Comparative Assessment of Dimensionality Reduction Techniques for Multivariate Analysis of Coral Bleaching Thermal Stress

Luana D. P. E. Casado[2], Ingrid G. C. Lira[2], João V. da C. Gomes[2], Maximiliano A. da S. Lopes[1], Angélica F. de Castro[2]

[1]Departamento de Informática, Faculdade de Ciências Exatas e da Terra Universidade do Estado do Rio Grande do Norte (UERN), Av. Prof. Antônio Campos, S/N, BR 110, Km 48 - Costa e Silva, Mossoró, RN – CEP: 59.610-090

[2]Programa de Pós-Graduação em Ciência da Computação – Universidade do Estado do Rio Grande do Norte (UERN) e Universidade Federal Rural do Semi-Árido (UFERSA), Av. Francisco Mota, s/n – Bairro Costa e Silva, Mossoró/RN, Brasil

{luana.casado,ingrid.lira,joao.gomes}@alunos.ufersa.edu.br, maximilianolopes@uern.br, angelica@ufersa.edu.br

Abstract. *This paper presents a comparative analysis of seven dimensionality reduction (DR) techniques applied to the BCO-DMO global coral bleaching dataset, investigating multivariate patterns of oceanic thermal stress from 1998 to 2019. Linear Discriminant Analysis proved most effective for supervised discriminant visualization of ocean basin structure, outperforming unsupervised techniques in revealing class-defined thermal patterns within this dataset. Temporal projections reveal progressive homogenization of thermal stress across major open ocean basins and an anomalous regime shift in the Arabian Gulf, with a 30% increase in maximum thermal stress between 1998–2004 and 2012–2019. These findings highlight the potential of supervised DR as a diagnostic tool for analyzing complex environmental data through an open-source web platform that supports interactive and reproducible exploration.*

1. Introduction

The exponential growth of environmental monitoring data has transformed climate science into a high-dimensional analytical challenge, rendering traditional approaches limited in performance and scalability [Badshah et al. 2024]. In the context of the climate emergency, this complexity is intensified by the Big Data era, demanding methods capable of handling highly complex multivariate data [Hassani et al. 2019]. In this scenario, oceanographic datasets integrate multiple variables over time series and different regions, making pattern identification difficult for conventional techniques.

Meanwhile, environmental monitoring has generated large volumes of meteorological and geospatial data, demanding approaches capable of capturing nonlinear relationships at scale [Hamdan et al. 2024]. Coral reefs sit at the intersection of this complexity and the climate emergency, being strongly impacted by rising sea surface temperatures. Bleaching events have become more frequent and severe since the 1980s, with a significant reduction in the interval between occurrences [Hughes et al. 2018]. Understanding the possible convergence of thermal stress patterns and regime shifts in regions such as the Arabian Gulf and the Red Sea therefore requires analytical approaches capable of revealing structures in high-dimensional data.

Dimensionality Reduction (DR) techniques offer a grounded pathway for this type of analysis, transforming high-dimensional data into lower-dimensional representations that preserve statistically significant structures, from linear approaches that capture global variance to nonlinear techniques that preserve local topological structure [Lopes et al. 2020]. However, different algorithms make fundamentally different assumptions about data geometry, and their relative effectiveness in revealing regional patterns in oceanographic thermal stress data has not been directly evaluated, which motivates the comparative approach adopted in this work.

This paper addresses the following research questions: (1) "Which DR technique best reveals the discriminant structure between ocean basins in a global coral bleaching dataset?" (2) "Do multivariate patterns of oceanic thermal stress show evidence of structural change between 1998 and 2019, and if so, which basins exhibit the most anomalous trajectories?" To answer them, we conducted a systematic comparative evaluation of seven algorithms, namely Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Neighbourhood Components Analysis (NCA), Kernel Principal Component Analysis (Kernel PCA), t-distributed Stochastic Neighbor Embedding (t-SNE), Locally Linear Embedding (LLE), and Isometric Mapping (Isomap), applied to the BCO-DMO dataset across three chronological periods, implemented through ReduzLab.

The main contributions of this work are: (i) a systematic comparative evaluation of seven DR techniques on a large-scale multivariate oceanographic dataset, using quantitative clustering metrics and structured visual inspection; (ii) empirical evidence of progressive homogenization of thermal stress among major open ocean basins between 1998 and 2019; and (iii) the identification of an anomalous escalation of thermal stress in the Arabian Gulf, with TSA_Maximum (maximum thermal stress) increasing by 30% during the analyzed period, constituting preliminary evidence of an accelerated regional climate regime shift.

This paper is organized as follows: Section 2 reviews related work on DR techniques and thermal stress in coral reef systems; Section 3 describes the dataset, preprocessing pipeline, and evaluation methodology; Section 4 presents and discusses the experimental results; and Section 5 presents conclusions and future perspectives.

2. Background and Related Work

The application of dimensionality reduction techniques to high-dimensional datasets has been explored in multiple domains. Linear methods such as PCA [Jolliffe and Cadima 2016] are widely used to capture global variance, while nonlinear approaches such as t-SNE [Van der Maaten and Hinton 2008], Kernel PCA [Schölkopf et al. 1998], Isomap [Tenenbaum et al. 2000], and LLE [Roweis and Saul 2000] preserve local topological structure in complex data. Shah and Joshi (2021) comparatively review these families, observing that algorithm selection is highly sensitive to data geometry and analytical objective.

Recent works have highlighted the interpretive challenges inherent to low-dimensional projections. Jeon et al. (2025) identify metric distortion and projection instability as central barriers to reliable interpretation, while Montambault et al. (2024) propose DimBridge, a tool that uses predicate logic to explain emerging visual patterns in DR projections. Both reinforce the need for multi-algorithm evaluation; in this context, Silhouette Score and Davies-Bouldin Index are established metrics for quantifying intracluster cohesion and intercluster separation [Rousseeuw 1987; Davies and Bouldin 1979], grounding qualitative observations in quantitative evidence independent of projected space geometry.

The global intensification of coral bleaching events is one of the most documented consequences of anthropogenic climate change. Hughes et al. (2018) demonstrate that mass bleaching events have become more frequent since the 1980s, with 1998–2016 recording the most severe episodes globally, and that the main factor is the accumulation of thermal anomalies, not gradual warming. The Arabian Gulf and the Red Sea are particularly vulnerable enclosed ecosystems: Chaidez et al. (2017) identified in the Red Sea warming rates above the global average, reaching up to 0.45 °C per decade in the northern portion, evidencing accelerated warming with potentially severe impacts on their ecosystems.

3. Materials and Methods

This section describes the materials and methods employed in the comparative evaluation of DR techniques. It covers the dataset, the preprocessing pipeline, the analytical platform, the techniques evaluated, their motivations, as well as the evaluation protocol and sampling strategy adopted to ensure computational viability while preserving class representativeness.

3.1. Dataset

The analysis is based on the "Coral Reef Global Bleaching" dataset, originally compiled by the Biological and Chemical Oceanography Data Management Office (BCO-DMO) and systematically documented by Van Woesik and Kratochwill (2022). Although accessible through the Kaggle[1] platform, the dataset is a globally consolidated and peer-reviewed resource, spanning records from 1980 to 2020. In its original form, it contains over 41,000 records and 62 variables, including coral identification metadata, geographic coordinates, dates, environmental factors, and thermal stress metrics, characterizing a high-dimensional multivariate structure that motivates the use of data reduction analysis.

3.2. Preprocessing Pipeline

A structured data cleaning pipeline was implemented using NumPy and Pandas. As DR algorithms operate on continuous numerical matrices, 24 variables were discarded: categorical attributes, unique identifiers, and free-text fields were removed; geographic coordinates and coastal distance were excluded to avoid spatial confounding; fine-grained temporal metrics (Date_Day, Date_Month) were discarded to eliminate seasonal noise; temperature columns on the Kelvin scale were removed due to multicollinearity with their Celsius equivalents; and Percent_Cover was excluded as a

[1] Database for the study: https://www.kaggle.com/datasets/mehrdat/coral-reef-global-bleaching

post-bleaching biological indicator rather than a direct thermal predictor. The categorical variable Ocean_Name was maintained exclusively as the supervised target variable, enabling temporal observation of how thermal anomaly patterns differ between ocean basins.

Records containing missing values (encoded as 'nd') were removed, ensuring statistical integrity. An initial density analysis revealed that statistically adequate records concentrate between 1998 and 2019; the year 2020 was excluded due to reduced observation count (n = 50). The remaining data were segmented into three chronological phases: (i) 1998–2004 (n = 6,029); (ii) 2005–2011 (n = 14,909); and (iii) 2012–2019 (n = 11,669). This segmentation enables temporal tracking of changes in discriminant structure over nearly two decades.

3.3. Analytical Platform

The comparative evaluation was conducted using ReduzLab[2], an open-source web platform which integrates seven dimensionality reduction algorithms in an interactive interface that supports tabular data in CSV and XLSX formats. The platform allows users to apply StandardScaler normalization to the loaded data, a critical step for dimensionality reduction, as it prevents variables with larger absolute scales from dominating distance calculations. The platform provides a straightforward interface for dataset upload, and upon data submission, displays alerts, a data preview, and the available algorithmic options. It also provides execution time metrics for all algorithms, cumulative explained variance for PCA, and quantitative clustering metrics such as Silhouette Score and Davies-Bouldin Index, in addition to supporting stratified sampling to manage computational constraints while preserving class representativeness.

3.4. Algorithms Evaluated

Seven DR techniques were evaluated, spanning the spectrum from linear to nonlinear. PCA [Jolliffe and Cadima 2016] was included as a linear baseline to assess whether global variance decomposition is sufficient to separate ocean basins. LDA [Martinez and Kak 2001], the only supervised method evaluated, was included to quantify the discriminant structure between basins and track its temporal stability. NCA [Goldberger et al. 2004] was included to verify whether supervised metric learning reveals structures not captured by linear methods. Kernel PCA [Schölkopf et al. 1998] was included to test whether nonlinear capture of global variance reveals clusters invisible to linear methods. t-SNE [Van der Maaten and Hinton 2008] was included to assess whether local topological relationships reveal clusters that global methods obscure. Isomap [Tenenbaum et al. 2000] was included as a global nonlinear baseline preserving geodesic rather than Euclidean distances. Finally, LLE [Roweis and Saul 2000] was included as a complementary local nonlinear method to contrast with Isomap's global geodesic approach.

3.5. Evaluation Protocol and Sampling Strategy

The baseline model selection for the first chronological period (1998–2004) was

[2] Platform repository: https://github.com/joao-victor-costa-gomes/ReduzLab.git.

performed through a hybrid criterion combining: (i) quantitative clustering metrics (Silhouette Score and Davies-Bouldin Index) calculated on projected representations; and (ii) structured visual inspection assessing the degree of class separation, absence of excessive overlap, and intraclass cohesion. The algorithm that performed best on both criteria in the reference period was then applied to the two remaining periods to track the temporal evolution of discriminant structure.

Given the dataset size, stratified sampling was applied to preserve the geographic representativeness of ocean basins (Ocean_Name) while managing each algorithm's computational complexity, as detailed in Table 1.

Table 1. Sampling and algorithmic complexity in the platform

Algorithms	Sampling	Complexity	Technical Justify / Limiter
PCA e LDA	100%	$O(d^2\, n)$	Low computational cost and full use of the dataset to maximize the quality of projections.
t-SNE	50%	$O(n\, log\, n)$	Avoid overplotting and optimize gradient.
NCA	40%	$O(d\, n^2)$	Avoid timeouts in a web environment.
KPCA, LLE e ISOMAP	30%	$O(n^3)$	Prevent RAM memory saturation.

4. Experiments and Results

This section presents the results obtained with the platform, including the comparison among seven algorithms that led to the selection of LDA as the baseline model. The projections of the three phases (1998–2019) are analyzed, relating the reduction of discriminant variance to the climatic convergence of ocean basins and to changes in thermal stress of the Red Sea and Arabian Gulf.

4.1. Baseline Dimensionality Reduction Model Selection

The dataset corresponding to the first chronological phase (1998–2004) was submitted to all seven algorithms integrated into the platform: PCA, t-SNE, LDA, NCA, LLE, Isomap, and KPCA. The baseline model selection followed a hybrid criterion combining quantitative clustering metrics and structured visual inspection. The visual protocol assessed: (i) the degree of separation between clusters corresponding to ocean basins; (ii) the absence of excessive overlap between classes; and (iii) the internal cohesion of identified clusters.

It is methodologically important to note that LDA occupies a distinct role among the evaluated algorithms: as a supervised technique, it explicitly uses the target variable (Ocean_Name) to calculate discriminant axes that maximize between-class separation and minimize within-class dispersion. Unlike unsupervised methods, LDA does not discover clusters independently; it projects data into a space optimized for predefined class separability. Its inclusion serves a specific diagnostic purpose: determining whether there is a linear discriminant structure between ocean basins in the thermal stress data and verifying whether this structure remains stable over time. It is worth noting that LDA's diagnostic effectiveness in this study is contingent on the linearly separable discriminant structure present in the BCO-DMO dataset; datasets with more complex, nonlinear basin differentiation may benefit from unsupervised or kernel-based approaches.

When analyzing the two-dimensional projections of unsupervised algorithms, none achieved satisfactory class separation for the 1998–2004 period. The data appeared as highly overlapping point clouds, indicating that global climatic variance obscured subtle regional distinctions from the perspective of purely topological or global variance maximization models. Figures 1 and 2 exemplify this absence of structural segregation, even in algorithms designed to capture nonlinear complexities. PCA, in particular, captured only a limited fraction of variance in the first two components, reflecting the intrinsically multivariate and nonlinear nature of oceanic thermal stress.

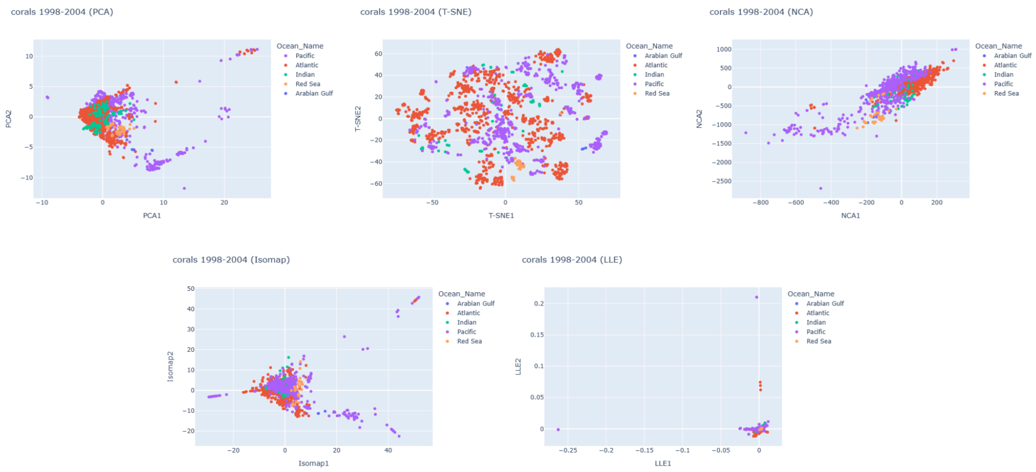


Figure 1. Visual projections generated by unsupervised and iterative algorithms (PCA, t-SNE, NCA, Isomap, and LLE, respectively).

Figure 2. Comparative visualization of the Kernel PCA (KPCA) algorithm and its different kernel functions (Cosine, RBF, Sigmoid, and Poly, respectively).

The LDA projection for this period, presented in Figure 3, clearly isolated the Arabian Gulf and the Red Sea in the lower portion of the vertical axis (LDA2), while establishing a perceptible boundary on the horizontal axis (LDA1): the Pacific Ocean concentrated mainly in the left quadrant, the Atlantic Ocean grouped to the right, and the Indian Ocean positioned in the lower central region.

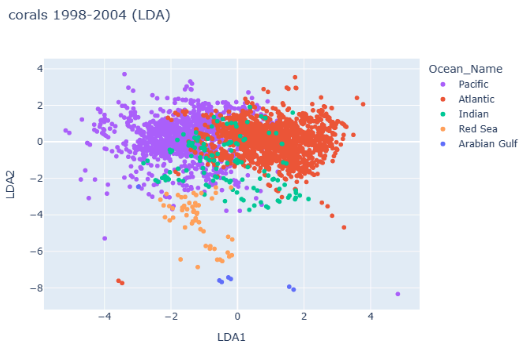
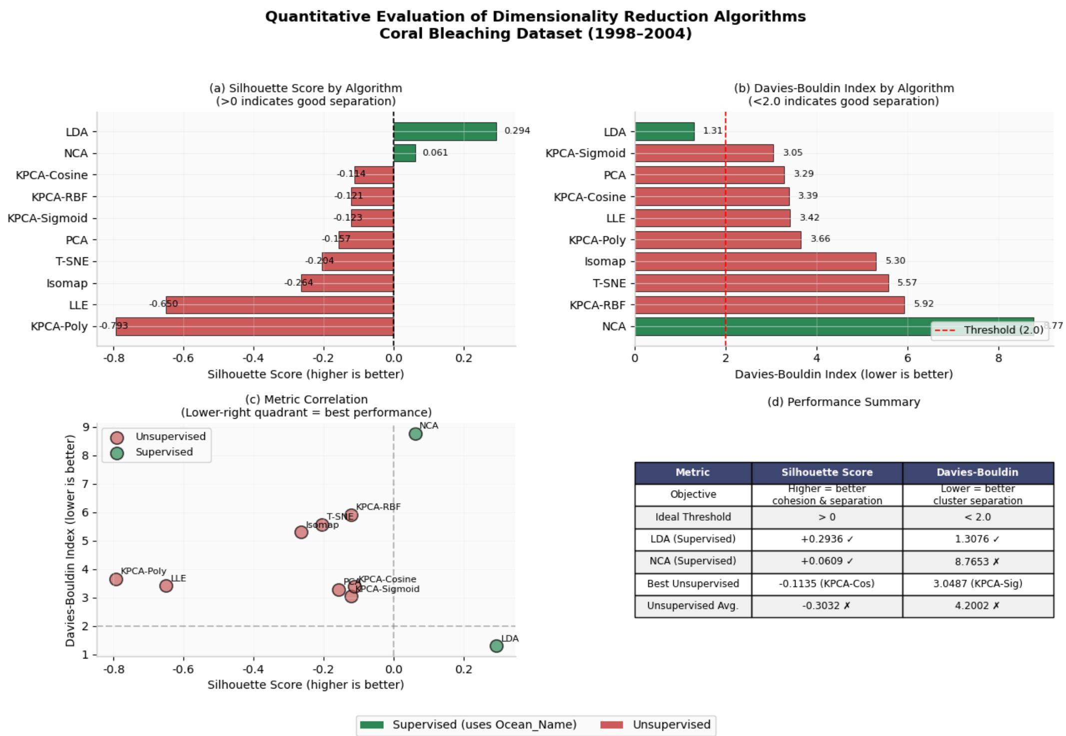


Figure 3. Two-dimensional projection generated by the LDA algorithm applied to the baseline period dataset (1998–2004).

To complement the visual evaluation, quantitative clustering metrics were calculated for all algorithms. LDA achieved a Silhouette Index of 0.2936 and a Davies-Bouldin Index of 1.3076 (Figures 4a and 4b), being the only algorithm to satisfy both acceptability thresholds (Silhouette > 0 and Davies-Bouldin < 2.0). In contrast, all unsupervised methods presented negative Silhouette Indices (mean: –0.30), indicating low cluster cohesion, and Davies-Bouldin indices consistently above 3.0 (mean: 4.20), evidencing high dispersion and overlap between ocean basins. The scatter plot (Figure 4c) positions LDA in the ideal lower-right quadrant, while the summary table (Figure 4d) confirms that no unsupervised algorithm approached acceptable separability thresholds. These results quantitatively establish LDA as the appropriate baseline model for comparative temporal analysis, confirming that there is a significant discriminant structure between ocean basins in the thermal stress data.



Metric	Silhouette Score	Davies-Bouldin
Objective	Higher = better cohesion & separation	Lower = better cluster separation
Ideal Threshold	> 0	< 2.0
LDA (Supervised)	+0.2936 ✓	1.3076 ✓
NCA (Supervised)	+0.0609 ✓	8.7653 ✗
Best Unsupervised	-0.1135 (KPCA-Cos)	3.0487 (KPCA-Sig)
Unsupervised Avg.	-0.3032 ✗	4.2002 ✗

Figure 4. Quantitative evaluation of dimensionality reduction algorithms (a) Silhouette Score; (b) Davies-Bouldin Index; (c) correlation between metrics); (d) performance summary comparing supervised and unsupervised methods.

4.2. Comparative Analysis of Climatic Evolution (1998–2019)

The LDA was applied to the datasets of the two subsequent periods to examine the temporal evolution of discriminant structure. A relevant methodological consideration must be made before interpreting the projections: LDA discriminant axes are recalculated independently for each period based on that period's covariance structure, meaning that projections do not share a common coordinate system. Comparison between periods is therefore necessarily qualitative, based on changes in separability, cohesion, and relative cluster positioning, rather than absolute coordinate values.

The projection corresponding to the 2005–2011 period, as seen in Figure 5a, reveals the first signs of structural transition. The major open ocean basins (Atlantic, Pacific, and Indian) begin to show greater densification in their central core, evidencing a progressive fusion of their climatic boundaries. Notably, the Red Sea loses the visual isolation it presented in the previous decade, integrating into this central grouping, while the Arabian Gulf remains largely isolated in the lower portion of the vertical axis (LDA2).

The projection for the most recent period (2012–2019), presented in Figure 5b, reveals an abrupt morphological transformation. The Pacific, Atlantic, and Indian Oceans now form an almost total overlap, constituting a single massive cluster and signaling a global homogenization of thermal stress. The most critical analytical observation is the inversion of relative positioning of both the Arabian Gulf and the Red Sea: both ecosystems, which historically occupied the lower limit of the discriminant space, abruptly shift to the upper positive quadrants, a spatial displacement opposite to their baseline position.

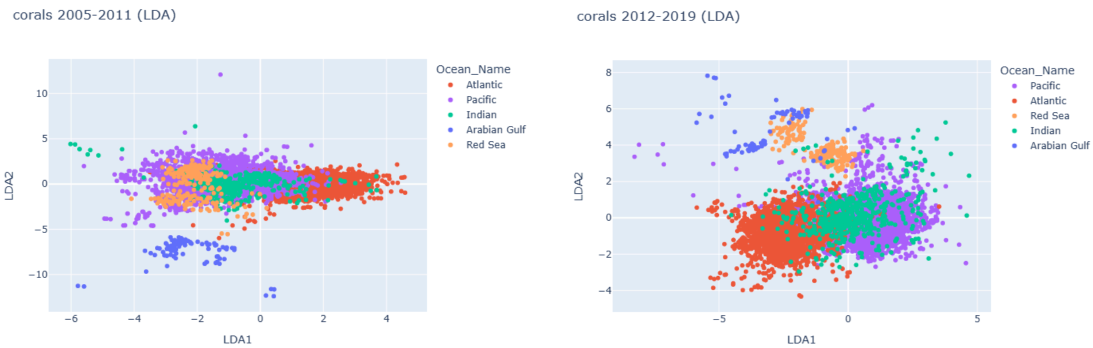


Figure 5. (a) LDA projection of the intermediate period (2005–2011) (b) LDA projection of the recent period (2012–2019).

The inversion of positioning of these enclosed marine ecosystems, progressing from the lower discriminant region through the central clustering of major open oceans and finally surpassing it in the upper quadrants, is consistent with the hypothesis that thermal stress variables in these basins progressively approached and then exceeded the profile of major open ocean basins.

4.3. Quantitative Validation: Degradation of Explained Variance

To ground visual observations in quantitative evidence, two independent analytical approaches were applied: examination of the explained variance metric by LDA across periods and direct inspection of raw predictor variables with the largest absolute loadings on the discriminant axes.

In the context of LDA, explained variance represents the proportion of interclass variance captured by the discriminant vectors retained in the two-dimensional projection, distinct from the total data variance measured by PCA. Table 2 shows a continuous degradation of this metric across the three periods: from 85.26% in 1998–2004 to 83.12% in 2005–2011, dropping sharply to 70.07% in 2012–2019. This drop of more than 15 percentage points indicates that the linear separability structure between ocean basins became progressively less pronounced, suggesting that climatic boundaries between oceanic regions became less defined over time, consistent with visual observations of increased cluster overlap among major open ocean basins.

Table 2. Evolution of LDA model explained variance by period

Chronological Period	Analyzed Records	Explained Variance	Observed Visual Impact
1998–2004	6.029	85,26%	Sharp regional separation
2005–2011	14.909	83,12%	Beginning of climatic convergence
2012–2019	11.669	70,07%	Homogenization and regime inversion

To directly validate whether the Arabian Gulf's discriminant shift reflects real changes in underlying data rather than axis recalibration artifacts, predictor variables with the largest absolute loadings on the LDA axes were examined. These coefficients revealed the predominance of Thermal Stress Anomaly (TSA) metrics, particularly TSA_Maximum, which measures the peak extreme thermal stress reached in a region, and TSA_Mean, which represents the continuous average of thermal anomalies experienced by a marine ecosystem. The arithmetic means of these variables were calculated by ocean basin in each chronological phase and are presented in Figure 6.

Figure 6a shows the TSA_Maximum trajectory across the three periods for all basins. While the Pacific, Atlantic, and Indian Oceans recorded reduction or stabilization in thermal stress peaks, with means moving from approximately 3.1 to 2.6, the Arabian Gulf showed a divergent trajectory, increasing from 2.86 in 1998–2004 to 3.78 in 2005–2011. Figure 6c confirms this trend when comparing the initial period with the most recent, in which the Arabian Gulf reached 3.72 in 2012–2019, representing a 30% increase from baseline, while other basins stabilized or reduced their values. Figure 6b complements the analysis with TSA_Mean, indicating that the Arabian Gulf evolved from -5.44 to -4.34 and -4.43, suggesting reduction of chronic negative anomaly and transition to more extreme and unstable thermal conditions, while other basins maintained relatively stable profiles or showed discrete improvements. Finally, Figure 6d synthesizes the percentage variations, evidencing that the Arabian Gulf is the only basin with simultaneous intensification of major thermal stress indicators, with a 30.1% increase in TSA_Maximum and 18.6% in TSA_Mean.

The convergence between the Arabian Gulf's mathematical displacement in LDA discriminant space and the 30% increase in TSA_Maximum documented in raw data constitutes the main finding of this analysis: supervised DR effectively detected and visualized a genuine and quantifiable escalation in regional thermal stress, independently confirmed through direct variable inspection. This result demonstrates both the diagnostic utility of LDA-based temporal projection and the anomalous trajectory of enclosed marine ecosystems in a context of global thermal stress peaks that would otherwise stabilize.

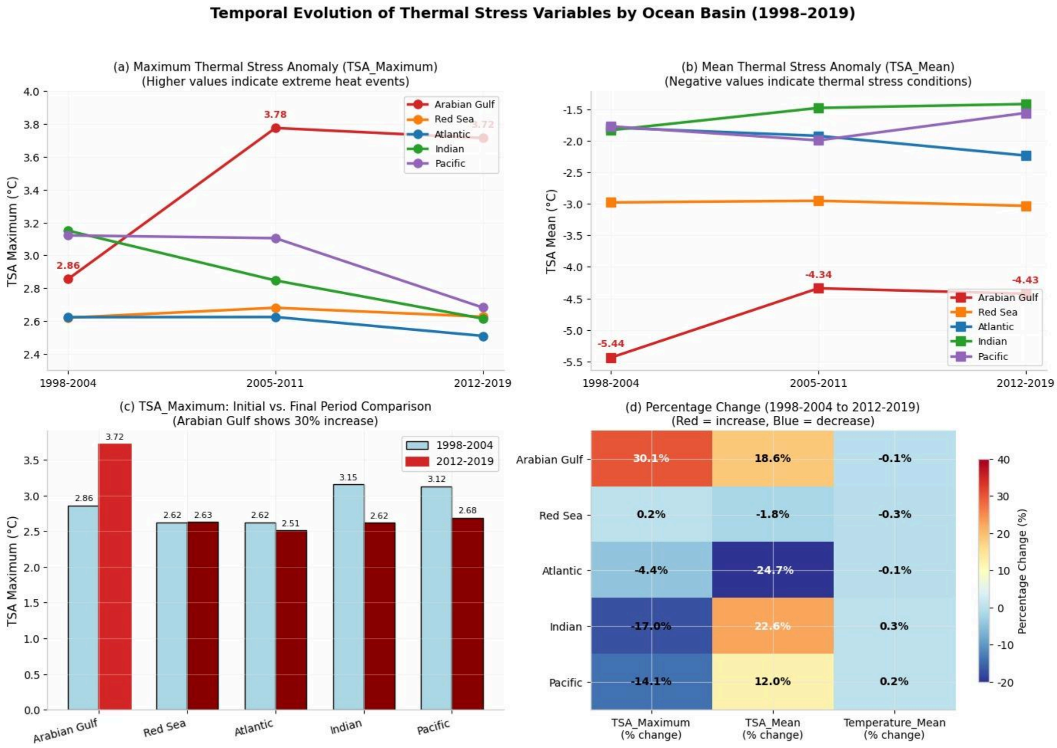


Figure 6. Temporal evolution of key thermal stress variables by ocean basin (1998–2019): (a) TSA_Maximum per basin across three chronological periods; (b) TSA_Mean per basin across three periods; (c) TSA_Maximum comparison between baseline (1998–2004) and recent period (2012–2019); (d) percentage change in TSA_Maximum, TSA_Mean, and Temperature_Mean from 1998–2004 to 2012–2019.

It is worth noting that the dataset covers the period up to 2020, with no records from 2021 onwards at the time of analysis. Given the progressive degradation of discriminant variance between classes and the documented escalation in the Arabian Gulf, it is reasonable to hypothesize that subsequent years, if incorporated, would present even more extreme values, with potentially severe consequences for coral reef health in these regions.

5. Conclusion

This paper presented a systematic comparative evaluation of seven dimensionality reduction techniques applied to the BCO-DMO global coral bleaching dataset (1998–2019), addressing the question of which methods best reveal the discriminant structure between ocean basins in high-dimensional thermal stress data and whether this structure changed over time.

Among the evaluated algorithms, LDA proved most effective for supervised discriminant visualization in the reference period, revealing class structure that unsupervised methods could not cohesively separate. This result reflects the linear discriminant structure present in this specific dataset, and does not imply that LDA is the optimal DR technique for thermal stress analysis in general. Its application across three phases revealed two main patterns: the progressive homogenization of thermal stress among the Pacific, Atlantic, and Indian Oceans, and an anomalous regime shift in the Arabian Gulf, diverging from the stabilization observed in other basins and confirmed by raw variable analysis. These results have implications for environmental monitoring, as the adopted workflow, combining supervised analysis, quantitative metrics, and direct validation, offers a replicable framework for detecting multivariate changes in large oceanographic datasets. The platform amplifies this potential by enabling this process in an accessible manner, even for researchers without programming experience.

Methodological limitations include the qualitative nature of LDA comparison between periods due to independent axis recalibration, variation in sample composition across periods, and dataset coverage ending in 2020. Future work should address these issues through: projection of all periods into a fixed LDA space, trained on the reference period, enabling geometrically valid temporal comparison; incorporation of post-2020 data to verify whether the Arabian Gulf trajectory continued; integration of UMAP as an additional nonlinear baseline; and validation of findings through independent oceanographic datasets and expert domain review.

6. Acknowledgments

The authors would like to thank the Coordination for the Improvement of Higher Education Personnel (CAPES) for the financial support provided through the Master's scholarship associated with the Graduate Program in Computer Science (PPGCC – Master's Degree), offered jointly by UFERSA and UERN.

References

Badshah, A., Daud, A., Alharbey, R. et al. (2024) "Big data applications: overview, challenges and future", Artificial Intelligence Review, 57(290).

Chaidez, V., Dreano, D., Agusti, S., Duarte, C. M. and Hoteit, I. (2017) "Decadal trends in Red Sea maximum surface temperature", Scientific Reports, 7(1), p. 8144.

Davies, D. L. and Bouldin, D. W. (1979) "A cluster separation measure", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1(2), p. 224-227.

Goldberger, J., Roweis, S., Hinton, G. and Salakhutdinov, R. (2004) "Neighbourhood components analysis", Advances in Neural Information Processing Systems, 17.

Hamdan, A., Ibekwe, K. I., Etukudoh, E. A., Umoh, A. A. and Ilojianya, V. I. (2024) "AI and machine learning in climate change research: A review of predictive models and environmental impact", World Journal of Advanced Research and Reviews, 21(1), p. 1999-2008.

Hassani, H., Huang, X. and Silva, E. (2019) "Big data and climate change", Big Data and Cognitive Computing, 3(1), p. 12.

Hughes, T. P., Kerry, J. T., Álvarez-Noriega, M., Álvarez-Romero, J. G., Anderson, K.

D., Baird, A. H., Babcock, R. C., Beger, M., Bellwood, D. R., Berkelmans, R. et al. (2018) “Spatial and temporal patterns of mass bleaching of corals in the Anthropocene”, Science, 359(6371), p. 80-83.

Jeon, H. et al. (2025) “Unveiling high-dimensional backstage: a survey for reliable visual analytics with dimensionality reduction”, In: CHI Conference on Human Factors in Computing Systems, ACM, New York, p. 1-24.

Jolliffe, I. T. and Cadima, J. (2016) “Principal component analysis: a review and recent developments”, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2065), p. 20150202.

Lopes, M. A. S., Dória Neto, A. D. and Martins, A. M. (2020) “Parallel t-SNE Applied to Data Visualization in Smart Cities”, IEEE Access, 8, p. 11482-11490.

Martinez, A. M. and Kak, A. C. (2001) “PCA versus LDA”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(2), p. 228-233.

Montambault, B., Appleby, G., Rogers, J., Brumar, C. D., Li, M. and Chang, R. (2024) “DimBridge: Interactive Explanation of Visual Patterns in Dimensionality Reductions with Predicate Logic”, IEEE Transactions on Visualization and Computer Graphics, 31(1), p. 207-217.

Rousseeuw, P. J. (1987) “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”, Journal of Computational and Applied Mathematics, 20, p. 53-65.

Schölkopf, B., Smola, A. and Müller, K.-R. (1998) “Nonlinear component analysis as a kernel eigenvalue problem”, Neural Computation, 10(5), p. 1299-1319.

Shah, S. and Joshi, S. (2021) “Study of various dimensionality reduction and classification algorithms on high dimensional dataset”, In: International Conference on Inventive Research in Computing Applications, IEEE, Coimbatore, p. 1005-1010.

Tenenbaum, J. B., Silva, V. and Langford, J. C. (2000) “A global geometric framework for nonlinear dimensionality reduction”, Science, 290(5500), p. 2319-2323.

Van der Maaten, L. and Hinton, G. (2008) “Visualizing data using t-SNE”, Journal of Machine Learning Research, 9(11).

Van Woesik, R. and Kratochwill, C. (2022) “A global coral-bleaching database, 1980–2020”, Scientific Data, 9, p. 20. doi: 10.1038/s41597-022-01121-y.