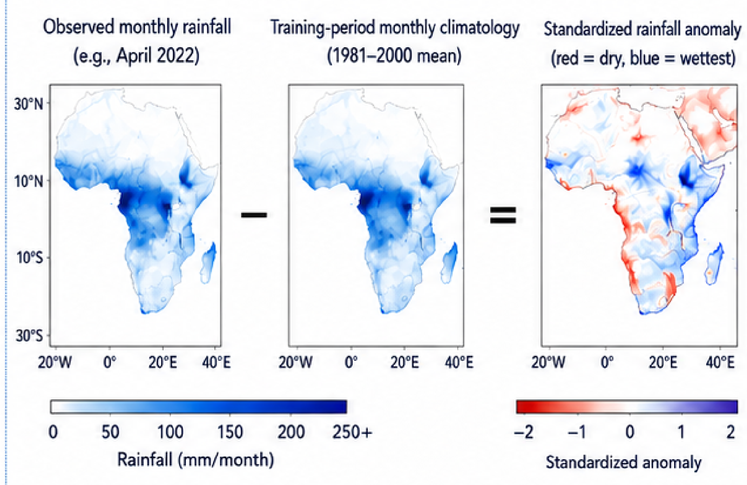
1. Large Numerical Rainfall Archives



- CHIRPS Africa monthly GeoTIFFs
- 1981–2025: 540 rainfall fields
- Rainfall measured in millimetres
- 2,258 valid land cells after preprocessing

Full-field storage and comparison become increasingly expensive as archives grow.

2. Seasonality Can Dominate Similarity



$$A_{y,m}(s) = \frac{R_{y,m}(s) - \mu_m(s)}{\max[\sigma_m(s), 5]}$$

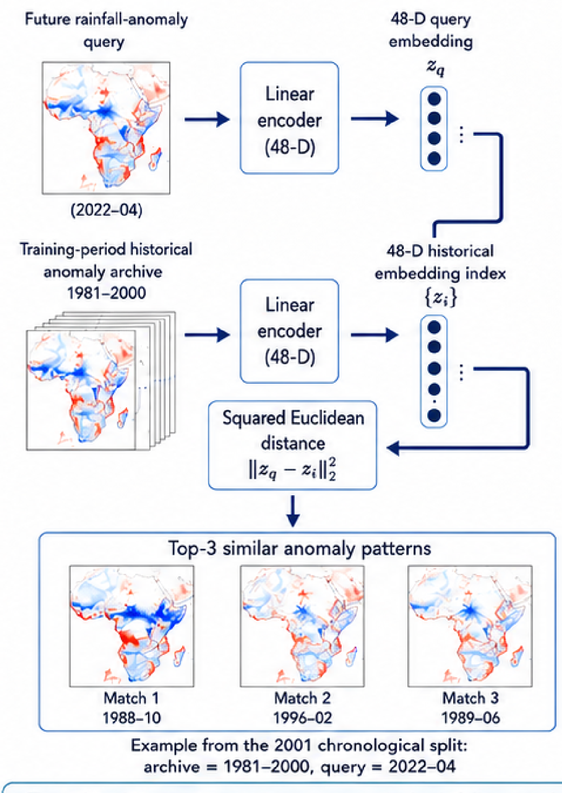
$R_{y,m}(s)$: observed monthly rainfall

$\mu_m(s)$: training-period monthly climatology

$\sigma_m(s)$: training-period monthly standard deviation

Removing monthly climatology highlights unusually dry and wet spatial patterns.

3. Historical Analogue Retrieval

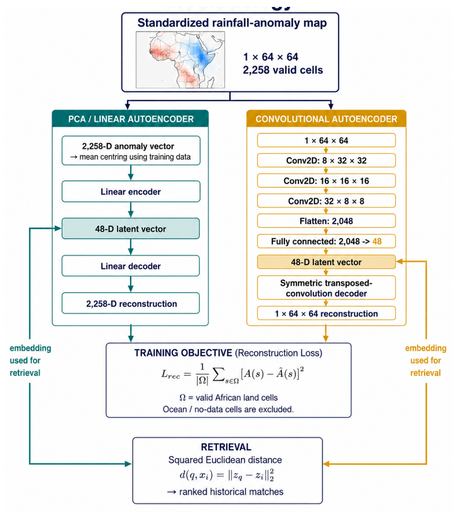


Example from the 2001 chronological split: archive = 1981–2000, query = 2022–04

The objective is to preserve numerical rainfall-anomaly similarity while reducing index size and retrieval cost.

Research Question — Can compact autoencoder representations retrieve similar historical African rainfall patterns?

Methodology



Key points

- Standardized anomaly maps are computed using training-period monthly climatology.
- Each rainfall anomaly map is compressed into a 48-dimensional latent vector.
- Historical archive months are ranked using Euclidean distance in latent space.
- The closest archive months are returned as historical rainfall analogues.

Experiment & Results

Dataset and protocol

- Dataset: CHIRPS v2 Africa monthly rainfall GeoTIFFs from 1981–2025.
- Maps are resized to 64×64 for model training and retrieval.
- Ocean and no-data cells are removed, leaving 2,258 valid African land cells.
- Five chronological train/test splits are used.
- Training-period months form the archive; future months are used as queries.

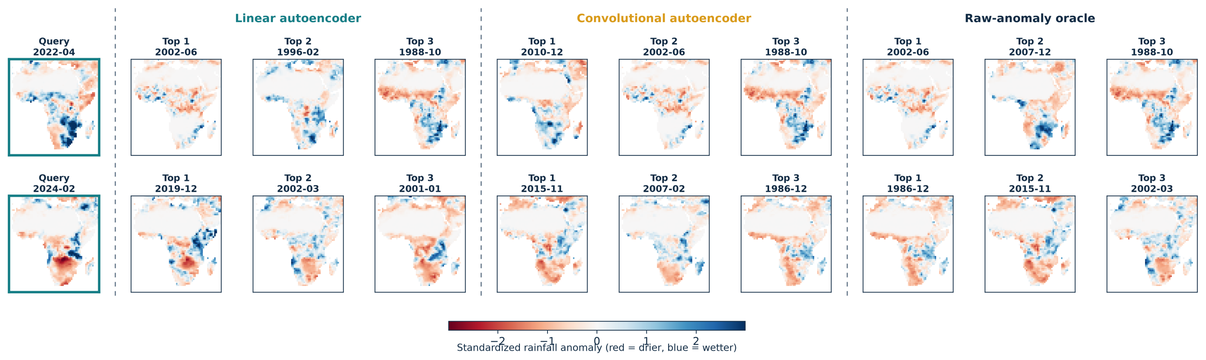
Model Performance Comparison

Model	nDCG@10	Recall@10	Recon. MSE
Raw anomaly oracle	100.00%	100.00%	–
Linear Autoencoder	*98.74%*	*70.10%*	0.1740
Convolutional Autoencoder	97.13%	56.40%	*0.1005*

Main result

- Linear Autoencoder achieves the highest nDCG@10 and Recall@10.
- Convolutional Autoencoder achieves the lowest reconstruction MSE.
- Better reconstruction does not always lead to better retrieval.
- For this task, retrieval quality is more important than reconstruction quality.

Historical Retrieval Example



Query month: April 2022. Historical archive: 1981–2020. The model retrieves the most similar historical anomaly maps. Red areas show drier-than-normal conditions; blue areas show wetter-than-normal conditions.

Conclusion & Takeaways

- Compact representations reduce storage and search cost.
- The 48-D linear autoencoder gives the best retrieval-quality trade-off among learned models.
- Reconstruction error is not a reliable proxy for retrieval performance.
- Linear autoencoders provide a strong baseline for African rainfall analogue search.

Future work: regional climate-zone evaluation, validation against documented drought/extreme-rainfall events, retrieval-aware contrastive learning, and interactive rainfall analogue search.