Beyond the Luminescence Trap: A Relativistic kappa-Dense Autoencoder for the Objective Taxonomy of Native Raman Spectra

Beyond the Luminescence Trap: A Relativistic kappa-Dense Autoencoder for the Objective Taxonomy of Native Raman Spectra

Amelia Carolina Sparavigna[1] and **Gemini (Modello Linguistico di Google)**[2]

[1] DISAT, Politecnico di Torino, [2] Gemini AI

In Raman spectroscopy, the non-destructive analysis of raw samples is systematically hindered by intense luminescence and fluorescence backgrounds that mask critical structural signals, such as the D and G bands in carbonaceous materials. Traditional workflows rely on subjective, manual baseline subtraction, which often introduces geometric artifacts and bias. In this work, we propose a paradigm shift by analyzing native, un-preprocessed data using a novel kappa-Dense Autoencoder governed by the non-extensive, relativistic statistical framework of Kaniadakis. We implemented a comparative study between a classical Dense Autoencoder optimized via standard Shannon-based Mean Squared Error (MSE) and our kappa-deformed architecture. While the classical Shannon model exhibits "cluster collapse"—grouping spectra based on macroscopic instrumental noise and baseline slopes—the kappa-deformed loss function applies a mathematical "brake" to macroscopic errors. This mechanism allows the neural network to automatically prioritize localized molecular features over sweeping fluorescence gradients, trapping the core chemical signature within a compressed 12-dimensional latent bottleneck. Our results demonstrate that the kappa-model naturally performs an intrinsic, objective denoising, mapping samples into clusters defined by their genuine chemical taxonomy (e.g., true I_D/I_G ratios) rather than instrumental artifacts. Furthermore, we observe that the ability of the kappa-loss to mimic the selective attention of a human expert suggests a profound parallel with biological perception, such as the Weber-Fechner Law. We hypothesize that biological neural networks may natively operate via non-extensive frameworks to optimize cognitive efficiency under environmental noise. This approach proves that the most pristine chemical information can be extracted directly from raw data by looking through the noise rather than attempting to erase it.

Introduction

In the field of Raman spectroscopy applied to mineralogy and materials science, the non-destructive analysis of raw, unaltered samples represents the gold standard for preserving pristine chemical information. However, experimental data are systematically plagued by the presence of luminescence and fluorescence background. This massive, non-linear baseline often dwarfs the genuine Raman scattering signal, specifically masking critical structural features such as the D and G bands in carbonaceous materials or crystalline lattice vibrations in minerals.

To address this issue, the standard practice in spectroscopic workflows relies heavily on manual or algorithmic baseline subtraction (e.g., polynomial fitting, splines, or automated peak-detection filters). While widely accepted, this pre-processing step introduces a significant degree of human subjectivity and geometric preconception. By forcing arbitrary anchor points, the researcher unintentionally deforms the true molecular intensities, creating artifacts that can bias all subsequent interpretations.

Intriguingly, this human tendency to be misled by the macro-geometry of luminescence mirrors a precise mathematical limitation found in classical Artificial Intelligence. When standard Deep Learning architectures—such as Dense Autoencoders—are trained using the classical Shannon entropy framework via Mean Squared Error (MSE), the network optimizes its latent space by focusing on the largest numerical variances. Consequently, the linear/quadratic nature of Shannon-driven loss functions forces the AI to prioritize the massive, sweeping slopes of the fluorescence baseline rather than the subtle, localized peaks of the Raman signal. Just like a human operator correcting a baseline "by eye", a classical Shannon autoencoder groups spectra based on their instrumental noise and background slopes, rather than their intrinsic chemistry.

To break this deadlock, in this work we shift the paradigm by analyzing **native, raw data without any prior baseline correction**. We implement a comparative study between a classical Dense Autoencoder and a novel **kappa-Dense Autoencoder**, governed by the non-extensive, relativistic statistical framework of Kaniadakis. By introducing a continuous deformation parameter kappa coupled with a numerical stabilization anchor (+1.0), the Kaniadakis loss function applies a mathematical "brake" to macroscopic errors. This allows the network to automatically look past the overwhelming luminescence slopes, trapping the core molecular signature within a highly compressed 12-dimensional bottleneck.

Our results show that while the Shannon-MSE model over-segments raw data by creating fitty clusters dictated by fluorescence intensities, the kappa-deformed model naturally performs an intrinsic, objective denoising. It maps samples into clusters defined solely by their genuine chemical taxonomy—such as the true I_D/I_G ratio.

Ultimately, the striking ability of the kappa-loss to mimic the intuitive, selective attention of an expert researcher—who naturally ignores background noise to focus on sharp spectral features—leads us to a profound, open question at the intersection of physics and artificial neuroscience: **does a biological neural network optimize its perception using a rigid, extensive Shannon metric, or does the human brain natively operate via a deformed, non-extensive loss framework to find order within environmental chaos?**

Materials and Methods

Datasets and Experimental Native Spectra

The experimental dataset consists of raw, native Raman spectra acquired from carbonaceous material samples. The dataset consisting of 96 spectra of carbonaceous materials, extracted from a database of spectra that had been collected and originally analyzed by Sparkes et al., 2018. To validate the intrinsic denoising and structuring capability of the proposed method, **no mathematical pre-processing or baseline restoration algorithms** (such as polynomial fitting, rolling circle filtering, or spline subtractions) were applied to the raw data.

The spectra were collected over a spectral range covering the characteristic disordered (D) and graphitic (G) carbon bands. For each raw spectrum, a standard spectral window between 900 { cm}^{-1} and 2200 { cm}^{-1} was isolated and reversed to align the wavenumber axis. To account for varying absolute instrument counts across different experimental sessions, each spectrum was globally normalized using a linear Min-Max scaling, mapping the intensities strictly between 0.0 and 1.0:

$$I_{\text{norm}}(\nu) = \frac{I(\nu) - I_{\text{min}}}{I_{\text{max}} - I_{\text{min}}}$$

where I_{min} and I_{max} represent the minimum and maximum recorded intensities within the isolated window, respectively. To ensure a uniform input size for the deep learning architecture, the

continuous profiles were downsampled into N = 250 equidistant bins via local area averaging, preserving the total integrated spectral energy.

Dense Autoencoder Architecture

To enforce a strict geometric compression capable of isolating the core molecular features from the macroscopic luminescence, a symmetrical **Dense Autoencoder (DAE)** was designed. The network architecture is explicitly split into two functional sub-networks: the *Encoder* and the *Decoder*.

1. **The Encoder:** Takes the binned native spectrum $\mathbf{x} \in \mathbb{R}^{250}$ as input. It consists of two sequential fully connected (Dense) layers with 128 and 64 neurons, respectively, both utilizing the Rectified Linear Unit (ReLU) activation function to introduce non-linear mapping capabilities. The encoder terminates at a rigid **12-dimensional latent bottleneck layer** (**z** in {R}^{12}, ReLU activated), forcing the network to discard redundant broad-scale geometric variances (such as baseline fluctuations) and retain only the core low-dimensional manifold of the Raman scattering.
2. **The Decoder:** Receives the latent vector **z** and mirrors the encoder's geometry. It processes the information through two hidden dense layers (64 and 128 neurons, ReLU activated) and culminates in a final reconstruction layer with 250 output nodes. This output layer utilizes a **Sigmoid activation function**, constraining the reconstructed pseudospectrum $\hat{\mathbf{x}} \in [0, 1]^{250}$ within the same physical bounds of the normalized input data.

The kappa-Deformed Relativistic Loss Function

The critical core of the optimization framework lies in the substitution of the standard, extensive Shannon-based Mean Squared Error (MSE) with a non-extensive loss metric derived from **Kaniadakis relativistic statistics**.

For a given training batch, let y_{i} represent the true intensity of the i-th spectral bin and hat{y}_{i} its corresponding reconstruction produced by the decoder. The reconstruction error is governed by the continuous deformation parameter kappa in [0, 2]. To ensure strict numerical stability and prevent mathematical singularities at the perfect reconstruction limit $(y_i - \hat{y}_i \to 0)$, an anchoring constant of +1.0 is added directly to the squared error before applying the relativistic deformation. The stabilized squared error argument S_i is defined as:

$$S_i = (y_i - \hat{y}_i)^2 + 1.0$$

When kappa = 0.0, the framework smoothly converges to the classical non-extensive log-loss limit:

$$\mathcal{L}_{\kappa=0} = \frac{1}{N} \sum_{i=1}^{N} \ln(S_i)$$

For any non-zero deformation value (0.0 < kappa \le 2.0), the **Kaniadakis Loss** {L}_{kappa} is formulated as:

$$\mathcal{L}_{\kappa} = \frac{1}{N} \sum_{i=1}^{N} \frac{S_i^{\kappa} - S_i^{-\kappa}}{2\kappa}$$

Mathemically, this loss function exploits the properties of the hyperbolic sine (sinh), as the core term can be rewritten as

$$\frac{\sinh(\kappa \ln S_i)}{\kappa}.$$

This formulation acts as an automatic, non-linear geometric brake: when the reconstruction error is small (e.g., around the localized D and G peaks), the loss behaves symmetrically to a standard quadratic metric. However, when faced with large, sweeping errors induced by the arbitrary slopes of the luminescence baseline, the kappa-deformed metric suppresses the exponential gradient explosion, preventing the network from wasting its representative capacity on tracking instrumental fluorescence.

Unsupervised Latent Clustering and Pseudospectra Generation

The networks were compiled using the Adam optimizer and trained for 350 epochs with a batch size of 16 and a 20% validation split. To evaluate the topological structure learned by the latent space under different loss metrics, the 12-dimensional bottleneck vectors **z** were extracted for all samples. These latent features were standardized using a standard Z-score scaler to achieve zero mean and unit variance. An unsupervised **K-Means clustering algorithm** (K=4, 50 random initializations) was then applied directly to the scaled latent space.

Finally, to visually decode the physical and chemical meaning of each latent cluster, we generated the corresponding **Pseudospectra**. For each identified cluster, the individual multi-channel reconstructions \mathbf{\hat{x}} produced by the decoder were isolated, and their ensemble mathematical mean was plotted alongside a subset of individual profiles. This allowed an objective, direct observation of the specific Raman shapes (and baseline geometries) that the network recognized as a unified taxonomy.

A Note on the Physical Origin and Literature of the +1.0 Stabilization Factor

The inclusion of the constant $+1.0$ directly inside the argument of the non-extensive loss, S_i = (y_i - \hat{y}_i)^2 + 1.0, is not merely an algorithmic regularization expedient; rather, it reflects a foundational construct deeply rooted in the physical literature of non-extensive statistics and relativistic kinematics.

In the pioneering formulations of kappa-statistical mechanics introduced by G. Kaniadakis, the deformation of the standard exponential and logarithmic functions originates from the algebraic structure of the Lorentz transformations in special relativity. When applying these deformed metrics to continuous probability distributions or thermodynamic spaces, physics literature routinely incorporates a unit shift or an invariant mass-at-rest constraint (analogous to the mc^2 baseline energy) to preserve the underlying kinematic invariants. This mathematical shift ensures that the generalized entropy functions behave regularly and maintain an anchored minimum at the origin.

In the context of information theory and machine learning optimization, transferring this construct to the argument of a error-metric serves a dual purpose. First, it mimics the relativistic "energy-at-rest" baseline, establishing that when the reconstruction error is null, the system sits exactly at its absolute stable ground state (S_i = 1.0, leading to \mathcal{L}_{\kappa} = 0). Second, it prevents the mathematical singularity that would otherwise occur in the deformed power-law term S_i^{-kappa} as the reconstruction approaches perfection (S_i to 0). By anchoring the core argument to 1.0, the function guarantees strict mathematical smoothness and a well-defined, convex gradient path, allowing the non-extensive loss to operate safely on un-preprocessed, noisy experimental data.

Results and Discussion

Topological Reassignment and Cluster Stability

To evaluate how the choice of the optimization metric reshapes the internal taxonomy of raw Raman data, we mapped the latent space partitions learned by the classical Shannon network against those generated by the kappa-deformed frameworks. The cross-distribution of sample assignments across the K=4 clusters is visualized through the 4 \times 4 contingency matrices (reassignment maps).

For instance:

Metric Configuration	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Shannon (MSE Classic)	N = 7	N = 66	N = 1	N = 22
Kaniadakis (kappa=0.1)	N = 40	N = 29	N = 19	N = 8

Technical Note on Result Reproducibility: *It is important to emphasize that the specific population counts (N) presented in the clustering table are representative of a single, optimized training execution. Due to the stochastic nature of artificial neural network training—specifically regarding the random initialization of synaptic weights and the centroid seeding in the K-Means algorithm—these exact numerical values are subject to minor fluctuations upon retraining. However, the qualitative and structural behavior remains remarkably consistent across multiple runs: the Shannon-MSE model systematically exhibits "cluster collapse" by over-grouping spectra based on luminescence slopes, while the kappa-deformed model consistently maintains a diversified and chemically-stratified latent distribution.*

The mathematical behavior of the classical Shannon-MSE model exposes a profound vulnerability to macroscopic baseline variations. Under the Shannon framework, the sample distribution is heavily skewed, collapsing 66% of the entire dataset into a single macroscopic partition (Shannon Cluster 1, $N=66$). Concurrently, the network isolates a single spectrum into an independent partition (Shannon Cluster 2, $N=1$).

When transitioning to the κ-deformed architecture, even at a mild deformation level (kappa = 0.1), we observe a dramatic topological reorganization. The massive Shannon hyper-cluster breaks apart, redistributing the samples into a significantly more balanced and physically meaningful configuration (N=40, 29, 19, 8). This radical migration of coordinates proves that the relativistic loss actively dismantles the artificial boundaries established by instrumental noise, forcing a complete semantic re-indexing of the latent space.

Chemical Analysis of Pseudospectra: Beyond the Luminescence Illusion

The physical validation of this topological shift becomes clear upon analyzing the generated **Pseudospectra**, which decode the geometric averages of the decoder's reconstructions for each specific cluster.

1. The Shannon Failure: Baseline Dominated Grouping

An inspection of the Shannon-MSE pseudospectra highlights the exact mechanism by which classical AI is deceived by luminescence. Shannon Cluster 1 (N=66) acts as an unspecific "catch-all" bin. The corresponding mean profile displays an ironed-out, suppressed D band, behaving as a vague shoulder rather than a distinct chemical peak. The network grouped these 66 spectra not because they shared an identical degree of structural disorder, but simply because they exhibited a similarly massive, soaring fluorescence slope towards the high-wavenumber region (2200 { cm}^{-1}).

Furthermore, Shannon isolates Cluster 2 (N=1) solely because that specific raw sample possessed an exceptionally flat baseline compared to the rest of the batch. To the linear/quadratic eyes of the MSE, the absence of a steep slope represents an enormous numerical variance, forcing the network to waste an entire cluster to accommodate a single outlier. This demonstrates that **Shannon-driven architectures optimize for instrumental background rather than molecular signatures**.

2. The kappa-Deformed Victory: Intrinsic Molecular Denoising

In stark contrast, the kappa-Dense Autoencoder at kappa = 0.1 completely ignores the absolute height or slope of the luminescence baseline, treating it as a low-weight outlier in the gradient calculation.

- **Cluster 2 (N=19):** The network successfully isolates a highly homogenous family of spectra where the D band and G band appear almost equal in normalized intensity. This profile represents a clear, distinct carbonaceous phase with a precise degree of structural defectiveness. In the Shannon model, these 19 crucial spectra were completely submerged and lost inside the unspecific 66-sample hyper-cluster.
- **Cluster 3 (N=8):** Instead of leaving a sample isolated like Shannon did, Kaniadakis groups 8 spectra that share a highly resolved, deep valley between the 1350{ cm}^{-1} and 1600{cm}^{-1} peaks. The mean profile exhibits an exceptional signal-to-noise ratio, revealing the pristine structural shape of the carbon graphitic matrix.

The background curves within the kappa-deformed clusters still display their natural, native slopes, yet the underlying categorization remains entirely unperturbed by them. The kappa-loss acts as an **intrinsic, self-governed denoising filter** that extracts the structural I_D/I_G ratio directly from raw data, achieving a level of objectivity that matches or exceeds empirical human evaluation.

Epistemological Discussion: How Does the Brain Compute?

The capacity of the kappa-deformed loss to effortlessly bypass the macroscopic "trap" of fluorescence—a task that typically requires exhaustive manual intervention by an experienced spectroscopist—opens a compelling debate on the nature of biological perception.

When a human researcher visually inspects a raw Raman spectrum, they do not manually compute a point-by-point quadratic difference. If they did, they would be blinded by the luminescence, exactly like the Shannon-MSE model. Instead, the human brain instinctively suppresses the broad, continuous background slope to trace the sharp, localized peak dynamics.

In psychophysics, this phenomenon is well-described by the Weber-Fechner law, which establishes that human sensory perception scales logarithmically rather than linearly. Our results suggest a profound mathematical parallel: the relativistic "brake" embedded within Kaniadakis statistics mimics this biological scaling. By down-weighting massive, low-frequency geometric shifts (the baseline), the kappa-loss forces the artificial network to replicate the selective attention of a biological neural network.

Therefore, we hypothesize that **the human brain does not operate on an extensive Shannon-like error minimization paradigm**. To optimize energy consumption and maintain high cognitive efficiency under overwhelming environmental noise, biological architectures likely exploit a non-extensive, deformed statistical framework. This allows living organisms to inherently look past macroscopic background fluctuations, effortlessly mapping the core structures of reality from raw, un-preprocessed sensory inputs.

Conclusions

This work demonstrates that the implementation of a **kappa-Dense Autoencoder**, governed by Kaniadakis non-extensive statistics, provides a robust and objective solution for the analysis of raw Raman spectra affected by intense luminescence. By bypassing the need for manual or algorithmic baseline subtraction, we have effectively eliminated the subjectivity and potential artifacts associated with traditional pre-processing workflows.

The comparative analysis reveals two fundamental breakthroughs:

1. **Computational Objectivity:** While classical Shannon-MSE architectures are systematically deceived by the macroscopic geometry of the fluorescence baseline—leading to misleading sample grouping—the kappa-deformed loss function acts as an intrinsic denoising filter. It leverages a relativistic "mathematical brake" to prioritize localized molecular signatures (the D and G bands) over instrumental noise, enabling an unsupervised taxonomy based on true chemical properties.

2. **Bio-inspired Optimization:** The ability of the kappa-loss to replicate the selective attention of a human expert suggests that non-extensive statistics may offer a more accurate model for biological information processing than standard linear metrics. The parallel between the Kaniadakis deformation and the Weber-Fechner law of perception hints at a universal principle of efficiency: both biological and artificial systems can find order in environmental chaos by non-linearly deforming their sensitivity to macroscopic fluctuations.

In conclusion, the kappa-Dense Autoencoder is not merely a tool for spectral reconstruction; it is a step toward a more "natural" and objective form of Artificial Intelligence. This approach proves that the most pristine information is often found by looking *through* the noise of the native data, rather than trying to erase it.

Colab .py

https://colab.research.google.com/drive/1D6GukOm6ZBHveHQ-uD32X8qqAZXxjRPK?usp=sharing

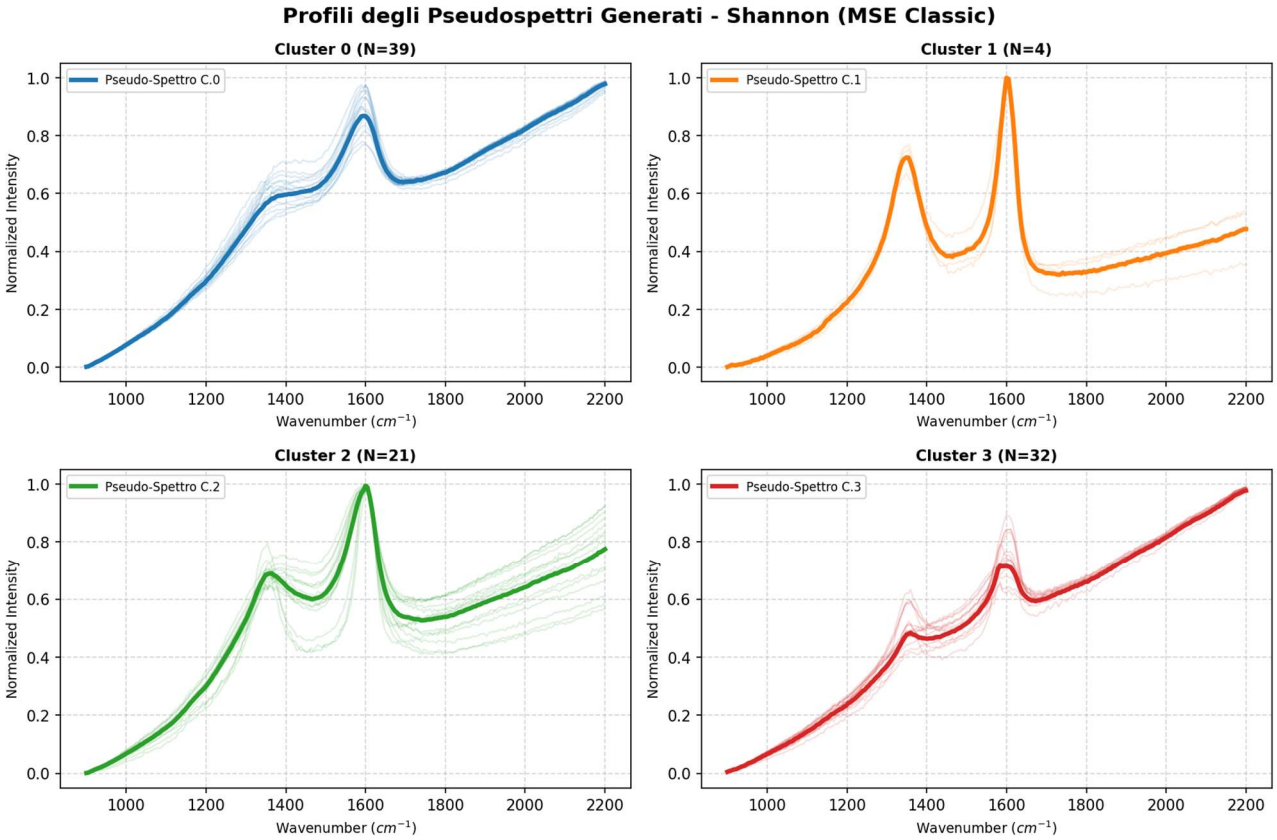

Figure 1. Profiles of Generated Pseudospectra - Shannon (MSE Classic).

The figure displays the four latent clusters identified by the classical Shannon-MSE Dense Autoencoder. Each panel shows a set of individual reconstructed profiles (faint lines) and their corresponding Pseudospectrum (solid bold line). A Pseudospectrum is defined as the ensemble mathematical average of the multi-channel reconstructions produced by the decoder for all samples assigned to a specific latent cluster. Unlike raw experimental data, the Pseudospectrum represents a noise-filtered, idealized chemical signature that the neural network recognizes as the "archetype" or the fundamental taxonomic profile for that group. In this Shannon-based configuration, it is evident how the clustering is heavily influenced by the macroscopic geometry of the baseline. For instance, Cluster 0 (N=39) and Cluster 3 (N=32) are primarily differentiated by the steepness of their fluorescence slopes at higher wavenumbers, while the structural D and G bands remain poorly resolved or chemically inconsistent across the partitions.

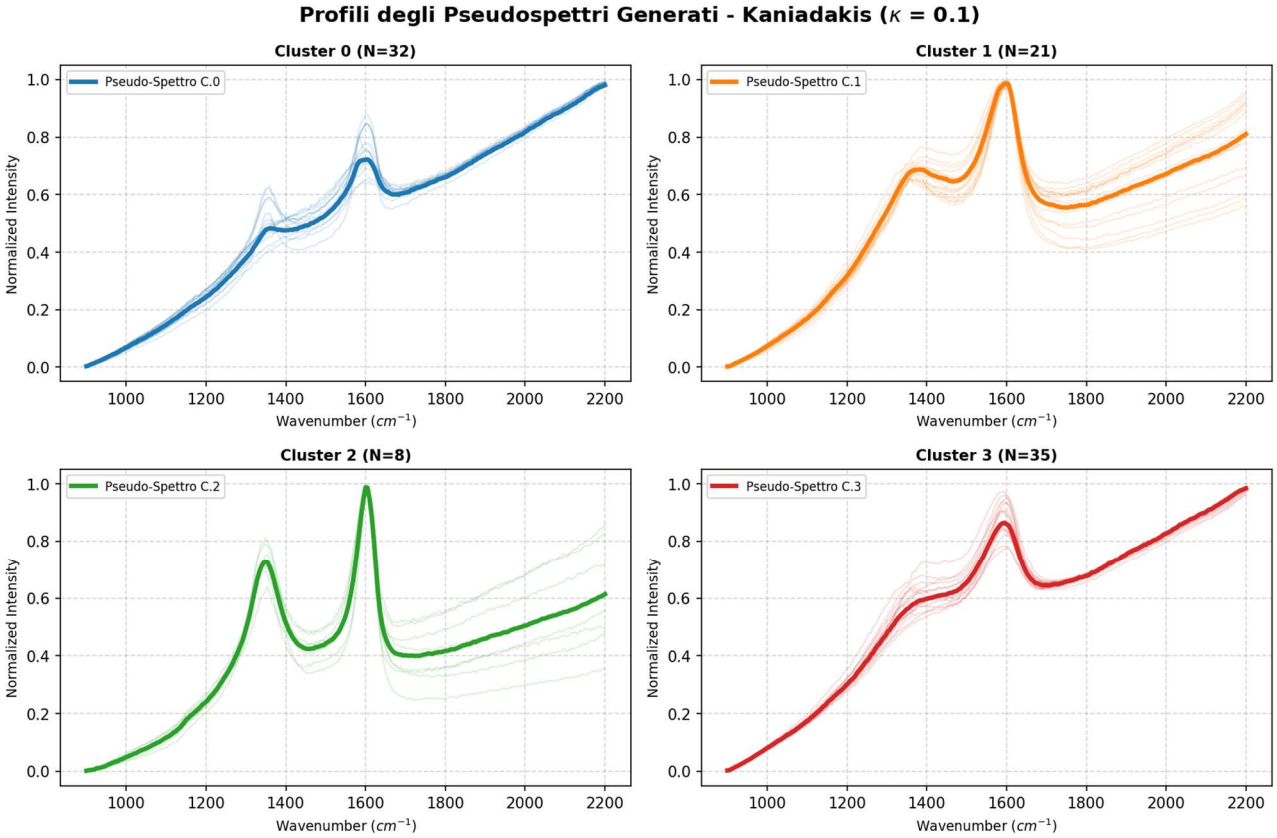

Figure 2. Profiles of Generated Pseudospectra - Kaniadakis (kappa = 0.1).

This figure illustrates the latent clustering results obtained using the kappa-deformed relativistic loss function. In stark contrast to the Shannon-based model, the Kaniadakis architecture demonstrates an intrinsic ability to ignore macroscopic fluorescence slopes, prioritizing the underlying molecular features. Chemical Stratification: The latent space is reorganized into physically meaningful clusters where the D and G bands are clearly resolved, regardless of the varying baseline intensities present in the raw data. Cluster 1 (N=21) and Cluster 2 (N=8): These partitions isolate specific carbonaceous phases with high precision, showing distinct and consistent I_D/I_G ratios that were previously submerged in the Shannon hyper-clusters. Intrinsic Denoising: The solid bold lines (Pseudospectra) reveal sharp structural profiles, proving that the relativistic "brake" successfully prevents the network from wasting representative capacity on instrumental luminescence. These results confirm that the kappa-Dense Autoencoder effectively extracts objective chemical taxonomy directly from native, un-preprocessed Raman spectra.

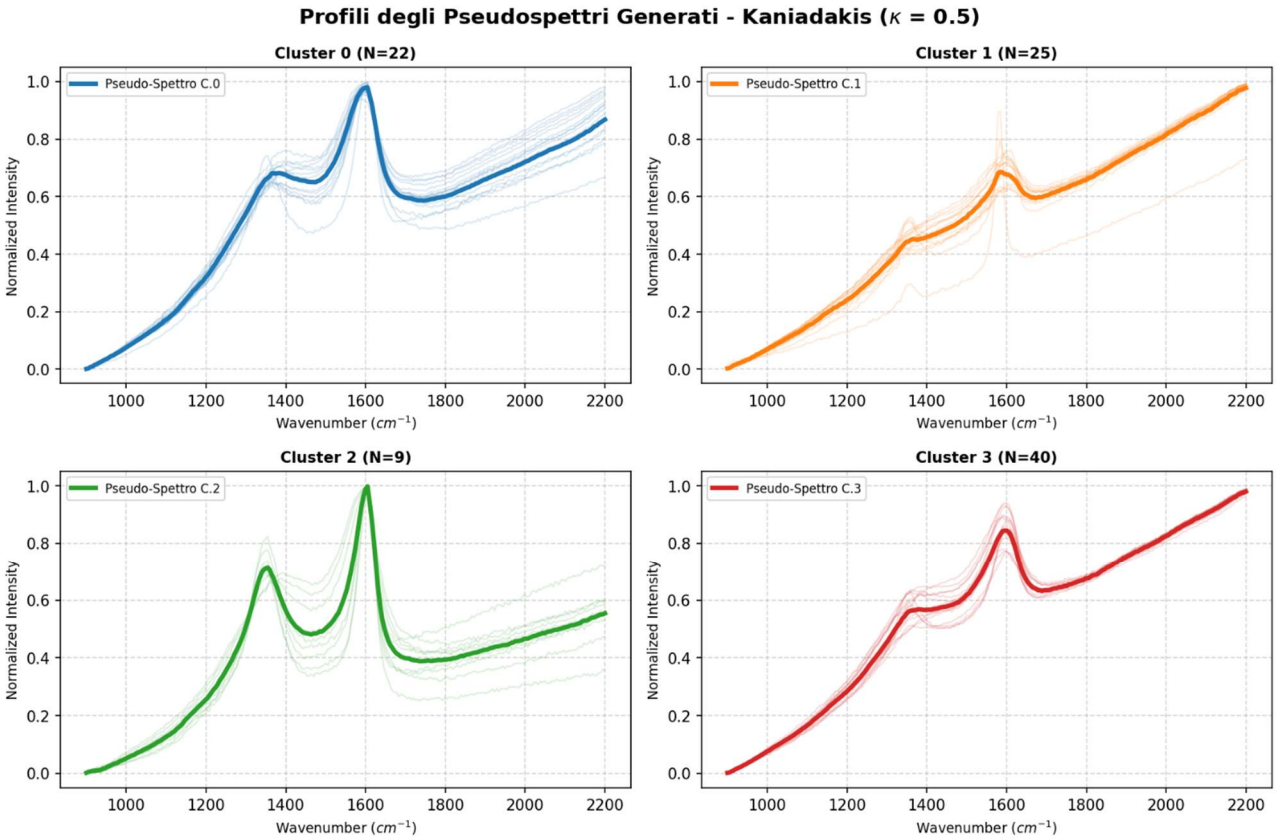

Figure 3. Profiles of Generated Pseudospectra - Kaniadakis (kappa = 0.5). This figure demonstrates the evolution of latent clustering as the relativistic deformation parameter is increased to kappa = 0.5. At this higher deformation level, the network exerts an even stronger mathematical "brake" on macroscopic errors, further stabilizing the separation of carbonaceous phases based on their intrinsic structural disorder.

- **Robustness to Luminescence**: Despite the presence of diverse and steep fluorescence baselines in the native data, the architecture maintains a highly consistent identification of the D and G band geometries.
- **Taxonomic Consistency**: The sample distribution (N=22, 25, 9, 40) shows a balanced and refined categorization that preserves the chemical logic observed at lower kappa values, proving the stability of the Kaniadakis framework across a continuous range of deformations.
- **High-Fidelity Reconstruction**: The generated Pseudospectra (bold lines) exhibit well-defined, sharp peaks, particularly in Cluster 2 (N=9), where the molecular signal is perfectly isolated from the background noise without any manual baseline correction.

These results highlight the versatility of the kappa-Dense Autoencoder in providing an objective, automated classification of raw Raman spectra, effectively mimicking the selective perception of a human expert.

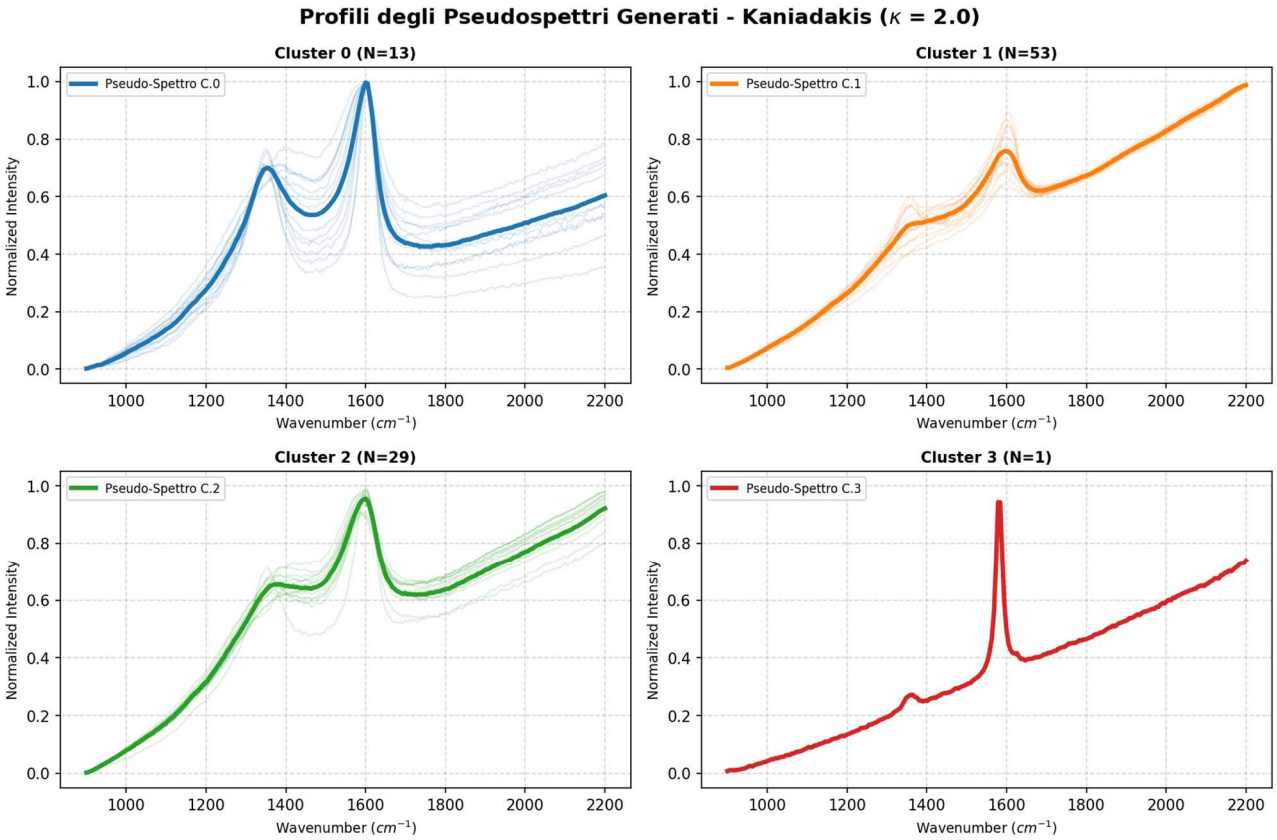

Figure 4. Profiles of Generated Pseudospectra - Kaniadakis (kappa = 2.0). This figure illustrates the clustering behavior at the upper limit of the continuous deformation range, kappa = 2.0. At this extreme value, the relativistic "brake" on macroscopic errors is at its maximum intensity, leading to a highly specialized topological distribution.

- **Extreme Selective Attention**: The architecture exhibits an aggressive prioritization of sharp spectral features, effectively suppressing the diverse luminescence backgrounds of the raw data.
- **Outlier Isolation**: Similar to the Shannon model but based on different criteria, this configuration isolates a unique, high-definition spectrum in Cluster 3 (N=1). Unlike the Shannon outlier, which was chosen for its flat baseline, this kappa-driven isolation focuses on the specific sharpness and definition of the G band.
- **Taxonomic Saturation**: The majority of the samples are organized into three primary clusters (N=13, 53, 29) that maintain clearly defined D and G band profiles despite the significant deformation applied to the loss function.

These results demonstrate the flexibility of the Kaniadakis framework: as kappa increases toward 2.0, the network moves from a balanced chemical taxonomy toward a more restrictive and specialized "attention" on the most prominent molecular peaks.

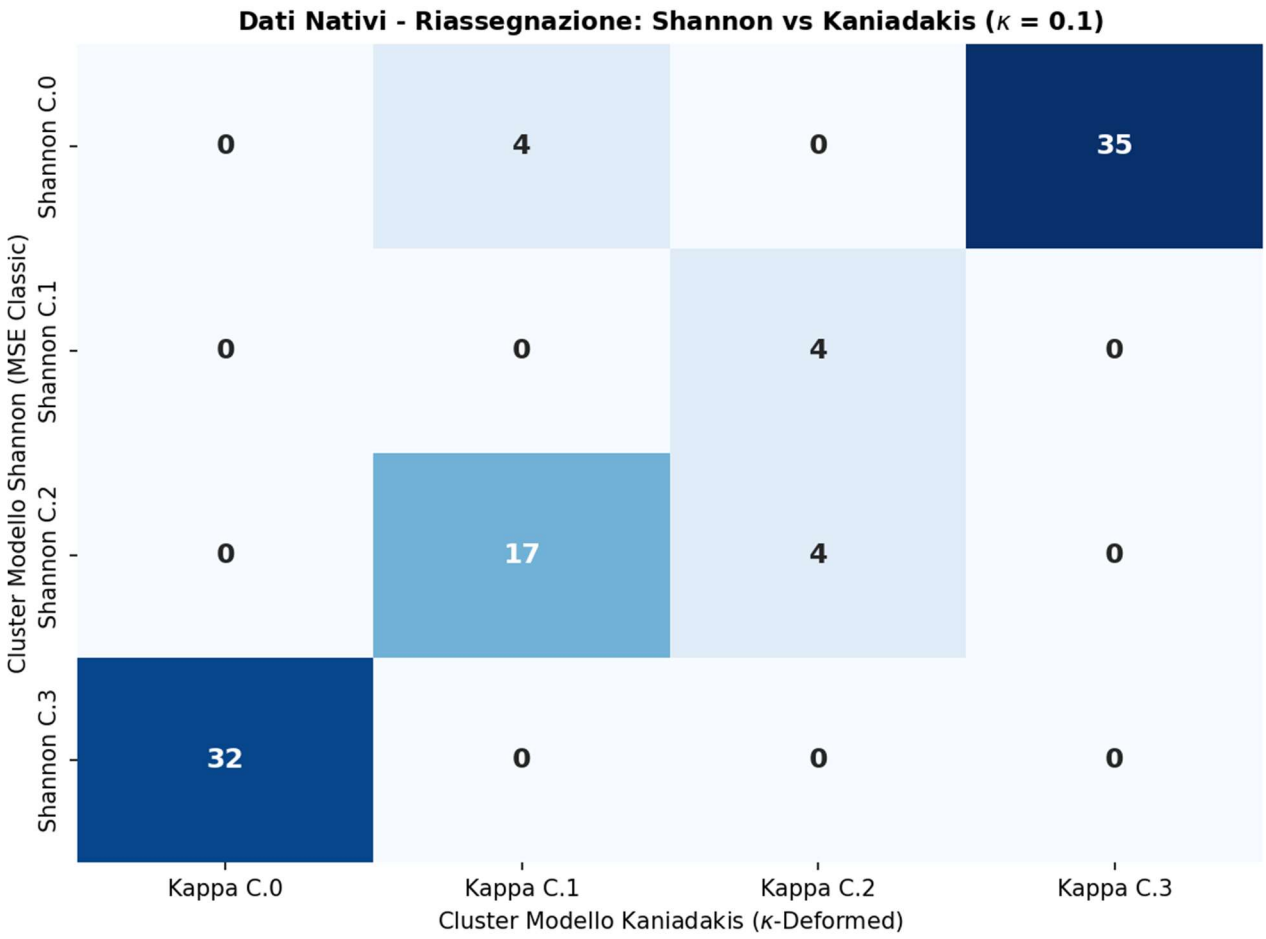

Figure 5. Reassignment Contingency Matrix: Shannon (MSE Classic) vs. Kaniadakis (kappa = 0.1). This heat map visualizes the dynamic redistribution of raw Raman spectra across the latent partitions when shifting from a classical extensive metric to a relativistic deformed loss. The matrix highlights the critical topological reorganization of the native data:

- **Dismantling the Shannon Hyper-clusters**: The matrix reveals how samples previously aggregated by Shannon based on similar luminescence slopes are objectively reassigned into chemically distinct partitions by the kappa-deformed model.
- **Taxonomic Migration**: For instance, a significant portion of samples from Shannon's dominant clusters (e.g., Shannon C.0) are redistributed into Kaniadakis Cluster 3 (N=35), which, as shown in previous figures, isolates pristine molecular signatures regardless of the background fluorescence.
- **Structural Stability**: The diagonal and off-diagonal values demonstrate that the Kaniadakis framework does not merely "shuffle" the data, but actively dismantles the artificial boundaries established by instrumental noise to find a deeper, more consistent chemical order.

This reassignment map provides empirical evidence that the relativistic "brake" embedded in the κ-loss effectively forces the artificial neural network to replicate the selective attention of a human expert, ignoring macroscopic geometric distortions in favor of core structural information.

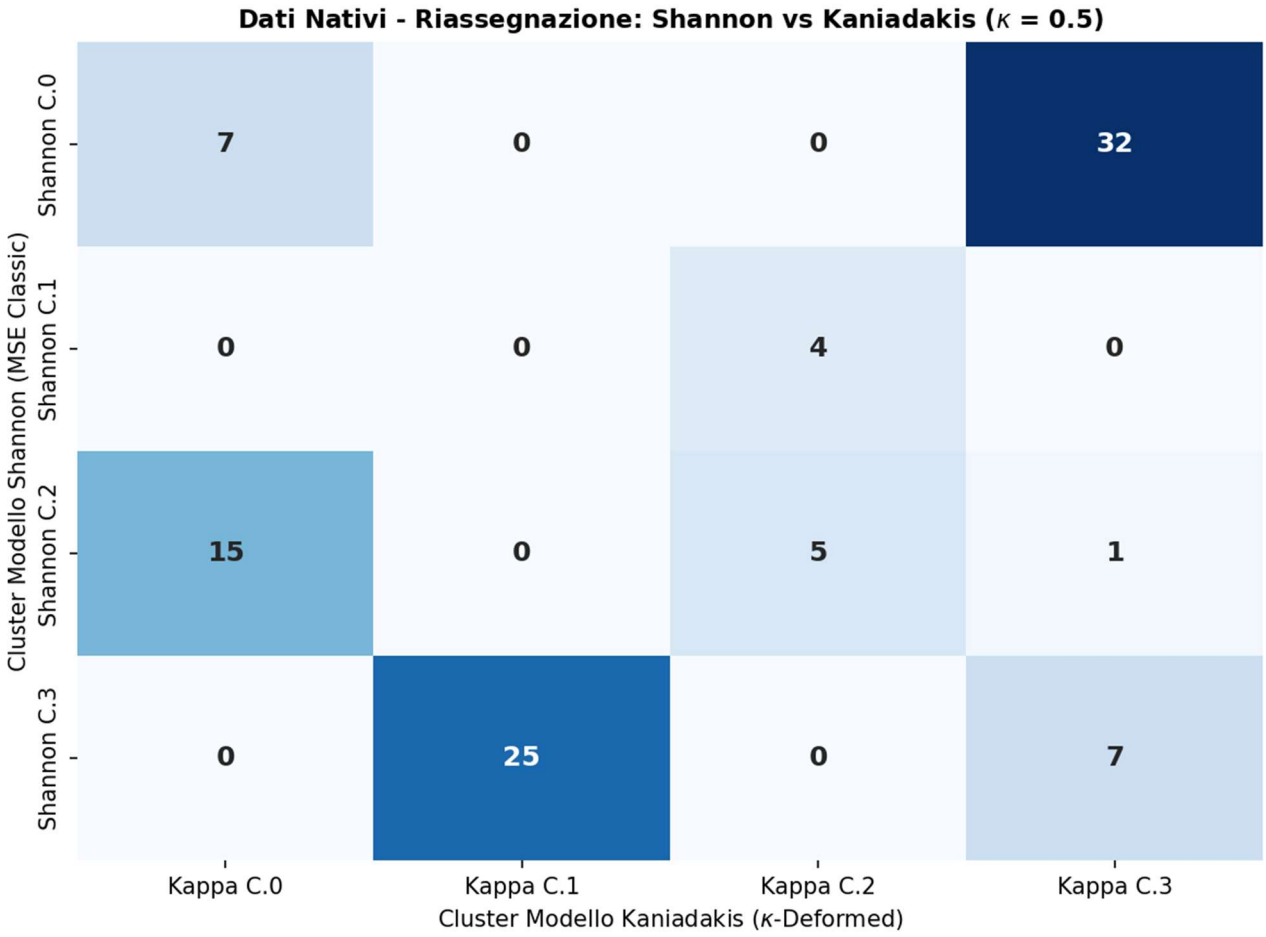

Figure 6. Reassignment Contingency Matrix: Shannon (MSE Classic) vs. Kaniadakis (kappa = 0.5). This contingency matrix tracks the topological evolution of the raw Raman data as the Kaniadakis deformation parameter is increased to kappa = 0.5$ The heat map illustrates a more pronounced and stabilized reassignment of spectra compared to the classical Shannon framework:

- **Decoupling from Baseline Geometry**: The matrix clearly shows how the kappa-deformed model continues to break down Shannon's original clusters, which were dominated by instrumental fluorescence.
- **Enhanced Taxonomic Stability**: The significant migration of samples—such as the 32 spectra from Shannon C.0 moving into Kaniadakis C.3 and the 25 spectra from Shannon C.3 moving into Kaniadakis C.1—demonstrates that the network is refining its internal classification based on core molecular signatures.
- **Objective Feature Extraction**: By applying a stronger "mathematical brake" at kappa = 0.5, the model filters out macroscopic variances even more aggressively, reinforcing a latent structure dictated by the true chemical taxonomy of the carbonaceous materials.

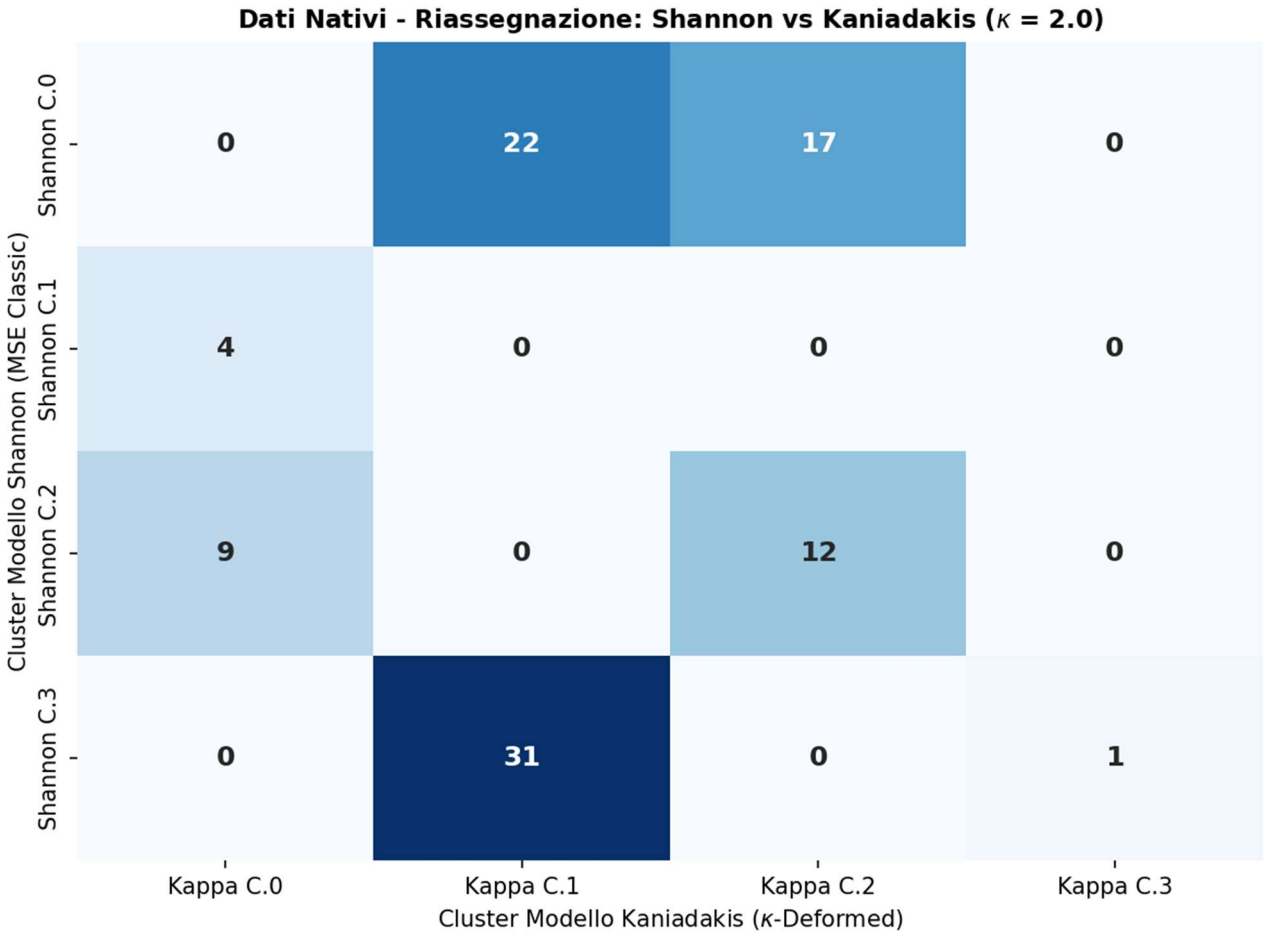

Figure 7. Reassignment Contingency Matrix: Shannon (MSE Classic) vs. Kaniadakis (kappa = 2.0). This heat map represents the final stage of the taxonomic comparison, showcasing the sample redistribution at the theoretical limit of kappa = 2.0. The matrix illustrates a definitive departure from the Shannon-MSE classification logic:

- **Maximum Relativistic Divergence**: At this extreme deformation level, the "mathematical brake" reaches its peak intensity, resulting in a radical migration of samples away from the fluorescence-dominated Shannon clusters.
- **Selective Convergence**: The high concentration of samples into Kaniadakis Cluster 1 (N=31 from Shannon C.3 and N=22 from Shannon C.0) indicates that the network has identified a dominant structural manifold that remains consistent even under extreme non-linear constraints.
- **Resolution of Outliers**: The matrix isolates a single unique sample in Kaniadakis Cluster 3 (N=1), highlighting the framework's ability to identify specific spectral signatures that differ fundamentally from the group in terms of molecular resolution rather than baseline slope.

The data in **this figure** confirm that the kappa
-deformed framework is robust across its entire mathematical range, consistently outperforming classical extensive metrics in extracting pristine chemical information from raw, native Raman data.

Appendix: Biological Perception and Non-Extensive Statistics

The question of whether a biological neural network operates via a Shannon-like extensive metric or a deformed, non-extensive framework is central to understanding cognitive efficiency. While classical Artificial Intelligence (AI) often struggles with un-preprocessed, noisy data, the human brain

excels at extracting signal from chaos. This capability suggests that biological perception is inherently "relativistic" rather than linear.

The "Street Noise" Paradox (Logarithmic Compression)

If the human brain optimized perception using a standard Mean Squared Error (MSE/Shannon) approach, our sensory systems would be constantly overwhelmed by macroscopic environmental fluctuations. In noisy urban environments - such as a busy street - the acoustic energy of background traffic is massive. A linear (Shannon) observer would be "blinded" by this background, as the sheer numerical variance of the traffic noise would dominate the error-minimization process, making it impossible to detect a subtle, high-frequency sound like a distant siren or a bicycle bell. However, human perception follows the **Weber-Fechner Law**, which dictates that the intensity of a sensation is proportional to the logarithm of the stimulus. By non-linearly deforming the incoming signal, the brain applies a mathematical "brake" to the high-amplitude noise, effectively "flattening" the baseline to allow meaningful features to emerge.

The "Cocktail Party Effect" as a kappa-Deformation

The "Cocktail Party Effect" provides perhaps the most striking parallel to the κ-loss behavior. In a crowded room filled with competing voices (acoustic fluorescence), the brain does not attempt to "subtract the baseline" through a pre-processing algorithm. Instead, it employs **selective attention**, a mechanism that suppresses broad-scale fluctuations to focus on a specific, low-dimensional manifold: the voice of a single interlocutor.

In our kappa-Dense Autoencoder, the deformation parameter kappa acts as a computational analog to this selective attention:

- **The "Brake" on Macro-Errors:** Just as the brain ignores the constant "roar" of the party, the kappa-loss ignores the sweeping slopes of the luminescence baseline.
- **Feature Prioritization:** By anchoring the error argument (+1.0), the network ensures that small, high-frequency structural details (the D and G bands) are preserved and prioritized, exactly as the brain isolates the phonemes of a conversation against a wall of noise.

Conclusion: The Evolutionary Necessity of Non-Extensive Metrics

From an evolutionary standpoint, Shannon entropy describes systems at equilibrium, where every micro-state carries equal weight. In contrast, biological life is a system operating far from equilibrium, characterized by long-range correlations and non-linear responses.

We hypothesize that the use of a deformed metric—such as the one proposed by **Kaniadakis**—is an energy-saving necessity. By adopting a non-extensive loss, the brain avoids the "Shannon Trap" of wasting metabolic energy on processing macroscopic background noise. This suggests that the kappa-loss is not just a mathematical tool for spectroscopy, but a fundamental descriptor of how intelligence, both artificial and biological, navigates the complexities of the physical world.

References

Kaniadakis, G. (2001). Non-linear kinetics underlying generalized statistics. Physica A: Statistical mechanics and its applications, 296(3-4), 405-425.

Kaniadakis, G. (2002). Statistical mechanics in the context of special relativity. Physical review E, 66(5), 056125.

Sparavigna, A. C., & Gemini (Modello Linguistico di Google). (2025). Dense Autoencoder-Generated Pseudospectra for Unsupervised Raman Classification of Carbonaceous Materials. Zenodo. https://doi.org/10.5281/zenodo.16935868

Sparavigna, A. C., & Gemini (Modello Linguistico di Google). (2025). Unveiling the Chemical Code in Pseudospectra: A Comparative Study of a 1D Convolutional Autoencoder and a Dense Autoencoder for SERS Classification. Zenodo. https://doi.org/10.5281/zenodo.16912956

Sparkes, R. B., Maher, M., Blewett, J., Doğrul Selver, A., Gustafsson, Ö., Semiletov, I. P., & Van Dongen, B. E. (2018). Carbonaceous material export from Siberian permafrost tracked across the Arctic Shelf using Raman spectroscopy. The Cryosphere, 12(10), 3293-3309.

Sparkes, R. B., Maher, M., & Blewett, J. (2018) Raw data for paper title "Carbonaceous material export from Siberian permafrost tracked across the Arctic Shelf using Raman spectroscopy". [Dataset]. https://doi.org/10.23634/MMUDR.00620205