

# Information-Theoretic Quality Metric of Low-Dimensional Embeddings

Sebastián Gutiérrez-Bernal<sup>1\*</sup>, Hector Medel Cobaxin<sup>1\*</sup>  
and Abiel Galindo González<sup>1</sup>

<sup>1\*</sup>Tecnologico de Monterrey, Escuela de Ingenieria y Ciencias, Ave.  
Eugenio Garza Sada 2501 Sur, Col: Tecnologico, Monterrey, N.L.,  
Mexico, 64700.

\*Corresponding author(s). E-mail(s): [sebastian.gtz.brnl@tec.mx](mailto:sebastian.gtz.brnl@tec.mx);  
[hmedel@tec.mx](mailto:hmedel@tec.mx);

Contributing authors: [A00836342@tec.mx](mailto:A00836342@tec.mx);

## Abstract

In this work we study the quality of low-dimensional embeddings from an explicitly information-theoretic perspective. We begin by noting that classical evaluation metrics such as stress, rank-based neighborhood criteria, or Local Procrustes quantify distortions in distances or in local geometries, but do not directly assess how much information is preserved when projecting high-dimensional data onto a lower-dimensional space. To address this limitation, we introduce the Entropy Rank Preservation Measure (ERPM), a local metric based on the Shannon entropy of the singular-value spectrum of neighborhood matrices and on the stable rank, which quantifies changes in uncertainty between the original representation and its reduced projection, providing neighborhood-level indicators and a global summary statistic. To validate the results of the metric, we compare its outcomes with the Mean Relative Rank Error (MRRE), which is distance-based, and with Local Procrustes, which is based on geometric properties, using a financial time series and a manifold commonly studied in the literature. We observe that distance-based criteria exhibit very low correlation with geometric and spectral measures, while ERPM and Local Procrustes show strong average correlation but display significant discrepancies in local regimes, leading to the conclusion that ERPM complements existing metrics by identifying neighborhoods with severe information loss, thereby enabling a more comprehensive assessment of embeddings, particularly in information-sensitive applications such as the construction of early-warning indicators.

**Keywords:** Manifold Learning, dimensionality reduction, embedding quality, entropy, singular values

## 1 Introduction

In high-dimensional spaces, data geometry deteriorates, implying that pairwise distances become less informative, neighborhoods degenerate, and many models see their performance collapse as the ambient dimension  $D$  grows, a phenomenon often described as a manifestation of the curse of dimensionality [1]. In consequence, both statistical estimation and learning algorithms, that rely on Euclidean distances or local neighborhood structure, may be affected. This motivates the search for representations in which the relevant structure of the data becomes more regular and accessible.

With these considerations in mind, various studies and proposals have been developed[2], mainly assuming that observations  $x \in \mathbb{R}^D$  are concentrated near a low-dimensional manifold  $\mathcal{M} \subset \mathbb{R}^D$  of intrinsic dimension  $d \ll D$ , and to construct a mapping  $f : \mathbb{R}^D \rightarrow \mathbb{R}^d$  such that preserves salient structural properties of  $\mathcal{M}$ , including connectivity, curvature and intrinsic distances. This manifold-learning perspective underlies a broad family of nonlinear dimensionality-reduction methods, including geodesic-based approaches such as Isomap [3], locally linear methods such as LLE and its variants [4, 5], and spectral or kernel methods such as Laplacian Eigenmaps and Kernel PCA [6, 7].

Therefore, it is extremely necessary to evaluate the quality of these low-dimensional representations. A variety of criteria have been proposed for this evaluation[2], such as global measures like Kruskal’s stress[8], Sammon’s stress[9], and Residual Variance[3], which quantify distortions of pairwise distances across the entire dataset. On the other hand, local criteria focus on neighborhood relations and assess changes in nearest neighbors, rank orders, and local shape through indices such as Trustworthiness, Continuity[10], Mean Relative Rank Error[11], and Local Procrustes[12].

However, these existing measures are defined purely in geometric terms and therefore quantify only how distances, ranks or local shapes change under the mappings, without having one that directly assesses how much information is lost or preserved when the projection is generated. This omission is particularly problematic when dimensionality reduction is used as a preprocessing step for sensitive tasks, such as reconstructing attractors from time series embeddings [13–15] or building early-warning indicators for regime shifts in dynamical and financial systems [16]. In such settings, the informational content carried by the embedding is as critical as its geometric fidelity, in recent work on stable rank and geometry-preserving embeddings[17] suggests that spectral properties of the data matrix can play a central role in capturing this content.

This work is motivated precisely by that gap. Our goal is to complement existing geometric criteria with a quality measure that incorporates an explicitly informational point of view on the effect of dimensionality reduction. By linking local spectral

structure to entropy, we aim to quantify not only how the manifold is geometrically distorted, but also how the uncertainty and effective number of spectral directions change when passing from the original representation to its low-dimensional embedding.

## 2 Case of study: Information-based manifold learning for early warnings

The motivation of this study arises from Huang's *et al.*[16] work, where a Manifold Learning methodology based on information criteria is proposed. This method takes a financial time series to reconstruct its phase space through Takens' theorem[13], building the delay vectors

$$X_t = (x_t, x_{t-\tau}, x_{t-2\tau}, \dots, x_{t-(m-1)\tau}), \quad (1)$$

choosing the delay  $\tau$  as the first minimum of the auto mutual information function [14] and the embedding dimension  $m$  through Cao's method[15], obtaining a matrix  $X \in \mathbb{R}^{n \times m}$  that represents the reconstructed phase space in high dimension.

Their idea centers in an information-theoretic perspective, in which instead of describing neighborhoods through Euclidean distances, each reconstructed state  $X_i$  is modeled as a discrete probability distribution  $p_i$ , estimated through kernel density, where the similarity between states is measured with the symmetrized Kullback–Leibler divergence

$$h(P_i, P_j) = \sum_{t=1}^m p_i^{(t)} \log \frac{p_i^{(t)}}{p_j^{(t)}} + \sum_{t=1}^m p_j^{(t)} \log \frac{p_j^{(t)}}{p_i^{(t)}}, \quad (2)$$

whose values form the matrix  $H$  that replaces the classical geometric neighborhoods.

From this, they construct a Manifold Learning method analogous to LLE[4], but with weights defined by the information metric, in which the low-dimensional embedding  $Y$  is obtained by minimizing

$$\min_Y \sum_{i=1}^n \left\| Y_i - \sum_{j=1}^n h_{ij} Y_j \right\|^2, \quad (3)$$

which reduces to an eigenvalue problem over  $M = (I - H)^\top (I - H)$ , whose  $d$  eigenvectors associated with the smallest eigenvalues form the reduced coordinates.

Finally, in the methodology they quantify the quality of the embedding using the Procrustes measure to compare the high-dimensional representation against the reduced one, using datasets from a stochastic simulation and the financial series of CSI800 and S&P 500, where they show that their information-based method has much lower Procrustes error compared with other relevant methods such as Isomap, LLE, or KPCA.

One of the things we highlight most about their work is its direct application, because once the manifold is reduced in dimensions, through clustering techniques and using Hidden Markov models, they detect regions associated with pre-crisis dynamics, which allows the reduced representation to be used as an early-warning tool. Furthermore, we consider their analysis of associating the curvature of the manifold with a global indicator of the underlying attractor’s rigidity or resilience to be extremely relevant.

However, we observe two relevant methodological limitations. First, the Procrustes metric measures global geometric preservation but is not sensitive to local distortions. Second, although the method is information-based, no metric was introduced to evaluate how well the embedding preserves divergence structure in the original space. Consequently, the comparative evaluation does not fully examine the method along the most relevant theoretical dimensions. This leads us to propose a local metric that can explicitly characterize aspects associated with the loss or preservation of information.

## 3 Background and Related Work

### 3.1 Manifold Learning Methods

Manifold learning methods are commonly categorized according to the type of geometric structure they are designed to preserve. In this work, we adopt the view that, despite differences in implementation, these methods share the common objective of recovering a low-dimensional representation that preserves the intrinsic distances or neighborhood relationships of an underlying manifold embedded in a high-dimensional space. Accordingly, we group them into three dominant paradigms: local methods, global methods, and kernel or spectral methods, each reflecting a distinct interpretation of the manifold hypothesis and of which aspects of the geometry should be preserved. For comprehensive reviews of these approaches, we refer the reader to [3–6].

Local methods assume that we can approximate the manifold by a collection of tangent planes. Given data points  $x_1, \dots, x_n$ , we construct neighborhoods  $N(i)$ , often via  $k$ -nearest neighbors, and impose that the local geometric relations within each neighborhood are preserved after embedding. For example, in Locally Linear Embedding (LLE)[4] we compute reconstruction weights  $w_{ij}$  such that

$$x_i \approx \sum_{j \in N(i)} w_{ij} x_j, \quad \sum_{j \in N(i)} w_{ij} = 1, \quad (4)$$

and then seek a low-dimensional representation  $Y$  satisfying

$$y_i \approx \sum_{j \in N(i)} w_{ij} y_j. \quad (5)$$

By construction, the embedding  $Y$  preserves local linear structure and remains consistent with the manifold’s tangent geometry. Related approaches, such as Local Tangent Space Alignment (LTSA)[5], preserve higher-order local properties by aligning

estimated tangent spaces rather than reconstruction weights. The Hessian Eigenmaps (HLE) [18] goes one step further by estimating a second-order local representation that is sensitive to the curvature of the manifold in each neighborhood and use it to obtain a smoother low-dimensional embedding.

Global methods focus on geodesic distances along the manifold. The most classical is Isomap [3], which approximates intrinsic distances by computing shortest paths on a neighborhood graph using Dijkstra’s algorithm:

$$d_{\mathcal{M}}(x_i, x_j) \approx \text{shortest path in graph } G. \quad (6)$$

We then embed these estimated geodesic distances into a low-dimensional space via classical multidimensional scaling, solving an eigenvalue problem. Because geodesics capture long-range curvature of the manifold, Isomap preserves global structure but may be sensitive to noise or disconnected neighborhoods.

Kernel or spectral methods interpret dimensionality reduction as constructing a nonlinear mapping by specifying a kernel function  $k(x_i, x_j)$  that implicitly embeds the data into a high-dimensional feature space. Kernel PCA [7] finds principal components by solving

$$K\alpha = \lambda\alpha, \quad (7)$$

where  $K_{ij} = k(x_i, x_j)$ . This allows linear methods to operate on nonlinear data by lifting them into a reproducing kernel Hilbert space. Closely related are graph Laplacian methods such as Laplacian Eigenmaps [6], which build a weighted adjacency matrix  $W$  and compute eigenvectors of the Laplacian

$$L = D - W. \quad (8)$$

These methods preserve diffusion geometry and are grounded in connections between discrete Laplacians and the Laplace-Beltrami operator on smooth manifolds.

Together, these families of methods illustrate the spectrum of approaches to approximating the geometry of an unknown manifold; from enforcing consistency with local linear approximations, to reconstructing global geodesic structure, to expressing the manifold through kernels that encode nonlinear similarity. In all cases, dimensionality reduction is achieved by solving constrained optimization or spectral problems designed to preserve intrinsic manifold relationships rather than ambient Euclidean distances.

### 3.2 Quality Metrics

To quantify how well the quality of an *embedding* is preserved after dimensionality reduction, it is natural to interpret such quality in terms of the geometric structure that is lost when we apply this mapping. In the work of Gracia *et al.*[2], a robust analysis of the state of the art of these metrics is presented, where it is shown, in a general manner, that most of them can be classified into local or global criteria.

From the global perspective, there exist measures such as Kruskal’s stress[8], Sammon stress[9], or the residual variance[3], which indicate how well pairwise distances

are preserved throughout the entire dataset. Meanwhile, metrics based on local criteria focus on neighborhood structure; some aim to evaluate the appearance and disappearance of neighbors, as well as changes in their relative order, such as Trustworthiness and Continuity[10] or the Mean Relative Rank Error[11]; and others compare geometrically the shape of each neighborhood, such as Local Procrustes[12].

A common feature of these criteria is that, despite their differences, they all assess the quality of the embedding through the geometric distortions introduced by dimensionality reduction, either through global distances or through changes in neighborhood structure. However, and considering the analysis in the case study of Section 2, we observe that these metrics do not allow us to directly evaluate how much information is retained throughout the process, which motivates the need to develop an additional criterion that explicitly incorporates this dimension of the analysis.

In what follows, we focus on two representative local metrics that will serve as benchmarks in Section 6: the Mean Relative Rank Error (MRRE) and the Local Procrustes statistic.

MRRE[11] adopts a local distance-preservation approach, constructing a co-ranking matrix  $\mathbf{Q} = q_{kl}$  that summarizes how distance ranks change between the original space and the low-dimensional embedding.

From  $\mathbf{Q}$  different types of neighborhood errors are identified according to the blocks defined in the Lee’s work:

- The block  $\mathbb{LL}_k$  counts the *hard intrusions*, which are points that did not belong to the neighborhood in high dimension but enter it in the low-dimensional embedding.
- The block  $\mathbb{UR}_k$  counts the *hard extrusions*, which are neighbors that existed in high dimension but cease to be neighbors in the low-dimensional embedding.
- And the block  $\mathbb{UL}_k$  groups the *mild intrusions* and *mild extrusions*, which are smaller rank changes in which, despite shifting, the points remain within the neighborhood.

This metric penalizes these errors through the following expressions:

$$W_n(K) = \frac{1}{H_K} \sum_{(k,l) \in \mathbb{UL}_k \cup \mathbb{LL}_k} \frac{|k-l|}{l} q_{kl}, \quad W_v(K) = \frac{1}{H_K} \sum_{(k,l) \in \mathbb{UL}_k \cup \mathbb{LL}_k} \frac{|k-l|}{l} q_{kl}, \quad (9)$$

where

$$H_K = N \sum_{k=1}^K \frac{|N - 2k + 1|}{k} \quad (10)$$

is a normalization factor.

The Local Procrustes measure[12], in turn, indicates the degree of geometric preservation for each neighborhood, where centered neighbor matrices are used to remove

translation effects:

$$\tilde{X}_i = (I - \frac{1}{k}\mathbf{1}\mathbf{1}^\top)X_i, \quad \tilde{Y}_i = (I - \frac{1}{k}\mathbf{1}\mathbf{1}^\top)Y_i, \quad (11)$$

and, with the SVD of  $\tilde{X}_i^\top \tilde{Y}_i = U_i S_i V_i^\top$ , a rotation matrix and a scaling factor are obtained:

$$A_i = U_i V_i^\top, \quad c_i = \frac{\text{tr}(S_i)}{\text{tr}(\tilde{Y}_i^\top \tilde{Y}_i)}, \quad (12)$$

with the purpose of aligning  $\tilde{Y}_i$  and  $\tilde{X}_i$  in the most optimal way. With this, the Procrustes statistic in its normalized conformal version is constructed:

$$R_{C,N}(X, Y) = \frac{1}{n} \sum_{i=1}^n \frac{G_{C_i}(X, Y)}{\|\tilde{X}_i\|_F^2}, \quad G_{C_i}(X, Y) = \|\tilde{X}_i - \tilde{Y}_i (c_i A_i^\top)\|_F^2, \quad (13)$$

which averages the quality of the local adjustments across all neighborhoods.

### 3.3 Information in terms of spectrum

Shannon [19] introduces the notion of information as a measure of the uncertainty associated with the occurrence of an event prior to observing it, which we can more intuitively relate to the average amount of information we need to receive in order to know the state of the system.

Mathematically, this average uncertainty is defined as the entropy of a discrete probability distribution

$$H = - \sum_j p_j \log p_j, \quad (14)$$

which measures the dispersion of the values  $p_j$  and therefore *the degree of lack of knowledge* about the system.

In the context of comparing an object before and after subjecting it to a dimensionality reduction process, we see the need to measure not only how much distance was preserved but also how much of this uncertainty was modified or lost, because we cannot naturally guarantee that preserving distances also preserves information.

As we know, in multivariate data, the eigenvalues of the covariance matrix measure variance along principal directions. These eigenvalues are proportional to the squared singular values of the centered data matrix, so the singular-value spectrum provides a compact summary of how variance is distributed across the dataset. We therefore define the entropy of a matrix  $M$  by normalizing the squared singular values to obtain a probability distribution  $p_j$  and evaluating the Shannon entropy in (14); quantifying how evenly the variance is spread across spectral directions: high values indicate that many directions contribute comparably, whereas low values indicate that only a few directions dominate. Consequently, when we apply a dimensionality-reduction

method that eliminates degrees of freedom or concentrates variance into fewer directions, we observe a change in spectral entropy, which we can interpret as a variation of uncertainty with respect to the original representation.

To illustrate this idea, let's consider a matrix  $M \in \mathbb{R}^{d \times n}$  of algebraic rank  $r_a = 3$ , whose variance is distributed across three orthogonal singular directions. If we apply a dimensionality-reduction method that produces a new representation  $\tilde{M} \in \mathbb{R}^{2 \times n}$ , then, by definition  $r_a(\tilde{M}) \leq 2$ , implying that one of the three original spectral directions is lost or combined with the remaining ones, concentrating variance into fewer directions. In this sense, we can interpret the loss as a reduction in entropy, meaning a loss of information content of  $\tilde{M}$  with respect to  $M$ .

### 3.3.1 Stable Rank as a measure of spectrum

The algebraic *rank* of a matrix  $M \in \mathbb{R}^{m \times n}$  is defined as the number of non-zero singular values, and it measures the dimension of the subspace spanned by its columns (i.e., the number of linearly independent directions contained in  $M$ ). However, this rank depends in a non-continuous way on the entries of the matrix, implying that an arbitrarily small perturbation may change the number of non-zero singular values[20], making it somewhat unstable in practice.

In order to obtain an indicator that is continuous, differentiable, invariant under scaling, and robust to perturbations, the *stable rank* is introduced[21]

$$r(M) = \frac{\|M\|_F^2}{\|M\|^2} = \frac{\sum_j \sigma_j(M)^2}{\sigma_1(M)^2}, \quad (15)$$

where  $\sigma_j(M)$  is the  $j$ -th singular value of  $M$ , and  $1 \leq r(M) \leq \text{rank}(M)$ .

Additionally, we consider the proposal of Eftekhari *et al.*[17], in which a condition is established to determine when the *delay-coordinate map* induces a *stable embedding* of the attractor using (15). From this perspective, stability is understood as the approximate preservation of distances between points on the attractor, leading to the conclusion that such stability requires the stable rank of the system to be proportional to the dimension of the attractor.

That said, if there exists a criterion that guarantees when a mapping is a stable embedding, then we consider that there must be an analogous notion of a *stable projection* between two representations, one that, in the context of dimensionality reduction, is capable of quantifying the extent to which some information parameter is preserved after the reduction.

## 4 Entropy Rank Preservation Measure (ERPM)

Starting from a matrix  $M \in \mathbb{R}^{d \times n}$  with singular values  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{r_a} > 0$  and algebraic rank  $r_a$ , we associate to its spectrum a probability distribution

$$p_j = \frac{\sigma_j^2}{\sum_{m=1}^{r_a} \sigma_m^2}, \quad (16)$$



which describes the variation that exists along the spectral directions of  $M$ . We conveniently introduce a scaling parameter  $\alpha_j = \sigma_j^2 / \sigma_1^2$ , which satisfies

$$\sum_{j=1}^r \alpha_j = \frac{\sum_{j=1}^r \sigma_j^2}{\sigma_1^2} = r(M), \quad (17)$$

providing an alternative form of writing the stable rank defined in (15). With this, it is evident that the expression (16) reduces to  $p_j = \alpha_j / r(M)$ , and that if we introduce these probabilities into (14), we obtain an entropy expression in terms of the stable rank:

$$H(M) = \log r(M) - \varepsilon(M), \quad (18)$$

where

$$\varepsilon(M) = \frac{1}{r(M)} \sum_{j=1}^r \alpha_j \log \alpha_j \quad (19)$$

is the term that reflects the internal distribution within the spectrum of matrix  $M$ .

Now, let  $X \in \mathbb{R}^{d \times n}$  be the high-dimensional embedding with  $n$  observations, and  $Y \in \mathbb{R}^{l \times n}$  its reduced representation. We define  $\bar{X}_i = X_i - \frac{1}{k} X_i \mathbf{1} \mathbf{1}^\top$  and  $X_i = [x_{i,j}]_{j=1}^k \in \mathbb{R}^{d \times k}$  as the centered matrix of the  $k$  nearest neighbors for a point  $x_i \in X$ , constructing analogously the matrix  $\bar{Y}_i$  in low dimension from the same set of neighbor indices.

We substitute these matrices into (18), obtaining uncertainty measures associated with the geometry of the neighborhood before and after dimensionality reduction. But, in order to quantify the change in the information of these representations, we compute the entropy difference, obtaining the *Entropy Rank Preservation Measure*:

$$\Delta H_i = H(\bar{Y}_i) - H(\bar{X}_i) = \log \left( \frac{r(\bar{Y}_i)}{r(\bar{X}_i)} \right) + \varepsilon(\bar{X}_i) - \varepsilon(\bar{Y}_i). \quad (20)$$

It is important to highlight that, in general, any system will have an intrinsic indicator of uncertainty completely determined by its singular values. Therefore, we are interested in measuring the *change* in this uncertainty, considering that even if the initial embedding has high or low entropy, what matters is that the variation  $\Delta H_i$  is as close to zero as possible to consider that the projection generated by the reduction method is stable and shows practically no information loss.

Finally, although the definition (20) is a local measure associated with each neighborhood, we can also define a global quantity that describes how much spectral information was modified throughout the embedding when the dimension was reduced.

For this, we average the entropy change over all points:

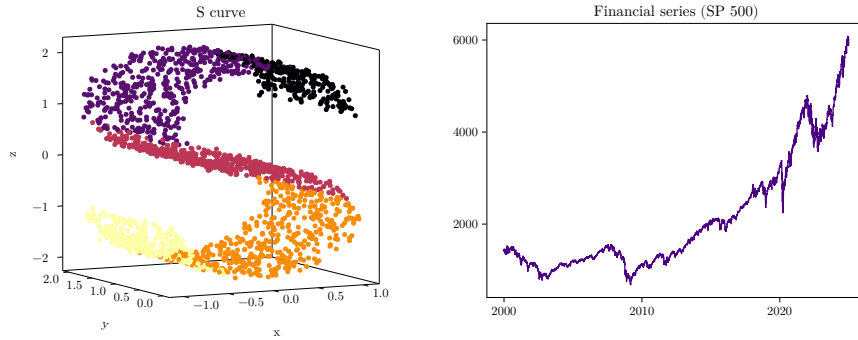
$$R_{\Delta H} = \frac{1}{n} \sum_{i=1}^n \Delta H_i, \quad (21)$$

where positive values of this parameter indicate a global increase in data dispersion, negative values indicate a net loss of information by reducing the effective number of significant spectral directions, and values close to zero suggest a dimensionality reduction that preserves, on average, the spectral structure of the system. Complementarily, it is also necessary to analyze the distribution of the values  $\Delta H_i$ , as a favorable global average of  $R_{\Delta H}$  may arise from a highly skewed or variable set of local behaviors. Such heterogeneity may produce neighborhoods in which the entropy changes sharply, despite the global measure suggesting that the overall distortion is relatively mild.

## 5 Experimental Setup

### 5.1 Data, reduction methods, benchmarks and hyperparameters

We chose to employ two datasets, shown in Fig. 1, in order to evaluate the behavior of the metrics under different scenarios. First, we used the S-curve with 2,000 points, a dataset widely studied in the Manifold Learning literature due to its nonlinear structure and its usefulness as a benchmark for dimensionality-reduction methods. Second, with the purpose of analyzing the information-theoretic aspects discussed in Section 2, we worked with the S&P 500 financial series, collecting daily closing data from January 1, 2000 to January 1, 2025, from which we generated a Takens embedding following the same procedure as in the case study to estimate the optimal delay and dimension, obtaining values of 57 and 6, respectively.



**Fig. 1** Visualization of the data used in the metric evaluation.

Regarding the dimensionality-reduction methods, we employed LLE, HLLE, Isomap, and KPCA with quadratic kernel, as described in Section 3.1. We also used

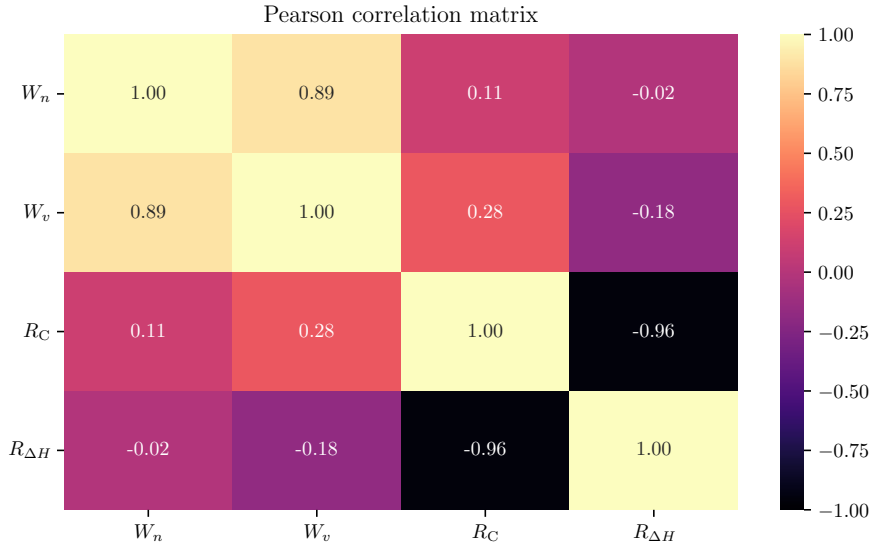
PCA as a linear baseline, which orders orthogonal directions according to their variance, discards those associated with the smallest variance, and linearly projects the data into a lower-dimensional space. To benchmark the ERPM introduced in Section 4, we used the MRRE and Local Procrustes measures described in Section 3.2, which allowed us to compare the resulting embeddings from complementary perspectives.

With respect to the hyperparameters associated with the local character of the metrics, we analyzed their behavior as a function of the number of neighbors  $k$  by performing a sweep over the interval  $k \in [1, 20]$ . For the dimensionality-reduction methods that explicitly depend on this parameter, we used the value  $k = k_{\max}$ , thus ensuring a consistent comparison across techniques. Finally, in the last analysis of the following section, we fixed  $k = 15$  based on the onset of the stabilization period, whose analysis we discuss in greater detail later.

## 6 Results

### 6.1 Global Correlation with Benchmarks

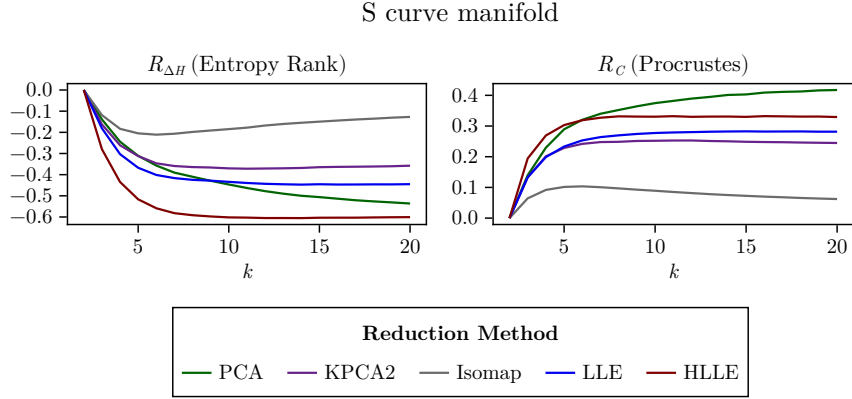
We begin by analyzing the relationship between the ERPM and the benchmark metrics based on the correlation plot in Fig. 2, where we observe that there is practically no correlation between distance-based metrics and metrics based on geometric and spectral properties, since both  $W_n$  and  $W_v$  exhibit extremely low correlation coefficients with  $R_C$  and  $R_{\Delta H}$ , which suggests that preserving distances does not directly imply preserving the geometric structure nor the information content of the embedding. Therefore, if we want a more robust analysis of projection quality, we must treat distance metrics and spectral metrics as complementary indicators.



**Fig. 2** Correlation plot between quality metrics.

On the other hand, we observe that the correlation between  $R_C$  and  $R_{\Delta H}$  is extremely high in absolute terms (the negative sign is due to their respective value ranges), which, from a relational perspective implies that both indicators consistently display the same type of local degradation in the embedded representation when viewed through their global measures.

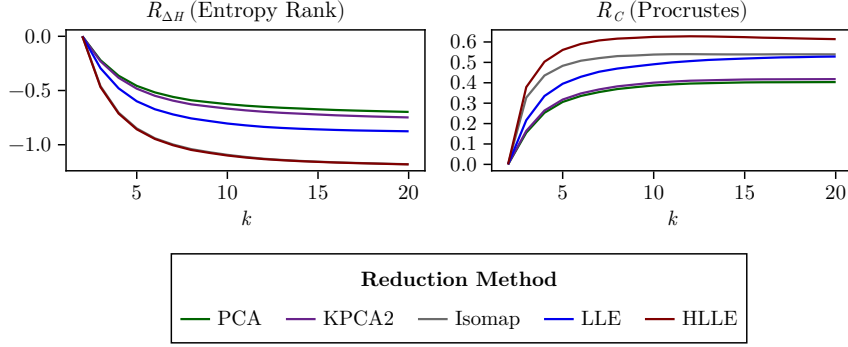
Considering this, and in order to further examine the correlation between both metrics, in Fig. 3 we analyze in a more granular manner the behavior of  $R_{\Delta H}$  and  $R_C$  for the S-curve dataset, where we first observe that both curves display a relatively sharp drop for small values of  $k$ , indicating a certain degree of local instability in the initial neighborhoods. Although the behavior becomes more stable as  $k$  increases, global performance continues deteriorating for all reduction methods except Isomap. Furthermore, the ERPM tends to take more extreme values than Local Procrustes, showing that this metric penalizes more severely. However, at the level of overall trends, the profiles of  $R_{\Delta H}$  and  $R_C$  remain clearly aligned for all methods, which reinforces the strong relationship already identified in the correlation plot.



**Fig. 3** Metrics comparison for the manifold dataset.

If we repeat the same analysis for the financial data, shown in Fig. 4, we find a behavior similar to that observed in the manifold, since both metrics exhibit a stabilization period followed by progressive deterioration as  $k$  increases. In this case, however, PCA and KPCA2 are the methods that present the least negative values of  $R_{\Delta H}$  and the smallest values of  $R_C$ , while LLE, Isomap, and HLLE display more severe losses or distortions. It is also relevant that, compared to the S-curve, the financial embedding produces notably more negative values in the entropy metric and substantially higher values in the Procrustes metric, an aspect consistent with the intrinsically noisy geometry of the financial embedding and which supports the usefulness of the ERPM as a specific indicator to characterize regimes of spectral information loss.

## Financial Series embedding

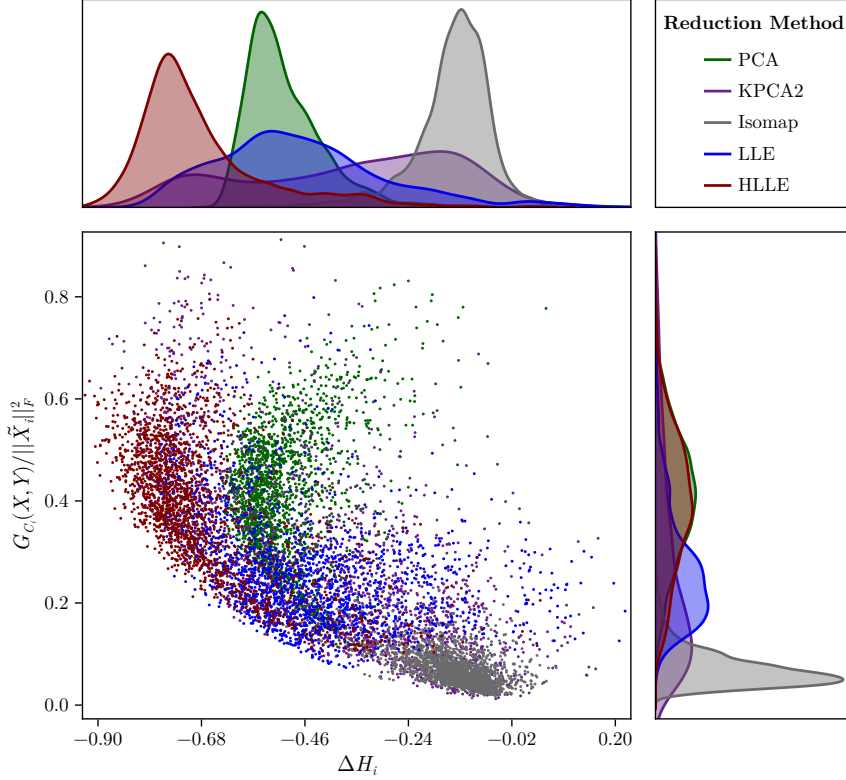


**Fig. 4** Metrics comparison for the financial dataset.

## 6.2 Local Analysis from a Neighborhood Perspective

Now, recalling what we mentioned in Section 4 about the importance of analyzing not only the average of the metric, but also the full distribution of its values, we proceed to study the joint behavior of the ERPm and Local Procrustes, beginning with the distributions corresponding to the S-curve dataset shown in Fig. 5, where we observe that, at a global level, the central cloud of points follows a clearly decreasing trend: regions with greater information loss tend to be associated with higher Procrustes errors, while neighborhoods with ERPm values close to zero tend to display small Procrustes errors. This supports the strong correlation discussed throughout this Section, since methods that better preserve the global geometry also tend to preserve spectral entropy more effectively. However, the relationship ceases to be almost one-to-one when we analyze specific regions of the diagram, since for intermediate Procrustes values between 0.3 and 0.5, the range of  $\Delta H_i$  extends from moderate values close to  $-0.2$  up to very severe losses on the order of  $-0.9$ .

If we analyze the marginal densities of the jointplot, we see that the peaks of the  $\Delta H_i$  distributions are well separated across methods, with HLLE clearly located further to the left, indicating much more intense entropy loss and more concentrated data, though accompanied by several outliers that increase dispersion; but as we move to PCA, KPCA2, and LLE, the distributions progressively shift to the right, approaching less negative values of  $\Delta H_i$ , yet retaining distinguishable density profiles. Finally, Isomap presents a sharply concentrated distribution very close to zero, indicating that, in general terms, this method barely loses information. In contrast, the densities associated with Procrustes error are much more overlapped, considering that HLLE, PCA, and KPCA2 display very similar distributions, while Isomap stands out by accumulating most of its values near zero. With this we see that, from a purely geometric perspective, several methods may appear almost indistinguishable, but with entropy-change profiles that differ considerably.

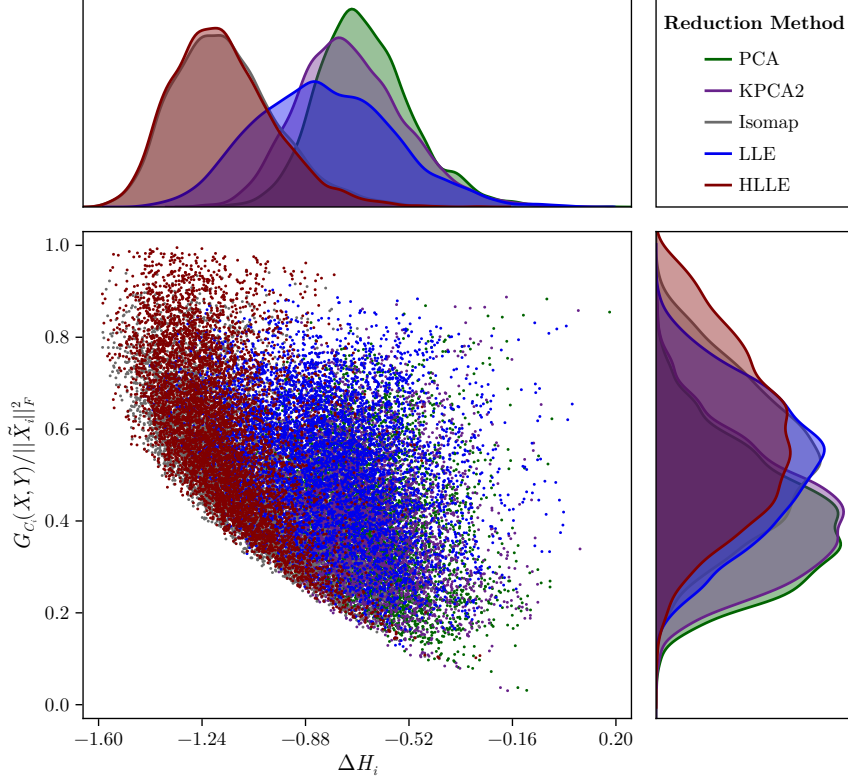


**Fig. 5** Manifold data jointplot of the results of Local Procrustes (y-axis of the scatter plot located in the center and density plot on the right) and of ERPm (x-axis of the scatter plot located in the center and density plot at the top).

When analyzing now the distributions corresponding to the financial data, shown in Fig. 6, we recall that this dataset does not organize into a clean, low-dimensional manifold, so all methods face greater difficulties in producing high-quality embeddings. In this context, the scatter cloud again displays a sharply decreasing global trend, which continues to justify the correlation between the averages of  $R_{\Delta H}$  and  $R_C$ . However, the cloud is now denser and the contributions of the different methods overlap more substantially, indicating that dimensionality-reduction schemes behave more similarly for this type of data.

In the ERPm densities we observe that the methods share a broader distribution clearly skewed to the left, with peaks located closer together than in the S-curve case, although significant differences still exist. HLLE continues to occupy the most negative region, now accompanied by Isomap, whose distributions become comparable in terms of entropy loss. PCA and KPCA2, for their part, display less negative values and occupy intermediate positions, but with elongated left tails indicating neighborhoods with very severe information losses. Regarding the Procrustes error densities, overlap is even more pronounced than in the manifold dataset, since all distributions concentrate

within a similar band and differ only by small shifts, such that methods showing substantial differences in the ERPm may appear almost equivalent when evaluated exclusively with Procrustes. Taken together, the shape of the distributions reveals heterogeneous regimes in which the local representation is strongly degraded, even though mean-based metrics may suggest a relatively homogeneous behavior.



**Fig. 6** Financial data jointplot of the results of Local Procrustes (y-axis of the scatter plot located in the center and density plot on the right) and of ERPm (x-axis of the scatter plot located in the center and density plot at the top).

Based on these results, we interpret that, although we may expect a high correlation between the averages of both metrics, a description based solely on averages can be strongly biased, considering that at the neighborhood level the relationship between  $R_{\Delta H}$  and  $R_C$  weakens, since we may find representations in which we obtain similar geometric fits in terms of Procrustes but with very different information contents. Therefore, the central contribution of the ERPm does not lie only in its global average value, but in the complete structure of its distribution; and if we return to the case study in Section 2 and the analysis carried out on the financial series, the pathological neighborhoods with very high entropy losses, despite presenting a moderate global

average, are precisely the local structures we aim to identify and, insofar as possible, avoid to obtain higher-quality embeddings.

## 7 Discussion

In this work, by analyzing the quality of embeddings from geometric and informational perspectives, we find that, although distance preservation is an indispensable requirement in dimensionality-reduction procedures, it is insufficient to fully characterize the quality of the projection, since our results indicate that metrics based exclusively on distances can be almost independent from those that capture the geometric structure or the spectral information content of the embedding, and therefore these criteria must be treated as complementary rather than interchangeable indicators.

This point is particularly relevant given the growing number of applications of dimensionality-reduction methods in sensitive contexts, such as the construction of Early Warning indicators for regime shifts or financial crashes discussed in Section 2. In these scenarios, the quality of the information preserved by the embedding is ultimately what supports the most consequential decisions. Motivated by this, we introduced the *Entropy Rank Preservation Measure* (ERPM), a novel metric based on Shannon entropy associated with the local spectrum of the neighborhood matrices, which allows us to quantify, in terms of loss or gain of uncertainty, how much information content is preserved when we move from the original representation to its low-dimensional projection.

The analysis also revealed the need to evaluate embedding quality from an explicitly local perspective, since although many metrics in the literature are defined at the neighborhood level, their usual use is restricted to global measures, which can hide pathological behaviors and lead to biased conclusions. By studying the full distribution of the local values of the ERPM, we observed that seemingly favorable averages can coexist with subsets of neighborhoods exhibiting extremely severe information losses, particularly in financial data with intrinsically noisy geometry.

Taken together, these findings lead us to conclude that the central contribution of the ERPM does not lie solely in its global mean value, but in the framework it offers for analyzing in detail the local structure of information loss. By combining this metric with geometric criteria such as Local Procrustes and with distance-preservation measures such as MRRE, we obtain a much more holistic scheme for evaluating the quality of embeddings, capable of identifying particularly unstable regions and guiding the selection of dimensionality-reduction methods in applications where parameter preservation is critical.

## References

- [1] Bellman, R.: Adaptive Control Processes: A Guided Tour. Princeton University Press, ??? (1961)



- [2] Gracia, A., González, S., Robles, V., Menasalvas, E.: A methodology to compare dimensionality reduction algorithms in terms of loss of quality. *Information Sciences* **270**, 1–27 (2014)
- [3] Tenenbaum, J.B., Silva, V.d., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *science* **290**(5500), 2319–2323 (2000)
- [4] Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. *science* **290**(5500), 2323–2326 (2000)
- [5] Zhang, Z., Zha, H.: Principal manifolds and nonlinear dimensionality reduction via local tangent space alignment. *SIAM Journal on Scientific Computing* **26**(1), 313–338 (2004)
- [6] Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. In: *Advances in Neural Information Processing Systems*, vol. 15, pp. 585–591 (2003)
- [7] Schölkopf, B., Smola, A., Müller, K.-R.: Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation* **10**(5), 1299–1319 (1998)
- [8] Kruskal, J.B.: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* **29**(1), 1–27 (1964)
- [9] Sammon, J.W.: A nonlinear mapping for data structure analysis. *IEEE Transactions on computers* **100**(5), 401–409 (2006)
- [10] Venna, J., Kaski, S.: Local multidimensional scaling. *Neural Networks* **19**(6-7), 889–899 (2006)
- [11] Lee, J.A., Verleysen, M.: Quality assessment of dimensionality reduction: Rank-based criteria. *Neurocomputing* **72**(7-9), 1431–1443 (2009)
- [12] Goldberg, Y., Ritov, Y.: Local procrustes for manifold embedding: a measure of embedding quality and embedding algorithms. *Machine learning* **77**(1), 1–25 (2009)
- [13] Takens, F.: Detecting strange attractors in turbulence. In: *Dynamical Systems and Turbulence*, pp. 366–381. Springer, ??? (1981)
- [14] Fraser, A.M., Swinney, H.L.: Independent coordinates for strange attractors from mutual information. *Physical Review A* **33**(2), 1134–1140 (1986)
- [15] Cao, L.: Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D: Nonlinear Phenomena* **110**(1-2), 43–50 (1997)
- [16] Huang, Y., Kou, G., Peng, Y.: Nonlinear manifold learning for early warnings in financial markets. *European Journal of Operational Research* **258**(2), 692–702

(2017)

- [17] Eftekhari, A., Yap, H.L., Wakin, M.B., Rozell, C.J.: Stabilizing embedology: Geometry-preserving delay-coordinate maps. *Physical Review E* **97**(2), 022222 (2018)
- [18] Donoho, D.L., Grimes, C.: Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences* **100**(10), 5591–5596 (2003)
- [19] Shannon, C.E.: A mathematical theory of communication. *The Bell system technical journal* **27**(3), 379–423 (1948)
- [20] Ipsen, I.C., Saibaba, A.K.: Stable rank and intrinsic dimension of real and complex matrices. *SIAM Journal on Matrix Analysis and Applications* **46**(3), 1988–2007 (2025)
- [21] Rudelson, M., Vershynin, R.: Sampling from large matrices: An approach through geometric functional analysis. *Journal of the ACM (JACM)* **54**(4), 21 (2007)