



MOX-Report No. 39/2026

K-Models: a Flexible and Interpretable Method for Ordinal Clustering with Application to Antigen-Antibody Interaction Profiles

Patanè, G.; Menafoglio, A.; Krauth, A.; Fechner, P.; Dede', L.; Colosimo, B.M.; Nicolussi, F.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it

<https://mox.polimi.it>

K-Models: a Flexible and Interpretable Method for Ordinal Clustering with Application to Antigen-Antibody Interaction Profiles

Giulia Patanè^{1*}, Alessandra Menafoglio¹, Alexander Krauth³,
Peter Fechner³, Luca Dede¹, Bianca Maria Colosimo²,
Federica Nicolussi¹

^{1*}MOX, Department of Mathematics, Politecnico di Milano, Piazza
Leonardo da Vinci 32, Milan, 20133, Italy.

²Department of Mechanical Engineering, Politecnico di Milano, Piazza
Leonardo da Vinci 32, Milan, 20133, Italy.

³Institute of Physical and Theoretical Chemistry (IPTC), Eberhard
Karls Universität Tübingen, Auf der Morgenstelle 18, Tübingen, 72076,
Germany.

*Corresponding author(s). E-mail(s): giulia.patane@polimi.it;

Abstract

Existing clustering methods for functional data often prioritize partitioning accuracy over interpretability, making it challenging to extract meaningful insights when the data-generating process follows a specific underlying structure and an ordinal relationship among clusters is suspected. This work introduces K-Models, a novel framework that integrates ordinal constraints and estimates key underlying elements of the random process generating the observed functional profiles, improving both interpretability and structure identification. The proposed method is evaluated through simulations and real-world applications. In particular, it is tested on Region of Interest (ROI) curves, which represent reaction profiles from a reflectometric sensor monitoring biomolecular interactions, such as antigen-antibody binding. These curves represent changes in reflected light intensity over time at multiple measurement spots with immobilized antigens during analyte exposure, capturing the binding dynamics of the system. The goal is to identify intrinsic signal patterns solely from the observed dynamics, making this dataset an ideal benchmark for assessing the added interpretability of the proposed approach. By incorporating structural assumptions into the

clustering process, K-Models enhances interpretability while maintaining performance comparable to state-of-the-art techniques, providing a valuable tool for analyzing functional data with an underlying ordinal structure.

1 Introduction

Functional data has become increasingly widespread in recent years, with applications in fields such as engineering [Zhou et al. \(2011\)](#), medicine [Ruan et al. \(2022\)](#); [Movahed et al. \(2021\)](#), psychometry [Li et al. \(2024\)](#), and many others. As a result, Functional Data Analysis (FDA, [Ramsay and Silverman \(2005\)](#); [Kolmogorov and Fomin \(1961\)](#); [Horváth and Kokoszka \(2012\)](#)) has gained significant attention, earning a central role in statistical research. For functional data we often, but not always, intend a set of curves defined in an infinite-dimensional space, where each curve represents a random function, which in turn is seen as the realisation of an underlying random process. Within this framework, functional clustering has also emerged as topic of interest, leading to the development of new methods aimed to partitioning curves based on their similarities [Zhang and Parnell \(2023\)](#). Formally, functional clustering refers to the unsupervised task of grouping a set of curves in distinct subsets, named clusters, in such a way that instances within a group are similar to each other while they are dissimilar to instances of other groups (see for instance [Ruhao et al. \(2022\)](#)). Various methods have been developed, each leveraging different characteristics of functional data. Given this broad range of techniques available, several classification schemes have been proposed in order to have a systematic description, such as those by [Jacques and Preda \(2014\)](#) or [Zhang and Parnell \(2023\)](#). Among the main categories, we can identify:

- *Raw data* methods, which cluster functions based on their observed values on a discrete grid of their domain. These approaches directly compare the functional observations without any prior transformation, making them conceptually straightforward. A well-known example is K-Means ([Oti et al. 2021](#); [Bock 2007](#)), a widely used algorithm which can be easily adapted to various settings.
- *Filtering* methods, which consist of two phases: the first phase involves approximating curves using a basis expansion, which leads to a finite-dimensional representation, and the second phase involves performing clustering on the resulting finite-dimensional objects. One such method, introduced by [Ruhao et al. \(2022\)](#), begins by applying functional Principal Component Analysis (fPCA), projecting the functions onto principal component curves to obtain principal scores, which are subsequently used for classical multivariate clustering.
- *Distance-based* methods, which rely on functional distance metrics to quantify similarities between curves. In reality, this category is overlapping with the previous two, as the choice of distance metric determines whether the method aligns more closely with raw data approaches or filtering approaches. An example is the PAM (Partitioning Around Medoids, [Kaufman and Rousseeuw \(1990\)](#); [Schubert and Rousseeuw](#)

(2019)) algorithm, or even K-Means itself, since they both are based on dissimilarity metrics that work directly with the observed values of the curves.

While existing functional clustering methods are designed to achieve well-separated groups, they rarely focus on uncovering the mechanisms driving the observed curves. We encountered this limitation when analysing a biosensor dataset of time-resolved biosensor signals, recording the interaction of an antibody with sensor spots functionalised with different antigen concentrations. The curves share a common pattern – an initial stable phase followed by a rapid decline – but differ in intensity. This variation is driven by differences in the immobilised antigen concentrations, since a single antibody concentration is applied uniformly across all antigens within each assay. The observed curves therefore reflect an ordinal latent structure, in which cluster membership indexes the magnitude of a common underlying effect rather than qualitatively distinct processes.

Traditional clustering methods, although capable of producing meaningful partitions, do not explicitly account for this ordinal nature. They divide the data into distinct groups but fail to estimate the latent components influencing curve generation. This challenge led to the development of K-Models, a new clustering approach tailored for functional data with an underlying ordinal structure. These curves were previously analyzed in Patanè et al. (2026) in a supervised setting with nominal labels; the present work extends that analysis to an unsupervised framework, explicitly accounting for the ordinal structure of the data. The proposed method groups functions based on a common latent effect, distinguishing clusters by the intensity with which this effect manifests. Additionally, it provides estimates of the underlying function and cluster-specific coefficients, offering a more structured and interpretable representation of the data.

This study is so organised as follows: in Section 2 we review well-established techniques used for comparison. Section 3 introduces the problem and presents our novel clustering method. Section 4 describes the validation metrics employed to assess clustering performance. Section 5 applies the methods to simulated datasets using a Monte Carlo approach. Finally, in Section 6 we evaluate the methods on a real-world datasets focusing specifically on biosensor signals, the primary motivation for the development of K-Models.

2 State-of-the-art methods

The key idea underlying our approach is that the observed functional data are generated by a latent functional phenomenon whose intensity varies progressively across groups, inducing a natural ordering among clusters. A wide range of clustering methods have been developed for functional data, primarily aimed at partitioning curves into homogeneous groups. Among these, we consider three widely used approaches that represent standard benchmarks in the functional clustering literature: K-Means, which partitions data by minimizing within-cluster variance; Partitioning Around Medoids (PAM), a medoid-based alternative more robust to outliers; and Functional Principal Component Analysis (fPCA) combined with K-Means, which reduces data dimensionality before clustering. These methods are well-established, interpretable, and flexible

enough to be applied across a broad range of functional data settings. Unlike our method, however, they do not incorporate any ordinal structure among clusters, nor do they provide a direct characterisation of each group in terms of the underlying functional shape.

Formally, we consider a set of functional observations $\{x_i(t)\}_{i=1}^n$, where each function takes values in \mathbb{R} and is defined on a common domain $\mathcal{T} \subset \mathbb{R}$. Given a predefined number of clusters K , the objective is to find an optimal partition $\mathcal{C}^* = \{\mathcal{C}_1^*, \dots, \mathcal{C}_K^*\}$ that divides the curves into K ordered groups. We assume that the clusters admit a natural ordering $\mathcal{C}_1 < \dots < \mathcal{C}_K$, induced by the progressive amplification of a shared functional effect across levels: each successive cluster exhibits a stronger expression of this component, so that the ordering reflects an intrinsic gradation in the underlying phenomenon rather than an arbitrary labelling.

2.1 Functional K-Means

K-Means is one of the most widely used clustering algorithm. As highlighted by [Bock \(2007\)](#), its earliest formulation dates back to 1950, although the formal definition was established later. Over the years, various adaptations of K-Means have been proposed, but they all share a common underlying principle: maximizing the within-cluster similarity between data points and their corresponding cluster centroids, where a centroid is defined as a representative element of the group. This is achieved through an iterative procedure that assigns each data point to the cluster with the closest centroid according to a predefined distance metric. The process ultimately seeks to minimise the within-cluster variance, thereby producing an optimal partition of the dataset.

A rigorous mathematical formulation is provided by [Bock \(2007\)](#), which can be adapted to functional setting by solving the following minimisation problem:

$$(\mathcal{C}^*, \mathcal{Z}^*) = \arg \min_{\mathcal{C}, \mathcal{Z}} \sum_{k=1}^K \sum_{x_i \in \mathcal{C}_k} d^2(x_i, z_k), \quad (1)$$

where \mathcal{C} is the clustering partition, K the number of clusters, $\mathcal{Z} = \{z_1, \dots, z_K\}$ represents the set of centroids, and $d(\cdot, \cdot)$ is a measure of dissimilarity (or equivalently, similarity) between two functions. Moreover, to mitigate the sensitivity to initialisation, a multiple-start strategy is commonly employed. From Equation (1) it is evident that K-Means is a highly flexible approach since it imposes no strict assumptions on the data structure and allows for the use of different dissimilarity measures d , making it adaptable to various contexts.

When dealing with functional data, one of the most commonly used metrics, and the one adopted in this study, is the L^2 -norm – i.e. for $x \in L^2(I)$, $\|x\|_{L^2} = \int_I x(t) dt$ – :

$$d(x_i, z_k) := \|x_i - z_k\|_{L^2}. \quad (2)$$

This measure ensures that clustering is based on the overall shape of the functions rather than specific pointwise differences, making it a suitable choice for functional data.

Despite its conceptual simplicity, functional K-Means inherits several well-known limitations from its multivariate counterpart, including sensitivity to initialisation and an implicit assumption of "spherical" clusters.

2.2 Functional PAM

Partitioning Around Medoids (PAM, [Kaufman and Rousseeuw \(1990\)](#)), also known as K-Medoids, is another widely used clustering algorithm. While its core idea is similar to that of K-Means, a key distinction lies in the definition of the representative element of a cluster. As discussed in Section 2.1, K-Means defines centroids as the mean of all elements within a cluster. While this is an effective representative in Euclidean spaces, it may not be the most suitable choice for other dissimilarity measures.

In contrast, PAM represents each cluster with a medoid, which is the data point that minimises the average (or equivalently, the total sum of) dissimilarities to all other points in the same cluster. Formally, given a cluster C_k , we define its medoid as:

$$m_k := \arg \min_{x_i \in C_k} \sum_{x_j \in C_k} d(x_i, x_j). \quad (3)$$

This formulation highlights a key difference from K-Means: while a centroid z_k can be any point in the data space, a medoid m_k must be an actual data point from the dataset. This property makes K-Medoids more robust to outliers and allows its application to any arbitrary dissimilarity measures.

Similar to K-Means, the K-Medoids method relies on minimizing an absolute error criterion, commonly referred to as total deviation (TD), which is defined as:

$$TD := \sum_{k=1}^K \sum_{x_i \in C_k} d(x_i, m_k), \quad (4)$$

where $d(\cdot, \cdot)$ is a dissimilarity measure, while $\{m_1, \dots, m_K\}$ represents the set of medoids for the partition $\{C_1, \dots, C_K\}$.

In this study, the PAM algorithm is implemented by using the L^2 -norm as dissimilarity measure for functional data. By adopting the L^2 -norm, the method ensures a clustering structure that reflects meaningful variations in the underlying functional patterns while preserving the robustness of the PAM framework.

2.3 Functional PCA combined with K-Means

FPCA combined with K-Means – drawing inspiration from [Peng and Müller \(2008\)](#) and [Ruhao et al. \(2022\)](#) – first reduces the dimensionality of the functional dataset using Functional Principal Component Analysis (fPCA), and only then performs the clustering. A key advantage of this method is that it allows clustering to be performed directly in the directions of greatest variability in the dataset, potentially leading to more meaningful partitions.

Let X be a stochastic process in $L^2(\mathcal{T})$. By exploiting the Karhunen-Loève expansion, X can be expressed as:

$$X(t) = \mu(t) + \sum_{l=1}^{\infty} \alpha_l \phi_l(t) \quad \forall t \in \mathcal{T}, \quad (5)$$

where $\mu(t) := \mathbb{E}[X(t)]$ is the mean function, $\{\phi_l\}_{l=1}^{\infty}$ are the principal functions and $\{\alpha_l\}_{l=1}^{\infty}$ the corresponding principal component scores, defined as:

$$\alpha_l := \int_{\mathcal{T}} (X(t) - \mu(t)) \phi_l dt \quad l = 1, 2, \dots \quad (6)$$

Retaining only the scores associated with the first p , such that the cumulative proportion of variance retained is, for instance, at least 95% principal components, we obtain a new multivariate dataset of dimensions $n \times p$, where each curve is now represented by the corresponding vector of scores, $\{\alpha_i\}_{i=1}^n$, with $\alpha_i \in \mathbb{R}^p$.

At this point, clustering is performed by applying the classical multivariate K-Means. The only modification lies in the choice of the dissimilarity measure: instead of using the L^2 -norm between functional observations, clustering is now based on the Euclidean distance between the extracted score vectors. This reformulation simplifies the problem by shifting from an infinite-dimensional functional space to a finite-dimensional multivariate setting, while still capturing the dominant modes of variability in the data.

3 K-Models

The aim of this work is to propose a new method for the unsupervised clustering of functional data. In general, clustering seeks to partition observations into distinct subgroups such that instances within the same group are highly similar, while those belonging to different groups are as dissimilar as possible.

In the context of functional data, clustering is particularly useful for identifying homogeneous and heterogeneous patterns among curves and for revealing potential structures in the underlying stochastic process. The notion of clustering's optimality depends on the clustering criterion adopted, which varies across methods according to the type of similarity (or dissimilarity) being emphasised.

Most existing clustering methods for functional data focus on identifying groups that differ in shape or location, treating clusters as nominal categories without any inherent structure between them. In contrast, the method proposed in this work introduces an additional structural assumption: clusters are not only distinct but also *ordered*. In other words, we aim to perform *ordinal clustering*, where the clusters represent increasing levels of a latent functional effect.

The objective of the method is therefore twofold: (i) to estimate the latent functional direction capturing the shared phenomenon, and (ii) to partition the curves according to the degree to which they amplify this phenomenon. Clusters thus correspond to different *levels* of the same underlying functional effect rather than to

unrelated groups of curves. This interpretation naturally yields an ordered structure among clusters.

Under this framework, we consider the following cluster-specific functional model:

$$x_i(t) = \mu_0(t) + \mu_k(t) + \varsigma_k \beta(t) + \varepsilon_i(t), \quad t \in \mathcal{T}, \quad (7)$$

subject to $\|\beta\|_{L^2} = 1$ and $\langle \mu_k, \beta \rangle = 0$, where x_i belongs to cluster \mathcal{C}_k , with $k \in \{1, \dots, K\}$. Here, $\mu_0(t)$ denotes the global mean function, $\mu_k(t)$ represents the deviation of cluster \mathcal{C}_k from the global mean, and ς_k is a scalar coefficient measuring the amplitude of the shared functional direction $\beta(t)$ within cluster \mathcal{C}_k . The function $\beta(t)$ captures a common functional component that is expressed with different intensities across clusters, while $\varepsilon_i(t)$ is a zero-mean random noise process.

The ordinal structure arises naturally from this decomposition: we assume $\mathcal{C}_i < \mathcal{C}_j$ whenever $\varsigma_i < \varsigma_j$, so that the ordering among clusters is fully encoded in the scalar coefficients ς_k . The constraint $\|\beta\|_{L^2} = 1$ fixes the scale of the ordinal component, while $\langle \mu_k, \beta \rangle = 0$ ensures a unique decomposition into an ordinal part $\varsigma_k \beta$ and a cluster-specific deviation μ_k .

The ordinal nature of the latent classes is determined by the values of the coefficients ς_k , which induce a hierarchical ordering of the clusters according to increasing levels of amplification of $\beta(t)$. There exist two equivalent orderings, corresponding to the two possible sign choices of β (and ς_k); these lead to reversed but otherwise identical interpretations of the clustering structure. In practice, one can select the orientation that best supports scientific communication and interpretability.

Therefore, unlike conventional clustering methods that simply partition the data into unrelated groups, our approach yields clusters that are intrinsically ordered, providing a natural and interpretable ranking of the curves according to the strength of the latent functional effect. The clustering procedure is formulated as an optimisation problem aimed at minimizing the within-cluster variability along the latent ordinal direction. Specifically, we define the objective function as

$$\mathcal{L}((\varsigma, \boldsymbol{\beta}), \mathcal{C}) = \sum_{k=1}^K \sum_{i \in \mathcal{C}_k} (\langle x_i - \mu_0, \beta \rangle - \varsigma_k)^2 + \lambda \mathcal{J}(\beta). \quad (8)$$

where the component μ_k does not appear due to the constraint $\langle \mu_k, \beta \rangle = 0$, and where the term $\mathcal{J}(\cdot)$ is a roughness penalty, ensuring that the estimated latent functions remain smooth and interpretable. In this study, we impose the penalty on the integral of the second derivative; namely, given a function $\gamma \in L^2$:

$$\mathcal{J}(\gamma) = \int_{\mathcal{T}} (\gamma''(s))^2 ds. \quad (9)$$

This loss function measures the within-cluster dispersion of the projections of the curves onto the direction β . By minimizing this quantity, the method identifies clusters whose members exhibit similar levels of amplification of the latent functional

effect. Consequently, the procedure performs *ordinal clustering*, organizing the functional observations along a latent ordered dimension, rather than producing a generic partition of the data as in classical approaches such as functional K -means.

The regularisation parameter λ controls the trade-off between model fit and smoothness. A higher value of λ enforces stronger regularisation, favoring smoother estimates of the functional parameters, at the cost of a potentially higher approximation error within clusters.

By minimizing this objective function, the method achieves a partitioning of the curves where each group exhibits a different level of the underlying phenomenon, while ensuring that β remains a smooth and well-structured function. This approach leads to an interpretable clustering solution that reflects a natural hierarchical ordering among the identified subgroups.

3.1 Implementation

The proposed clustering procedure iteratively alternates between the following two key steps:

1. Estimation of the functions $\mu_k(t)$ and quantity ς_k for each cluster independently, while the common functions $\mu_0(t), \beta(t)$ are calculated using the entire data set.
2. Updating the cluster assignments by grouping together curves that exhibit similar patterns, based on the estimated coefficients.

A former approach, [Conde et al. \(2024\)](#), proposes a clustering algorithm for functional data based on cluster-specific functional regression models. Our method differs in a fundamental aspect: in [Conde et al. \(2024\)](#), the observed curves x_i are modeled through a set of predefined functional covariates

$$\mathbf{X}_i(t) = [X_{i1}(t), \dots, X_{iL}(t)]^\top$$

$t \in \mathcal{T}$, which influence the estimation process. In contrast, our approach does not rely on any external covariates. Instead, we estimate both the latent function $\beta(t)$ and the cluster-specific parameters $(\mu_k(t), \varsigma_k)$ directly from the data, without any additional input variables. This ensures that the clustering is driven purely by the intrinsic shape of the observed curves.

Algorithm 1 implements the proposed K-Models procedure for ordinal clustering of functional data. Each functional observation $x_i(t)$ and the latent direction $\beta(t)$ are represented in a common basis expansion

$$x_i(t) = \sum_{j=1}^{n_J} a_{ij} \phi_j(t), \quad \beta(t) = \sum_{j=1}^{n_J} b_j \phi_j(t),$$

where $a_i = (a_{i1}, \dots, a_{in_J})^\top$ and $b = (b_1, \dots, b_{n_J})^\top$ denote the corresponding coefficient vectors and $\{\phi_j(t)\}_{j=1}^{n_J}$ is a chosen functional basis. Coherently with this finite structure, we define two useful matrices:

- $D_2 = \left\{ \int_I \frac{\partial^2 \phi_j(t)}{\partial t^2} \frac{\partial^2 \phi_k(t)}{\partial t^2} dt \right\}_{j,k}$
- $D_0 = \left\{ \int_I \phi_j(t) \phi_k(t) dt \right\}_{j,k}$

This representation allows all computations to be carried out in the finite-dimensional coefficient space, while preserving the functional nature of the data.

We initialise the clusters randomly, to enable exploration of possible clustering configurations, and $\beta(t)$ through fPCA, to suggest a plausible quantitative – ideally ordinal – direction. The algorithm alternates between three main steps in each iteration. First, given the current latent direction $\beta(t)$, cluster-specific parameters are estimated: the cluster means in the coefficient space \bar{a}_k , the cluster-specific intercepts β_{0k} , and the ordinal amplification coefficients ς_k . The amplification coefficient ς_k represents the strength with which the common latent effect $\beta(t)$ manifests in cluster k .

Second, the latent direction $\beta(t)$ is updated using the current cluster assignments and amplification coefficients. The update Equation (10) is proved in Appendix 7. This step ensures that $\beta(t)$ captures the direction of variation associated with the ordinal latent effect.

Third, curves are reassigned to clusters based on the proximity of their projection onto $\beta(t)$ to the current cluster levels ς_k . Specifically, the projection score of each curve is compared to each cluster’s amplification coefficient, and the curve is assigned to the cluster minimizing the squared distance. After reassignment, clusters are reordered to enforce the ordinal structure $\varsigma_1 \leq \dots \leq \varsigma_K$.

To guarantee identifiability, β is L^2 -normalised by imposing $\|\beta\|_{D_0} = 1$.

To prevent empty or degenerate clusters, a repopulation safeguard is applied: if any cluster contains fewer than a minimum number of curves, a small number of curves are randomly reassigned to that cluster. This mechanism ensures stability of the alternating optimisation and avoids convergence to trivial solutions.

Finally, the procedure can be combined with a multi-start strategy to improve robustness. Multiple runs with random initialisations are performed, and the clustering solution with the highest quality score (e.g., a ratio of between-cluster to within-cluster variation) is retained as the final output. The algorithm iterates until the cluster assignments converge, producing the latent function $\beta(t)$, the ordered cluster amplification coefficients $\{\varsigma_k\}$, and the final cluster partition \mathcal{C} .

Algorithm 1: K-Models: Ordinal Functional Clustering

- 1 **Input:** Functional data $\{x_i(t)\}_{i=1}^n$, number of clusters K , maximum iterations I_{\max}
2 **Basis representation:** Let $\{\phi_j(t)\}_{j=1}^{n_J}$ be a basis of the functional space.

$$x_i(t) = \sum_{j=1}^{n_J} a_{ij} \phi_j(t), \quad \beta(t) = \sum_{j=1}^{n_J} b_j \phi_j(t), \quad \mu_k(t) = \sum_{j=1}^{n_J} c_{kj} \phi_j(t)$$

Denote by

$$\mathbf{a}_i = (a_{i1}, \dots, a_{in_J})^\top, \quad \mathbf{b} = (b_1, \dots, b_{n_J})^\top, \quad \mathbf{c}_k = (c_{k1}, \dots, c_{kn_J})^\top$$

the corresponding coefficient vectors.

3 **Initialisation:**

4 Center the data: $x_i(t) \leftarrow x_i(t) - (\frac{1}{n} \sum_{i=1}^n x_i(t))$. Save $\mu_0 = (\frac{1}{n} \sum_{i=1}^n x_i(t))$.

5 Initialise clusters $\{C_k\}$ randomly and estimate $\beta(t)$ using the first fPCA component.

6 **for** $it = 1, \dots, I_{\max}$ **do**

7 **Cluster update:**

8 **for** $k = 1, \dots, K$ **do**

9 Compute cluster mean $\bar{\mathbf{a}}_k = \frac{1}{n_k} \sum_{i \in C_k} \mathbf{a}_i$, where $n_k = |C_k|$ for any $k \in \{1, \dots, K\}$.

10 Compute cluster amplitude of the effect $\varsigma_k \leftarrow \bar{\mathbf{a}}_k^\top D_0 \mathbf{b}$;

11 $\mathbf{c}_k \leftarrow \bar{\mathbf{a}}_k - \varsigma_k \mathbf{b}$;

12 **Latent direction update:**

13 Find \mathbf{b} such that:

14

$$\left(\sum_{k=1}^K n_k \varsigma_k^2 D_0 + \lambda D_2 \right) \mathbf{b} \leftarrow \sum_{k=1}^K n_k \varsigma_k (\bar{\mathbf{a}}_k - \mathbf{c}_k) \quad (10)$$

15 **Ordinal enforcement:**

16 Order clusters such that $\varsigma_1 \leq \dots \leq \varsigma_K$.

17 **Reassignment:**

18 **for** $i = 1, \dots, n$ **do**

19 $\rho_i \leftarrow \frac{\mathbf{a}_i^\top D_0 \mathbf{b}}{\mathbf{b}^\top D_0 \mathbf{b}}$;

20 $C_i \leftarrow \arg \min_k (\rho_i - \varsigma_k)^2$

21 **Normalisation for identifiability:**

22

$$\mathbf{b} \leftarrow \mathbf{b} \cdot (\mathbf{b}^\top D_0 \mathbf{b})^{-1}$$

23 **Repopulation safeguard:**

24 If any cluster contains fewer than a minimum number of curves, randomly reassign a small set of observations to that cluster.

25 **Convergence check:**

26 Stop if cluster assignments do not change.

27 **Output:** $\beta(t)$, $\{\varsigma_k\}$, cluster partition \mathcal{C}

4 Validation in clustering

To assess the performance of the proposed clustering methods, we employ two widely used evaluation metrics: the Between-Cluster to Within-Cluster Variability Ratio and the Silhouette Score. These two measures provide complementary perspectives on clustering quality, offering both a global and a local evaluation of the resulting partitions. The first metric quantifies how well-separated the clusters are relative to their internal cohesion. It is defined as the ratio between the Between-Cluster Sum of Squares (BSS) and the Within-Cluster Sum of Squares (WSS). Given a partition $\mathcal{C} = \{C_1, \dots, C_K\}$

of the functional observations, the ratio is computed as follows:

$$\begin{aligned} \text{Ratio} &= \frac{\text{BSS}}{\text{WSS}}, \quad \text{with} \\ \text{BSS} &:= \sum_{k=1}^K n_k \|\mu_k - \mu_0\|_{L^2}^2, \\ \text{WSS} &:= \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|_{L^2}^2. \end{aligned} \tag{11}$$

Where $n_k = |C_k|$ denotes the number of curves in cluster k , $\mu_k(t) := \frac{1}{n_k} \sum_{x_i \in C_k} x_i(t)$ is the pointwise mean of the functions within cluster C_k , and $\mu_0(t) := \frac{1}{n} \sum_{i=1}^n x_i(t)$ is the global pointwise mean. A higher BSS/WSS ratio indicates well-separated clusters with high internal homogeneity, which are desirable properties in clustering.

The second metric used is the Silhouette Score. As described in [Rousseeuw \(1987\)](#), it evaluates clustering quality by balancing intra-cluster cohesion and inter-cluster separation. For each observation x_i , the Silhouette Score is defined as:

$$s_i := \frac{d_i - w_i}{\max(d_i, w_i)}, \quad i = 1, \dots, n. \tag{12}$$

where w_i is the average L^2 -distance between x_i and all other curves within the same cluster (a measure of intra-cluster similarity), while d_i is the smallest average L^2 -distance between x_i and all curves in a different cluster (a measure of inter-cluster separation). Then the overall Silhouette Score is computed as the mean of s_i across all observations:

$$S = \frac{1}{n} \sum_{i=1}^n s_i. \tag{13}$$

This score ranges between -1 and 1, with higher values corresponding to better partitions of the curves.

These two metrics evaluate clustering quality from different but complementary perspectives:

- **BSS/WSS Ratio:** provides a global measure of clustering performance by comparing the between-cluster sum of squares (BSS) to the within-cluster sum of squares (WSS). A higher ratio indicates better separation between clusters relative to their internal variability. This metric is naturally optimised by most clustering methods, but in multiple-start procedures it is also used to select the best solution among different initialisations.
- **Silhouette Score:** captures local clustering behavior by assessing how well each individual curve is assigned to its respective cluster. Specifically, it compares the average distance of a curve to other curves in the same cluster with the average distance to curves in the nearest neighboring cluster. This metric is particularly useful to identify potentially misclassified curves that lie near cluster boundaries and to compare the quality of different clustering methodologies.

By combining these two evaluation approaches, the BSS/WSS ratio ensures that clusters are compact and well-separated on a global scale, while the Silhouette Score validates the assignment consistency at an individual level. This dual perspective leads to a more robust assessment of the clustering solutions in functional data analysis.

Now, we present a simulation study designed to validate the K-models algorithm. The aim of the simulation study is to evaluate the ability of the method to correctly recover the underlying ordinal clustering structure under controlled settings. In particular, synthetic functional datasets are generated with known ordinal patterns and varying levels of noise and cluster separation, allowing us to systematically investigate the robustness and accuracy of the proposed method.

5 Simulation Study

To evaluate the empirical performance of the considered clustering approaches, we conduct a simulation study based on synthetic functional data generated from a realistic structure derived from the dataset of our case study. The aim of the study is to assess how accurately the different methods recover the underlying clustering structure under varying proportions of ordinal and global variability.

The simulated data are generated according to a functional model constructed from the empirical functional principal components of the original dataset. In particular, the mean function, the first principal component (PC1), and the second principal component (PC2) are taken directly from the case study. The second component is used to generate a global smooth variability shared across all observations, whereas the first component is used to introduce the cluster structure.

More precisely, let K denote the number of clusters and n the number of functional observations. Each curve is generated as a perturbation of the empirical mean function plus a linear combination of the first two functional principal components. The score associated with the second component is drawn from a Gaussian distribution representing global functional variability common to all curves. In contrast, the score associated with the first component is generated from cluster-specific Gaussian distributions whose means determine the cluster centers. Consequently, clusters differ primarily along the first principal direction.

To control the relative importance of the cluster structure with respect to the global variability, we introduce a parameter q that scales the variability associated with the first principal component. This parameter regulates the ratio between the ordinal (cluster-related) variability and the global variability while maintaining a clear separation between clusters. In the simulation study we consider the values

$$q \in \{0.025, 0.05, 0.1, 0.2\}.$$

As an illustrative example Figure 1 displays one example of simulated dataset per considered value of q .

For each configuration we generate multiple Monte Carlo samples and apply the clustering methods under comparison. Since the true cluster labels are known by construction, the performance of each method is evaluated by measuring the agreement between the estimated partition and the true clustering structure. The clusters are

constructed to guarantee a clear separation between groups, so that the true partition is well defined and not subject to ambiguity.

To quantify the agreement between the true and the estimated partitions, we compute the correlation between the true cluster labels and the estimated labels. In the case of K-models, the clusters are naturally ordered, enabling a direct computation of the correlation. For our competitors, we allow for label switching, considering all possible permutations of the estimated labels and retain the maximum correlation over these permutations. This ensures that the measure is invariant to the arbitrary labeling of clusters, and ensures a fair comparison with K-models. In Equation 14, we define the correlation estimated $\hat{\mathcal{C}}$ and true \mathcal{C} clustering configuration, for generic clustering methods.

$$\text{Corr}(\hat{\mathcal{C}}, \mathcal{C}) = \max_{\pi \in \text{Sym}^*(K)} \text{Corr}(k_i, \pi(\hat{k}_i)) \quad (14)$$

where $k_i = k \cdot \mathbb{1}_{i \in C_k}$, and $\text{Sym}(K)$ is the group of permutations on $\{1, \dots, K\}$. While alternative measures such as the Adjusted Rand Index (ARI) could also be used, we prefer this criterion because, in addition to being permutation invariant, it provides a more stringent notion of agreement. Indeed, by relying on the numerical values of the labels after optimal alignment, it captures how well the overall labeling structure matches the true one, rather than only pairwise group membership. This is particularly relevant in our simulation setting, where clusters are generated along a continuous latent direction. In Figure 2, the results are summarised using boxplots of the correlations obtained across simulation runs for each value of q and for each clustering method. This representation allows us to analyze how the performance of the competing approaches varies as the relative contribution of ordinal variability changes. In particular, K-Models approach consistently achieves high correlations even when q is small, indicating its ability to detect subtle cluster structures. By contrast, competing methods show more variability and generally lower correlations in these challenging scenarios. As q increases, the differences between methods tend to diminish, but the proposed approach maintains a robust performance across all considered values of q , illustrating its stability and reliability in recovering the true clustering structure.

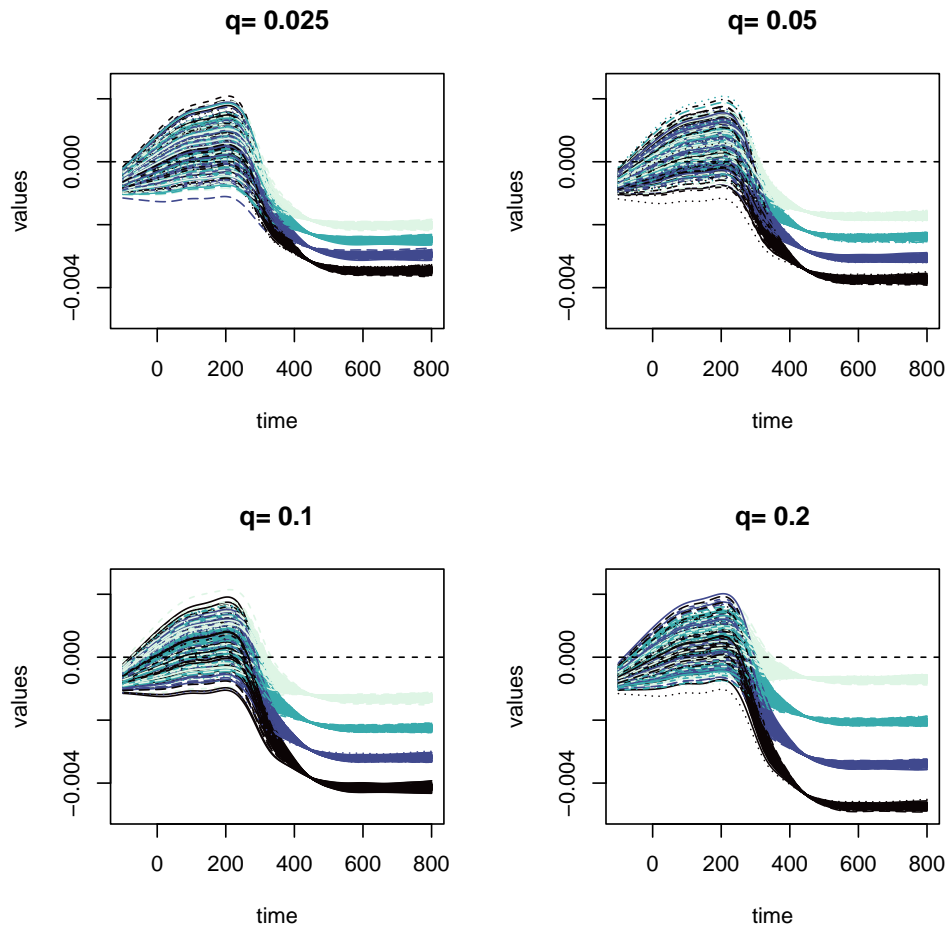


Fig. 1: Examples of simulated scenarios depending on q .

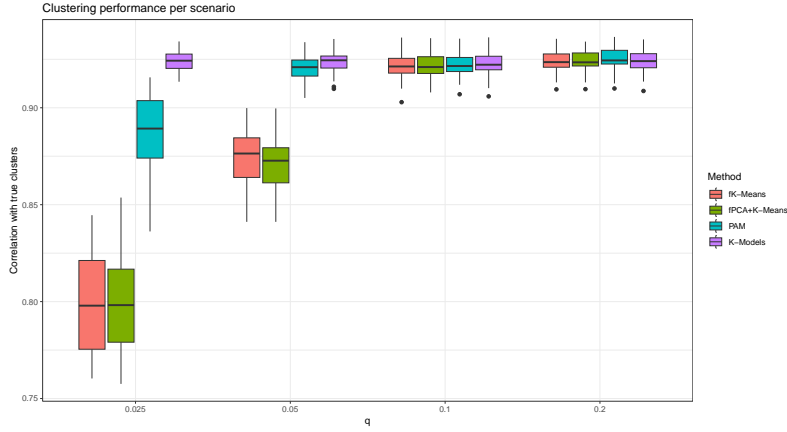


Fig. 2: Boxplots of the correlations obtained across simulation runs for each value of q and for each clustering method.

6 A case study: ordinal clustering of optical biosensor signals

In this section, we apply the proposed method to our case study, based on biosensor signals. The signals are treated as functional data obtained through the preprocessing procedure described in [Patanè et al. \(2026\)](#), where the same dataset is analyzed in a supervised setting.

The functional dataset, displayed in [Figure 3](#), consists of biosensor signal profiles from 1026 Regions of Interest (ROIs), i.e., there are 1026 spots with eight different antigen concentrations immobilized. Each signal represents the temporal evolution of light intensity within the corresponding ROI. As already discussed in [Patanè et al. \(2026\)](#), using an optical reflectometric sensor, light intensity is used as an indicator of interactions between antigens and antibodies [Gauglitz \(2010\)](#).

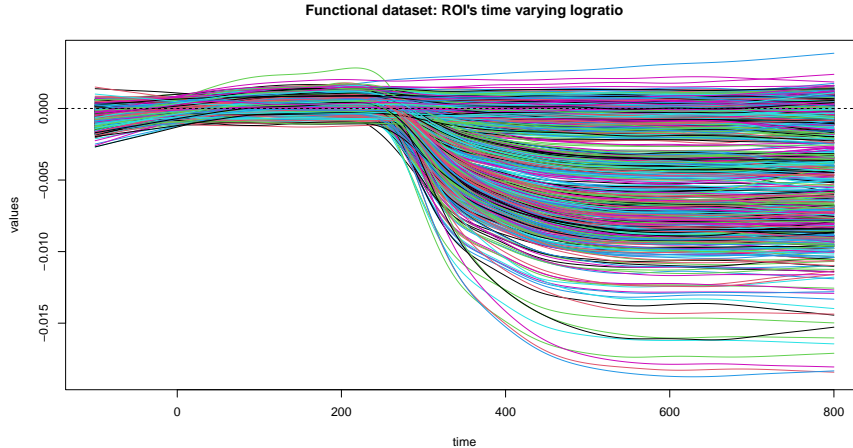


Fig. 3: Unlabeled ROI signal curves.

The K -models algorithm is applied to the biosensor signals by considering a number of clusters K ranging from 2 to 7. Figure 4 illustrates the resulting partitions of the dataset for different values of K . In this real-data application, both the underlying ordinal structure and the appropriate number of clusters are unknown. Therefore, model selection is performed using the Silhouette Score. Indeed, since the algorithm already maximises the between-cluster to within-cluster sum of squares ratio (BSS/WSS), the Silhouette Score provides a complementary criterion for assessing clustering quality. According to the Silhouette Score plot reported in Figure 5, the elbow rule suggests selecting $K = 4$ as the optimal number of clusters, fully signal-based. This differs from the 8 nominal concentration levels known from the experimental design – used to supervise the ordinal classification in Patanè et al. (2026) – indicating that the light intensity signal alone is not sufficient to discriminate all antigen concentrations, as some levels appear indistinguishable from a purely signal-based perspective.

Once K is fixed, the model outputs can be more fully leveraged for interpretation. In particular, the K -models provide an estimate of the ordinal behavior β (Figure 6). Moreover, the clusters identified by the method are naturally ordered according to the values of ς_k , which can be interpreted as different levels of amplification of β . This ordering—reversible upon changing the sign of β —is given by C_1, C_2, C_3, C_4 . Cluster C_1 corresponds to the lowest amplitude, likely representing an absent or negligible antibody–antigen interactions. At the opposite extreme, cluster C_4 captures a highly pronounced and heterogeneous class of larger signals representing higher antibody–antigen interactions. The remaining clusters represent intermediate levels of response intensity.

As illustrated in Figure 6, the ordinal behavior β clearly characterises the reaction profile: it exhibits a marked decay approximately between time 200 and 400, followed by a horizontal asymptote after time 600. This pattern indicates that the proposed

method successfully captures the ordinal latent structure of the data, revealing the underlying level-based organisation of the biosensor signals.

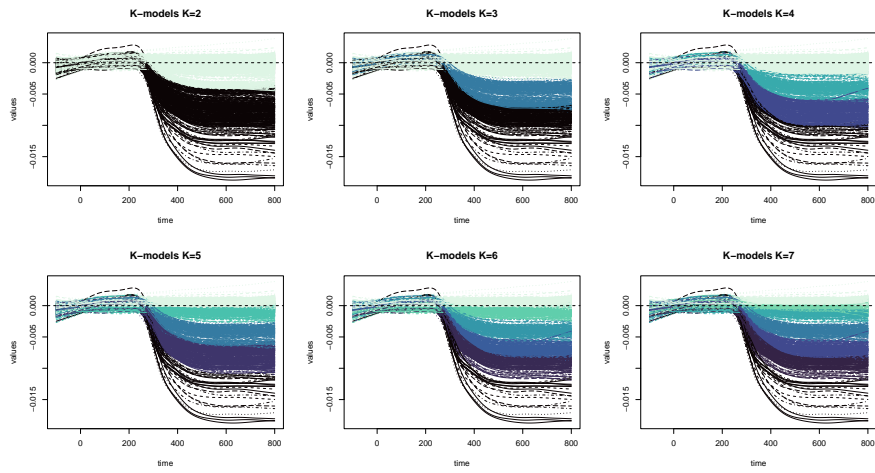


Fig. 4: The levels assigned to the biosensor signals by the K-models $K = 2, \dots, 7$.

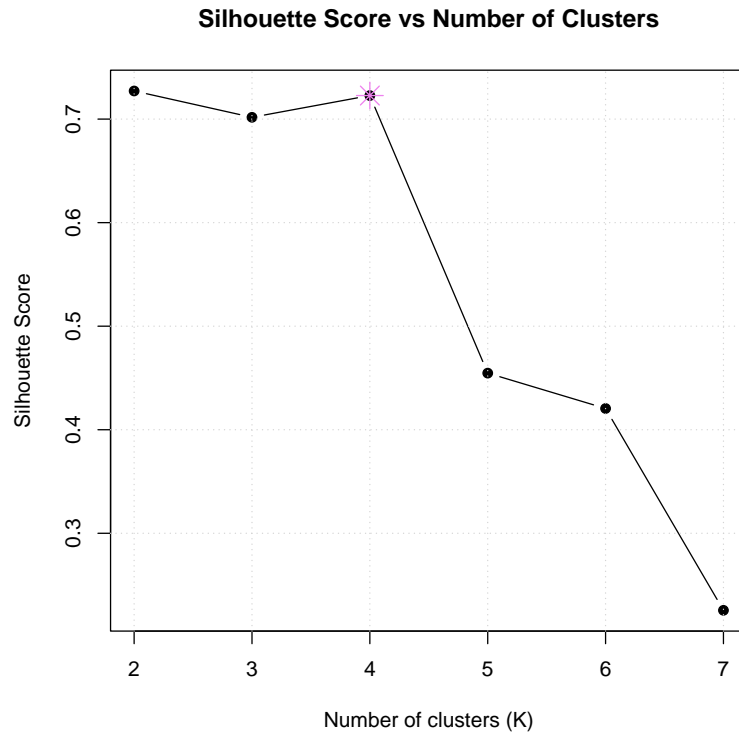


Fig. 5: Silhouette score of the projections of x_i on β , by the K -models $K = 2, \dots, 7$. We choose $K = 4$ according the elbow rule.

K-models, with K=4

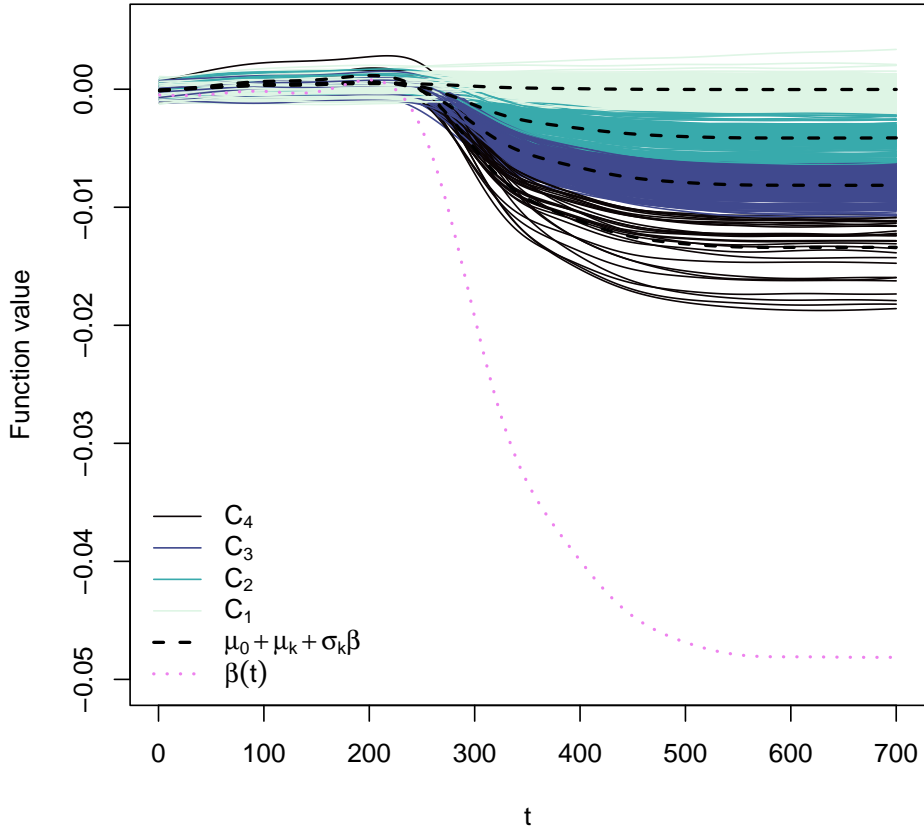


Fig. 6: The ordinal pattern β which increases with k according to ς_k .

7 Conclusions

The great variety of clustering methods for functional data in the existing literature highlights the significant interest that this field has received within the mathematical and engineering communities. Despite this, in several contexts – such as the biosensor signals presented in the case study – the existing techniques do not always provide the most suitable tools for addressing the desired analyses. In many cases, clustering procedures are needed not only to partition the curves, but also to offer a comprehension of the underlying mechanisms behind the generation of these functions, and how they contribute to the formation of distinct subgroups.

In this regard, the K-Models method proposed by this work allows for the simultaneous achievement of these two objectives. By defining a clustering framework based on a linear functional model, the K-Models approach enables the estimation of a latent phenomenon underlying the stochastic process that generated the observed curves. Furthermore, it evaluates how this shared effect manifests within each cluster through its corresponding linear coefficient. A key advantage of this formulation is that it defines an ordinal partition of the data, an absent feature in the traditional functional clustering procedures. The ordering of the clusters is indeed directly determined by the magnitude of the common effect, providing a structured interpretation of the underlying relationships within the dataset.

The simulation study presented in this work confirms these theoretical advantages. Across scenarios with subtle cluster-related variability, the K-Models approach consistently recovers the true clustering structure more accurately than competing methods, while maintaining the ordinal interpretation of the clusters. These results illustrate not only the reliability of the method in practical applications, but also its ability to offer a meaningful characterisation of the latent processes driving the observed functional data.

Acknowledgements

This research has received funding by the European Commission under the “HORIZON-CL4-2021-DIGITALEMERGING-01 project BioProS - Biointelligent Production Sensor to Measure Viral Activity” (grant agreement no. 101070120), 2022-2026”. GP, AM, LD and FN acknowledge the initiative “Dipartimento di Eccellenza 2023–2027”, MUR, Italy, Dipartimento di Matematica, Politecnico di Milano. LD acknowledges his membership to INdAM GNCS - Gruppo Nazionale per il Calcolo Scientifico (National Group for Scientific Computing), Italy. The authors thank Alessandro Ganelli and Johanna Hutterer for their support in this research.

Declarations and statements

Conflict of interest: the authors have no relevant financial or non-financial interests to disclose.

Proof of Equation (10): in the second step, where μ_k , ς_k and the clusters assignment are fixed, we aim at minimising the conditional loss \mathcal{L}_c , where

$$\mathcal{L}_c(\beta|\mu_k, \varsigma_k, \mathcal{C}_1, \dots, \mathcal{C}_k) = \sum_{k=1}^K \sum_{i \in \mathcal{C}_k} \|x_i - \mu_k - \varsigma_k \beta\|_{L^2}^2 + \lambda \mathcal{J}(\beta)$$

By discretising the problem,

$$\sum_{k=1}^K \sum_{i \in \mathcal{C}_k} \|x_i - \mu_k - \varsigma_k \beta\|_{L^2}^2 \approx \sum_{k=1}^K \sum_{i \in \mathcal{C}_k} (\mathbf{a}_i - \mathbf{c}_k - \varsigma_k \mathbf{b})' D_0 (\mathbf{a}_i - \mathbf{c}_k - \varsigma_k \mathbf{b}) + \lambda \mathbf{b}' D_2 \mathbf{b}$$

The gradient of the discretised \mathcal{L}_c with respect to \mathbf{b} is given by:

$$\begin{aligned}\nabla_{\mathbf{b}}\mathcal{L}_c(\mathbf{b}|\mathbf{c}, \boldsymbol{\varsigma}, \mathcal{C}_1, \dots, \mathcal{C}_K) &= D_0 \cdot \left(\sum_{k=1}^K \sum_{i \in \mathcal{C}_k} \varsigma_k (\mathbf{a}_i - \mathbf{c}_k - \varsigma_k \mathbf{b}) \right) + \lambda D_2 \cdot \mathbf{b} \\ &= \left(\sum_{k=1}^K n_k \varsigma_k^2 D_0 + \lambda D_2 \right) \mathbf{b} - \left(\sum_{k=1}^K n_k \varsigma_k (\bar{\mathbf{a}}_k - \mathbf{c}_k) D_0 \right)\end{aligned}$$

Imposing the gradient equal to 0, we obtain Equation (10).

□

References

- Bock HH (2007) Clustering Methods: A History of k-Means Algorithms, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 161–172. https://doi.org/10.1007/978-3-540-73560-1_15
- Conde S, Tavakoli S, Ezer D (2024) Functional regression clustering with multiple functional gene expressions. PLOS ONE 19(11):e0310991. <https://doi.org/10.1371/journal.pone.0310991>
- Gauglitz G (2010) Direct optical detection in bioanalysis: an update. Analytical and Bioanalytical Chemistry 398:2363–2372. <https://doi.org/10.1007/s00216-010-3904-4>
- Horváth L, Kokoszka P (2012) Inference for functional data with applications. Springer
- Jacques J, Preda C (2014) Functional data clustering: A survey. Advances in Data Analysis and Classification 8:231–255. <https://doi.org/10.1007/s11634-013-0158-y>
- Kaufman L, Rousseeuw PJ (1990) Partitioning around medoids (program PAM). In: Finding Groups in Data: An Introduction to Cluster Analysis. John Wiley & Sons, p 68–125, <https://doi.org/10.1002/9780470316801.ch2>
- Kolmogorov AN, Fomin SV (1961) Elements of the Theory of Functions and Functional Analysis. Graylock Press
- Li Y, Li B, Ding J, et al (2024) A novel framework for forecasting mental stress levels based on physiological signals. In: Neural Information Processing. ICONIP 2023, Communications in Computer and Information Science, vol 1963. Springer, Singapore, pp 287–297, https://doi.org/10.1007/978-981-99-8138-0_23
- Movahed RA, Pirzad GJ, Shahyad S, et al (2021) A major depressive disorder classification framework based on EEG signals using statistical, spectral, wavelet, functional connectivity, and nonlinear analysis. Journal of Neuroscience Methods 358:109209. <https://doi.org/10.1016/j.jneumeth.2021.109209>

- Oti EU, Olusola MO, Eze FC, et al (2021) Comprehensive review of K-means clustering algorithms. *International Journal of Advances in Scientific Research and Engineering* 7(8):64–69. <https://doi.org/10.31695/IJASRE.2021.34050>
- Patanè G, Nicolussi F, Krauth A, et al (2026) Functional-ordinal canonical correlation analysis with application to data from optical sensors. *Technometrics* 68(2):384–398. <https://doi.org/10.1080/00401706.2025.2584479>
- Peng J, Müller HG (2008) Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. *The Annals of Applied Statistics* 2(3):1056–1077. <https://doi.org/10.1214/08-AOAS172>
- Ramsay JO, Silverman BW (2005) *Functional Data Analysis*. Springer, New York
- Rousseeuw P (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Ruan H, Dai X, Chen S, et al (2022) Arrhythmia classification and diagnosis based on ECG signal: A multi-domain collaborative analysis and decision approach. *Electronics* 11(19):3251. <https://doi.org/10.3390/electronics11193251>
- Ruhao W, Bo W, Aiping X (2022) Functional data clustering using principal curve methods. *Communications in Statistics - Theory and Methods* 51(20):7264–7283. <https://doi.org/10.1080/03610926.2021.1872636>
- Schubert E, Rousseeuw PJ (2019) *Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms*, Springer International Publishing, pp 171–187. https://doi.org/10.1007/978-3-030-32047-8_16
- Zhang M, Parnell A (2023) Review of clustering methods for functional data. *ACM Transactions on Knowledge Discovery from Data* 17(7):1–34. <https://doi.org/10.1145/3581789>
- Zhou R, Serban N, Gebrael N (2011) Degradation modeling applied to residual lifetime prediction using functional data analysis. *Annals of Applied Statistics* 5(2B):1586–1610. <https://doi.org/10.1214/10-AOAS448>

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

- 38/2026** Clemente, A.; Arnone, E.; Mateu, J.; Sangalli, L.M.
Nonparametric estimators over metric graphs
- 37/2026** Centofanti, E.; Ziarelli, G.; Scacchi, S.; Pavarino, L.F.
A Neural Latent Dynamics Approach for Solving Inverse Problems in Cardiac Electrophysiology
- 36/2026** Botti, M.; Mascotto, L.; Mosconi, M.
A nonconforming method for a generalized Darcy-Forchheimer model
- 35/2026** Caon, B.; Corti, M.; Bonizzoni, F.; Antonietti, P.F.
High-fidelity and Network-based Spatio-temporal Mathematical Models of Alzheimer's Disease Progression and their Validation Against PET-SUVR Imaging Data
- 34/2026** Mancinelli, F. M.; Torzoni, M.; Maisto, D.; Donnarumma, F.; Corigliano, A.; Pezzulo, G.; Manzoni, A.
Multi-Agent Digital Twins for strategic decision-making using Active Inference
- 33/2026** Franzoni, G.; Mirabella, S.; Dabek, A.; Ferro, N.; Antona, A.; Carlessi, M.; Cinquemani, S.; Matteucci, M.; Cocetta, G.; Perotto, S.
Integrating Environmental Control and Hyperspectral Imaging to Assess Light and Nutrient Effects on Lettuce Post-Harvest Quality in Vertical Farming
- Franzoni, G.; Mirabella, S.; Dabek, A.; Ferro, N.; Antona, A.; Carlessi, M.; Cinquemani, S.; Matteucci, M.; Cocetta, G.; Perotto, S.
Integrating Environmental Control and Hyperspectral Imaging to Assess Light and Nutrient Effects on Lettuce Post-Harvest Quality in Vertical Farming
- 32/2026** Antonietti, P.F.; Bonizzoni, F.; Perugia, I.; Verani, M.
A Multilevel Monte Carlo Virtual Element Method for Uncertainty Quantification of Elliptic Partial Differential Equations
- 31/2026** Guastamacchia, C.; Piersanti, R.; Giardini, F.; Coppini, R.; Ferrantini C.; Dede' L.; Sacconi L.; Regazzoni F.
The functional impact of myofiber macroscopic organization and disarray in computational models of the murine heart
- 30/2026** Regazzoni, F.
The internal law of a material can be discovered from its boundary