CLV: A Novel Framework for Enhanced Anomaly Detection and Attribution in Multivariate Time Series Data

Tolulope Ale[1], Nicole−Jeanne Schlegel[2], Vandana P. Janeja[1]

1. Information Systems Department, University of Maryland, Baltimore County; 2. Geophysical Fluid Dynamics Laboratory NOAA Princeton, NJ

iharp.umbc.edu

Motivation

Anomalies in climate data, such as those leading to unprecedented melt events, often result from complex interplays of conditions that are not readily discernible through univariate analysis. For instance, the 2019 melt was due to a series of anomalous conditions such as abnormally low winter snow cover, spring heat waves, and clear summer skies, which were identified as critical contributors to the unprecedented levels of ice melt. Therefore, we develop Cluster-LSTM-VAE (CLV), a multivariate anomaly detection and feature attribution framework to unveil these interactions, providing a clearer picture of the correlative factors behind extreme climate events.

Problem Definition

This research addresses the complex challenge of analyzing multivariate time series data $T = t_1, \ldots t_m$, where each time series are unique observations. The primary aims of this work are: first, to detect anomalous periods within the multivariate time series; second, to analyze the historical trend of climate anomalies; and third, to identify the principal features that drive these identified anomalies.

Dataset

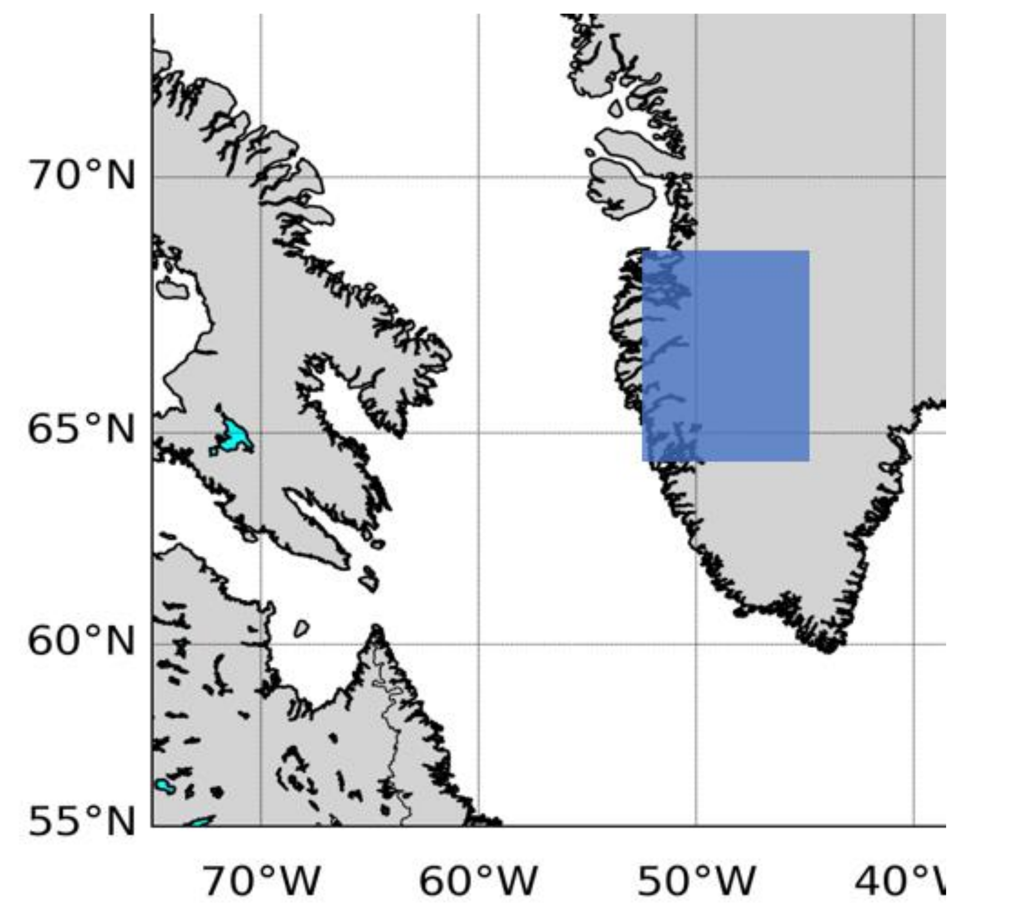

Figure 1: Southwest Greenland. The blue-shaded area is the study region.

Dataset: ERA5 Reanalysis data
Spatial Dimension: 33 long x 17 lat
Temporal Dimension: 1941 - 2020

Features	Description
u10	Eastward component of the 10m wind
v10	Northward component of the 10m wind
t2m	Air temperature at 2m above
ssrd	Amount of solar radiation that reaches a horizontal plane at the surface of the earth.
strd	Amount of thermal radiation emitted by the atmosphere and
skt	Temperature of the surface of the Earth.
asn	Snow albedo
sd	Snow depth
smlt	Snowmelt
sp	Pressure of the atmosphere at the surface of land and sea.
msl	Pressure of the atmosphere at the surface of Earth adjusted to the height of mean sea level
tcc	Total cloud cover
tp	Total precipitation

Overview of Framework

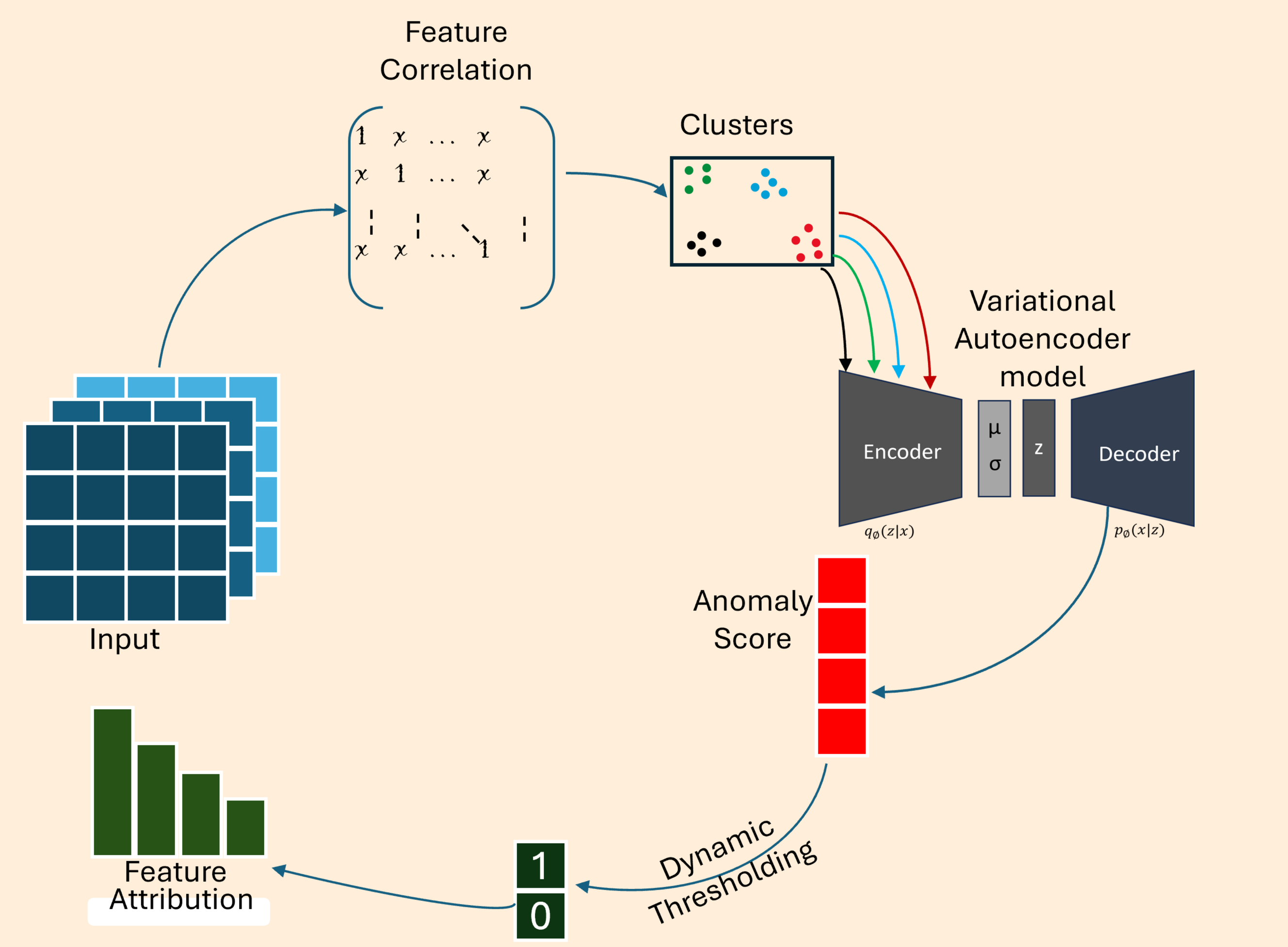

Figure 2: Proposed Methodology

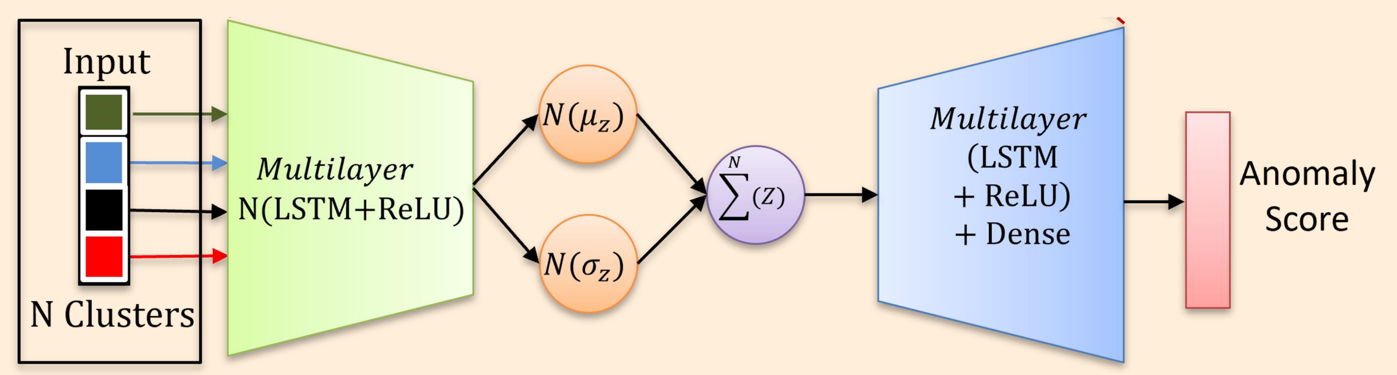

Figure 3: Variational Autoencoder Architecture

Ablation Study

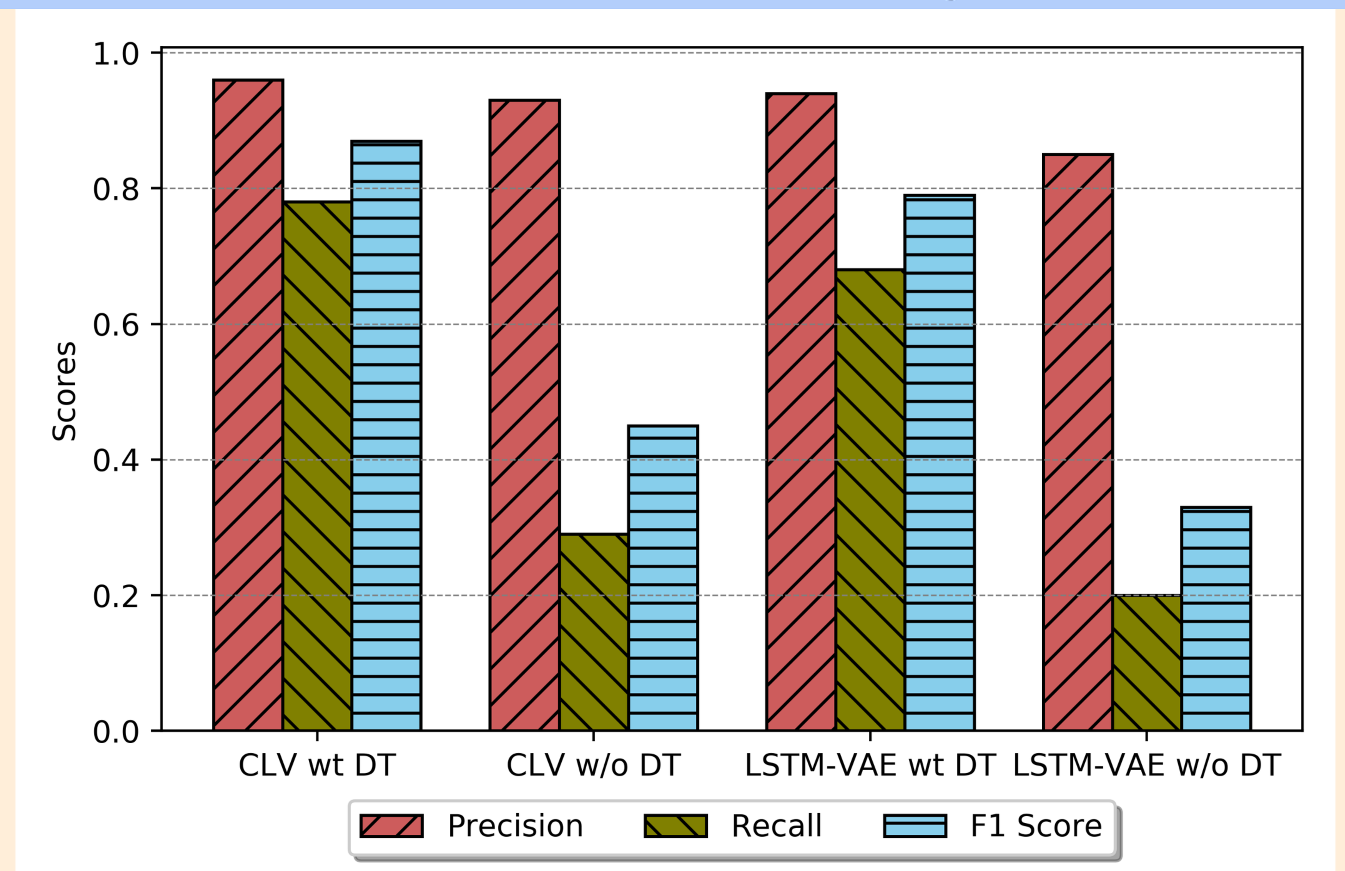

Figure 4: Ablation analysis of the components of the methodology.

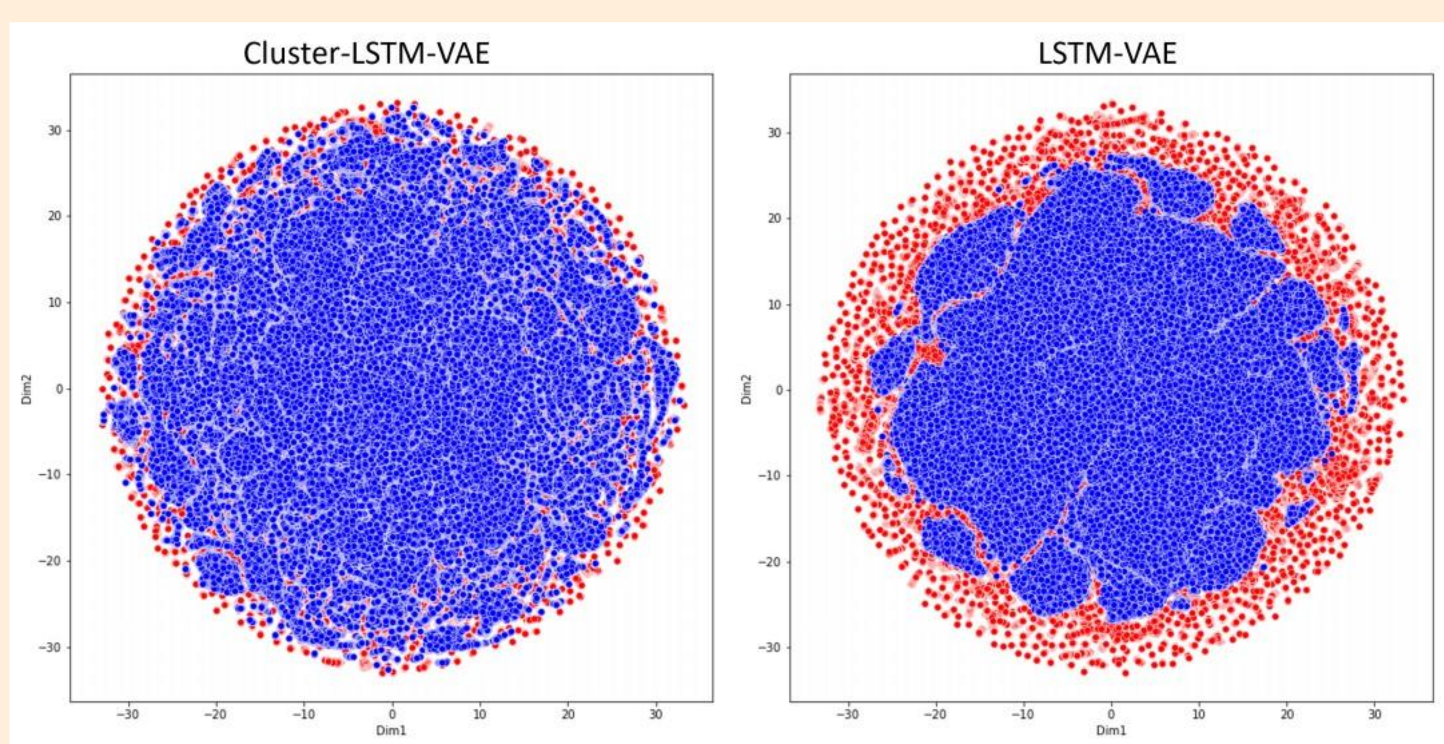

Figure 5: Analyzing model reconstruction performance in clustered and un-clustered inputs.

Statistically Significant Test

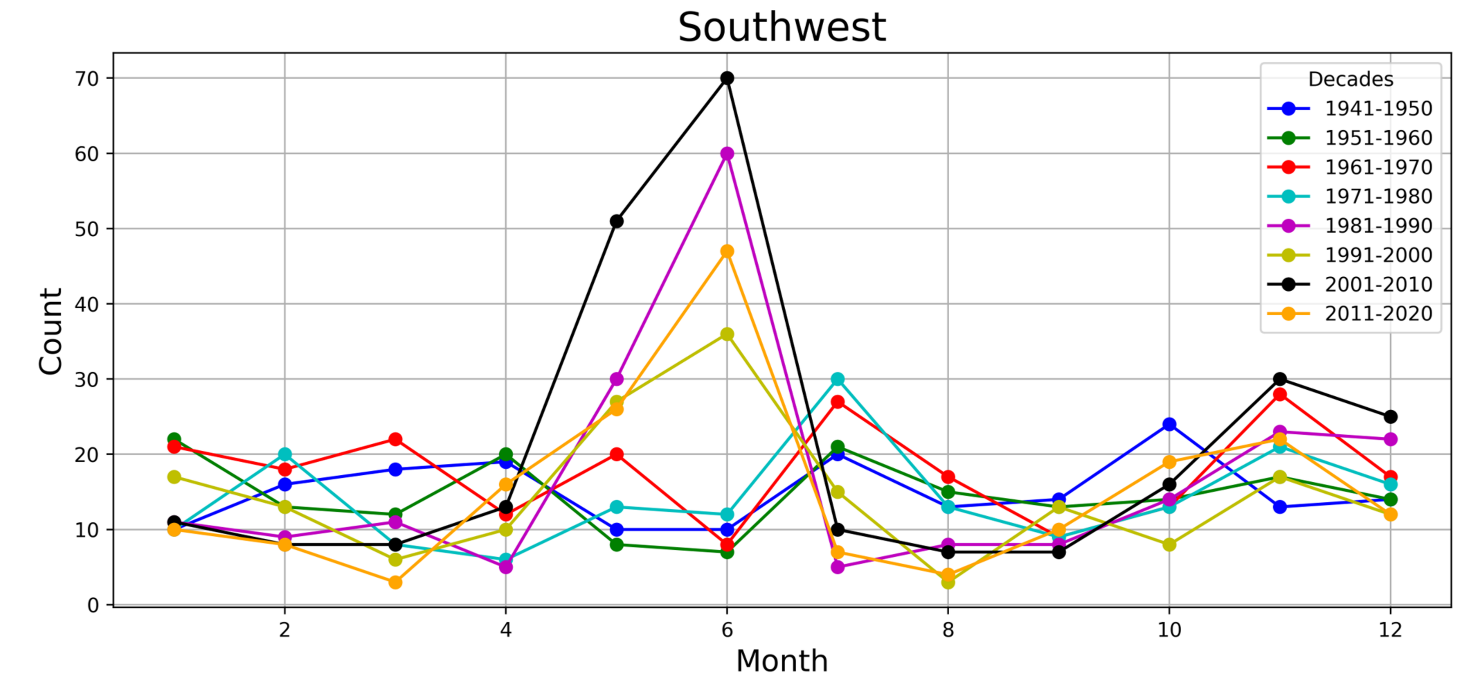

Figure 6: Month-by-month per decade aggregate of identified anomalies in Southwest

Table 2: Analyzing the differences in T-test scores of AMJJ (April, May, June, and July) and SON (September, October, and November) for Pre- and Post- 1981.

Period	T-Stat	P-Value
AMJJ	-7.42	1.26E-13
SON	0.05	0.96

Feature Attribution

Table 3: Per decade feature attribution count for Southwest region for AMJJ months

1941-1950	1951-1960	1961-1970	1971-1980	1981-1990	1991-2000	2001-2010	2011-2020
asn → 35	asn → 29	asn → 37	asn ↓ 13	asn ↓ 6	asn ↓ 4	asn ↓ 5	asn ↓ 10
skt ↓ 16	skt ↓ 15	skt ↓ 21	skt → 28	skt ↑ 49	skt → 38	skt ↑ 66	skt ↑ 59
tcc ↓ 8	tcc ↓ 11	tcc ↓ 7	tcc ↓ 15	tcc → 23	tcc → 23	tcc → 33	tcc ↓ 9
		sp ↓ 1	sp ↓ 2	sp ↓ 5	sp ↓ 9	sp ↓ 11	sp ↓ 7
	u10 ↓ 1			u10 ↓ 2	u10 ↓ 4	u10 ↓ 5	u10 ↓ 1
			strd ↓ 2	strd ↓ 14	strd ↓ 9	strd ↓ 20	strd ↓ 9
		t2m ↓ 1	t2m ↓ 1		t2m ↓ 1	t2m ↓ 4	t2m ↓ 1
				msl ↓ 1			

Evaluation

Table 4: Model accuracy in terms of precision, recall, and F1 score

Method	WADI			SMD		
	Prec	Rec	F1	Prec	Rec	F1
DAGMM	0.54	0.27	0.36	0.59	0.87	0.7
LSTM-VAE	0.88	0.15	0.25	0.79	0.7	0.78
MAD-GAN	0.41	0.34	0.37			
GDN	0.98	0.4	0.57			
OmniAnomaly				0.83	0.94	0.89
CLV wt DT	**0.96**	**0.78**	**0.87**	**0.97**	**0.75**	**0.85**

Conclusion

- The research highlighted a pivotal shift in the occurrence of anomalies before and after the year 1981. Anomalies were less pronounced in the early period, while a significant increase was observed post-1980.
- skt and tcc are influential features in the last decade in the southwest region. In contrast, asn and skt were the prevailing features in the earlier decades. The findings correspond with the intensification of Arctic warming since 1979, according to the recorded literature.

Reference

- Ale, Tolulope, Nicole-Jeanne Schlegel, and Vandana P. Janeja. "Harnessing Feature Clustering For Enhanced Anomaly Detection With Variational Autoencoder And Dynamic Threshold." arXiv preprint arXiv:2407.10042 (2024).
- Zhang, Q., Huai, B., van Den Broeke, M. R., Cappelen, J., Ding, M., Wang, Y. & Sun, W. Temporal and spatial variability in contemporary Greenland warming (1958–2020). Journal of Climate 35, 2755–2767 (2022).
- An, J. & Cho, S. Variational autoencoder-based anomaly detection using reconstruction probability. Special lecture on, i.e., 2, 1–18 (2015).

iHARP NSF HDR INSTITUTE

MData lab