
MOX-Report No. 49/2023

# A Spearman Dependence Matrix for Multivariate Functional Data

Ieva, F.; Ronzulli, M.; Romo, J.; Paganoni, A.M.

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

# A Spearman Dependence Matrix for Multivariate Functional Data

Francesca Ieva[1,3] | Michael Ronzulli[1] | Juan Romo[2] | Anna Maria Paganoni*[1]

**Abstract**

We propose a nonparametric inferential framework for quantifying dependence among two families of multivariate functional data. We generalize the notion of Spearman correlation coefficient to situations where the observations are curves gen-erated by a stochastic processes. In particular, several properties of the Spearman index are illustrated emphasizing the importance of having a consistent estimator of the index of the original processes. We use the notion of Spearman index to define the Spearman matrix, a mathematical object expressing the pattern of dependence among the components of a multivariate functional dataset. Finally, the notion of Spearman matrix is exploited to analyze two different populations of multivariate curves (specifically, Electrocardiographic signals of healthy and unhealthy people), in order to test if the pattern of dependence between the components is statistically different in the two cases.

**KEYWORDS:**
Spearman correlation coefficient, multivariate functional data, ECG signals.

[1]MOX Lab, Department of Mathematics, Politecnico di Milano, Milan 20133, Italy

[2]Department of Statistics, University Carlos III of Madrid, Madrid 28903 , Spain [3]Health Data Science Center, Human

Technopole, Milan 20157, Italy

## 1 | INTRODUCTION

Nowadays, the statistical analysis of complex and high dimensional data is experiencing a notable growth for application in different fields of science such as medicine, finance, criminology, quality control, and many others. This leads to rethink the way that classical statistics approaches the analysis of such data, since methodologies commonly implemented until now for both descriptive and inferential purposes are increasingly limited or inefficient. Data dimensionality often leads multivariate analysis to be not feasible and its results not easily interpretable. Functional Data Analysis (FDA) (see[1,2,3], among others, for complete overview) is clearly the main field of research in statistics which tried to overcome this issue. Despite the fact that several multivariate methods are not usually well suited for functional datasets, many multivariate techniques have inspired advances in FDA. For example, to quantify the relationship of dependence between two or more groups of functional data. In fact the investigation of the dependence among curves is relatively a new issue in statistics. We aim at providing a non parametric measure of dependence for families of multivariate curves, as well as a suitable corresponding inferential framework for testing the presence of dependency among components and possible differences among patterns of dependency. For instance in[4] the authors provided a generalization of the Pearson correlation coefficient for functional data that allows to quantify the dependence among two families of curves. This measure is called the concordance correlation coefficient and was used to evaluate the reproducibility of repeated-paired curve data. In[5] is defined a Kendall's $\tau$ coefficient for functions considering pre-orders that permit the sorting of the functional observations and the identification of the concordant and discordant pairs in a bivariate sample of curves. In this work, we will consider the definition of the Spearman index for functions introduced by[6] to set a proper inferential framework for assessing dependency among families of multivariate curves. The paper is organized as follows:

[0]**Abbreviations:** ANA, anti-nuclear antibodies; APC, antigen-presenting cells; IRF, interferon regulatory factor

**Correspondence**
*Anna Maria Paganoni, MOX Lab, Department of Mathematics, Politecnico di Milano, p.za Leonardo da Vinci 32, 20133 Milano, Italia. Email: anna.paganoni@polimi.it

Section 2 recalls some basic notions about depths, the definition of the Spearman index and Spearman Matrix and the statistical properties of the related sample estimators. Section 3 presents the whole inferential framework we propose for assessing the presence of dependency in h-variate functional data, and the related application to a real case study considering two populations of multivariate Electrocardiographic signals from healthy and unhealthy patients. Results are discussed in Section 4, together with possible further developments. All the analyses are carried out using[7]. Codes are embedded in the roahd package, detailed in[8].

## 2 | SPEARMAN INDEX IN THE FUNCTIONAL FRAMEWORK

Spearman index is a non-parametric measure of association between two random variables X and Y. It presents significant advantages over the classical Pearson correlation coefficient that quantifies linear dependence. In fact its sample version is less sensitive to outliers than the Pearson correlation coefficient, and it is able to capture also non linear dependences among two random variables. Let us consider $(X_1, Y_1), (X_2, Y_2)$ and $(X_3, Y_3)$ be three independent copies of the random vector (X,Y) with joint cumulative distribution function $F_{XY}$ and margins $F_X$ and $F_Y$, respectively. The Spearman index[9] between the variables X and Y, denoted by $\rho_s(X, Y)$, is defined as:

$$\rho_s(X, Y) = 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]) \tag{1}$$

The Spearman index is proportional to the difference between the probability of concordance and the probability of discordance for two vectors $(X_1, Y_1)$ and $(X_2, Y_3)$.
However, we are interested in the equivalent definition of $\rho_s$ by computing the Pearson correlation coefficient, indicated with $\rho_p$, between the random variables $U = F_X(X)$ and $V = F_Y(Y)$, that is:

$$\rho_s(X, Y) = \rho_p(U, V) = \frac{E(UV) - E(U)E(V)}{\sqrt{Var(U)}\sqrt{Var(V)}}. \tag{2}$$

$U$ and $V$ are called the *grades* of $X$ and $Y$. For this reason, the Spearman index is also called *the grade correlation coefficient*.
Suppose now to have two samples of size $n$ from the random variables $X$ and $Y$, say $\mathbf{x} = (x_1, x_2, ..., x_n)$, $\mathbf{y} = (y_1, y_2, ..., y_n)$. Consider the vectors of the estimated grades $\mathbf{u} = (u_1, u_2, ..., u_n)$ and $\mathbf{v} = (v_1, v_2, ..., v_n)$, defined evaluating each observation in the empirical cumulative distribution function of the corresponding sample. So, we have

$$u_i = \hat{F}_X(x_i) = \frac{1}{n}\sum_{j=1}^{n} I(x_j \leq x_i)\},$$
$$v_i = \hat{F}_Y(y_i) = \frac{1}{n}\sum_{j=1}^{n} I(y_j \leq y_i)\},$$

for $i = 1, ..., n$. Notice that $u_i$ (resp. $v_i$) can be interpreted as the relative position of the observation $x_i$ (resp. $y_i$) in the set $\mathbf{x}$ (resp. $\mathbf{y}$).
The sample version of the Spearman index is defined as the sample Pearson correlation coefficient of $\mathbf{u}$ and $\mathbf{v}$:

$$\hat{\rho}_s(\mathbf{x}, \mathbf{y}) = \hat{\rho}_p(\mathbf{u}, \mathbf{v}) = \frac{\sum_{i=1}^{n}(u_i - \bar{u})(v_i - \bar{v})}{(\sum_{i=1}^{n}(u_i - \bar{u})^2 \sum_{i=1}^{n}(v_i - \bar{v})^2)^{\frac{1}{2}}}, \tag{3}$$

where $\bar{u}$ and $\bar{v}$ stand for the sample means of $\mathbf{u}$ and $\mathbf{v}$, respectively.

The Spearman index $\rho_s$ defined for random variables can be extended, in a rigorous way, to the case of two stochastic processes $X_t$ and $Y_t$[6], on the basis of (2) and the notion of grade of a stochastic process $X_t$ with respect to another process $Z_t$ (see[10,11]). Let $C(I)$ be the space of the continuous functions defined in a compact interval $I$ and consider a stochastic process $X_t$, with distribution $\mathcal{L}$ and sample paths in $C(I)$.

**Definition 1.** Let $X_t$ and $Z_t$ be two stochastic processes. Then,

$$IL-grade(X_t)_{Z_t} = \frac{1}{\lambda(I)} E_{Z_t}[\lambda\{t \in I \,:\, X_t \geq Z_t\}],$$
$$SL-grade(X_t)_{Z_t} = \frac{1}{\lambda(I)} E_{Z_t}[\lambda\{t \in I \,:\, X_t \leq Z_t\}],$$

where $\lambda$ stands for the Lebesgue measure on $R$. Consider now a functional dataset $x_1(t), ..., x_n(t)$, with $t \in I$, composed by $n$ realizations of the process $X_t$. If we fix any curve $x = x(t)$ of the dataset, the sample version of both *IL-grade* and *SL-grade* can be easily obtained by substituting the expectation with the sample mean as follows:

$$IL_n-grade(x) = \frac{1}{n\lambda(I)} \sum_{i=1}^{n} \lambda\{t \in I \,:\, x(t) \geq x_i(t)\},$$
$$SL_n-grade(x) = \frac{1}{n\lambda(I)} \sum_{i=1}^{n} \lambda\{t \in I \,:\, x(t) \leq x_i(t)\}.$$

$IL_n-grade(x)$ and $SL_n-grade(x)$ quantify the relative position of $x$ with respect to the other curves of the sample, therefore they represent two possible generalizations of the notion of grade in the infinite dimensional framework. The sample version of the Inferior and Superior Length grade provide an effective way for ordering a set of curves. In fact, we can give the following criterion:

**Definition 2.** Consider the functional dataset $x_1(t), ..., x_n(t)$, $t \in I$, composed by $n$ realizations of a stochastic process $X_t$. Then,

$$x_i(t) \preceq x_j(t) \iff IL_n-grade(x_i) \leq IL_n-grade(x_j), \tag{4}$$

The alternative definition can be deduced by replacing the $IL_n-grade(x)$ with $SL_n-grade(x)$. The relation given by Definition 2 meets important properties such as reflectivity and transitivity, but, unfortunately, it does not satisfy the antisymmetry property. Therefore, the relation introduced is a pre-order, which is less restrictive than a partial order and allows to compare any pair of functions in the sample. Notice that if the curves do not cross each other, Definition 2 corresponds to the pointwise order.

## 2.1 | Spearman index for bivariate functional data

According to equation (2), as in[6] we define the Spearman index for two stochastic processes as the Pearson correlation coefficient between the random variables $IL-grade(X_t)$ and $IL-grade(Y_t)$, as follows:

**Definition 3.** (*Spearman index for stochastic processes*) Let $(X_t, Y_t)$ be a stochastic process with law $\mathcal{L}$ taking values on the space $C(I;R^2)$ of the continuous functions $(f(t), g(t)) : I \to R^2$, with $I$ a compact interval of R. The Spearman index for $(X_t,Y_t)$ is defined as

$$\rho_s(X_t, Y_t) = \rho_p((IL-grade(X_t), IL-grade(Y_t)), \tag{5}$$

where $\rho_p$ denotes the Pearson correlation coefficient and *IL-grade*$(\cdot)$ is the grade associated to a stochastic process, as in Definition 2.

The corresponding sample version is the following:

**Definition 4.** (*Spearman index for stochastic processes*)

Consider the bivariate functional dataset,

$$\begin{bmatrix}\mathbf{x} & \mathbf{y}\end{bmatrix} = \begin{bmatrix} x_1(t) & y_1(t) \\ x_2(t) & y_2(t) \\ \vdots & \vdots \\ x_n(t) & y_n(t) \end{bmatrix}_{t\in I}$$

composed by $n$ realizations of the stochastic process $(X_t,Y_t)$ as above. Then, the sample Spearman index, denoted by $\hat{\rho}_s(\mathbf{x}, \mathbf{y})$, is defined as

$$\hat{\rho}_s(\mathbf{x}, \mathbf{y}) = \hat{\rho}_p(IL_n-grade(\mathbf{x}), IL_n-grade(\mathbf{y})), \tag{6}$$

where $\hat{\rho}_p$ is the sample Pearson correlation coefficient and

$$IL_n-grade(\mathbf{x}) = (IL_n-grade(x_1), IL_n-grade(x_2), ..., IL_n-grade(x_n)),$$
$$IL_n-grade(\mathbf{y}) = (IL_n-grade(y_1), IL_n-grade(y_2), ..., IL_n-grade(y_n)).$$

An alternative definition of the Spearman index for functions can be obtained by replacing $IL_n-grade$ by $SL_n-grade$.

We can prove some asymptotic properties of the Spearman coefficient from the fact that it can be expressed as a UB-statistic. Let $\boldsymbol{B}$ be a real separable Banach space. A UB-Statistic[12] is defined as

$$U_n = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \ldots < i_m \le n} \Phi\{(X_{i_1}, ..., X_{i_m})\} \tag{7}$$

where $\Phi : X^m \to \boldsymbol{B}$ of m variables given on $X^m$ and taking values in $\boldsymbol{B}$, is an integrable symmetric function (kernel).

**Definition 5.** (*Functional* $\rho$) If $(X_1, Y_1)$, $(X_2, Y_2)$ and $(X_3, Y_3)$ are copies of a bivariate stochastic process $\{(X_t, Y_t) : t \in I\}$, the population version of this dependence measure is

$$\rho = 6[P(X_1 \prec X_2, Y_1 \prec Y_3) + P(X_2 \prec X_1, Y_3 \prec Y_1)] - 3. \tag{8}$$

where $\prec$ is the preorder $\preceq$ defined in 4 in the case without considering ties. Consider a sample $(x_1, y_1), ..., (x_n, y_n)$ of a two-dimensional random process $(X, Y) = \{(X_t, Y_t) : t \in I\}$ within the compact interval $I$, with $X, Y \in C(I)$. Then Spearman's extended correlation coefficient is estimated by the empirical version:

$$\hat{\rho}_n = \binom{n}{3}^{-1} \sum_{1 \le i_1 < i_2 < i_3 \le n} 6I(x_{i_1} \prec x_{i_2}, y_{i_1} \prec y_{i_3}) + 6I(x_{i_2} \prec x_{i_1}, y_{i_3} \prec y_{i_1}) - 3. \tag{9}$$

Now, consider $(X_1, Y_1), ..., (X_n, Y_n)$ to be independent copies of the bivariate stochastic process (X,Y) with identical distribution $P$ and whose realizations or paths are pairs of functions that take values in the measurable space $(C[a, b] \times C[a, b], \chi)$.
Then, the functional $\hat{\rho}$ given in (9) can be expressed as a UB-statistic,

$$U_n = \binom{n}{3}^{-1} \sum_{1 \le i_1 < i_2 < i_3 \le n} \Phi\{(X_{i_1}, Y_{i_1}), (X_{i_2}, Y_{i_2}), (X_{i_3}, Y_{i_3})\}, \tag{10}$$

where $\Phi : C^2[a, b] \times C^2[a, b] \times C^2[a, b] \to R$ is a Bochner integrable symmetric function according to[13], and given by

$$\Phi[(x_i, y_i), (x_j, y_j), (x_z, y_z)] = 6I(x_i \prec x_j, y_i \prec y_z) + 6I(x_j \prec x_i, y_z \prec y_i) - 3,$$

where $I$ denotes the indicator function.

## 2.2 | Properties of functional $\rho_s$

Let $(X_t, Y_t)$ be a bivariate stochastic process and let $\rho_s(X_t, Y_t)$ the corresponding Spearman index. Then:

1. $\rho_s$ is well defined for any $(X_t, Y_t)$.
2. $\rho_s(X_t, Y_t) = \rho_s(Y_t, X_t)$.
3. $-1 \le \rho_s(X_t, Y_t) \le 1$.
4. $\rho_s(X_t, g(X_t)) = 1$ for any increasing function g.
5. $\rho_s(X_t, g(X_t))$ = -1 for any decreasing function g.
6. The Spearman index is invariant under strictly increasing and continuous transformations of the processes, that is:

   $$\rho_s(\alpha(X_t), \beta(Y_t)) = \rho_s(X_t, Y_t)$$

   where $\alpha(\cdot)$ and $\beta(\cdot)$ are strictly increasing functions.
7. If $X_t$ and $Y_t$ are stochastically independent, then $\rho_s(X_t, Y_t) = 0$.
8. The sample Spearman index is a consistent estimator of the index of the original processes.

The proofs of properties 1,2 and 3 are trivial from the definition of $\rho_s$.
The details for the proofs of properties 4, 5 and 6 are reported in[6].
Property 7 is based on the fact that, if $X_t$ and $Y_t$ are independent, than the random variables $IL - grade(X_t)$ and $IL - grade(Y_t)$ are also independent. Therefore

$$\rho_s(X_t, Y_t) = \rho_p(IL - grade(X_t), IL - grade(Y_t)) = 0,$$

by the well known property of the Pearson correlation coefficient.
Expressing the functional $\hat{\rho}$ given in (9) as a UB-statistic, we can get the consistency of functional $\hat{\rho}_s$ applying Theorem 2 of[5] obtaining an asymptotic result in the functional field also for the Spearman's coefficient.
**Theorem** (*Asymptoticity of* $\hat{\rho}_n$) Let $(x_1, y_1), ..., (x_n, y_n)$ be a sample of independent and identical functional observations from (X,Y). Then,

$$\hat{\rho}_n \to \rho_s \text{ a.s. as } n \to \infty$$

It is easy to check that the function

$$\Phi[(x_i, y_i), (x_j, y_j), (x_z, y_z)] = 6I(x_i \prec x_j, y_i \prec y_z) + 6I(x_j \prec x_i, y_z \prec y_i) - 3,$$

which represents the kernel for the $\hat{\rho}_n$ belongs to the interval [-3, 11]. Then, the functional $\hat{\rho}$, given in (9) and expressed as a UB-statistic in (10), has associated a kernel $\Phi$ such that $E||\Phi||$ is finite.
Therefore, from Theorem 1 in[12], we have that, if $\Phi$ is such that $E||\Phi|| < \infty$, then the UB-statistic will converge almost surely to the parameter $\rho$. Observe that the above theorem is valid in general for any well-defined preorder $(\prec_m, \prec_i)$.

## 2.3 | Spearman Matrix for *h*-variate functional data

We introduce a new mathematical object for expressing the pattern of dependence among the components of a multivariate functional dataset. Given the h-variate stochastic process $X_t$ as above, we define the *Spearman Matrix* (*SM* in the following) as the $h \times h$ symmetric matrix

$$SM(\mathbf{X}_t) = \begin{bmatrix} \rho_s(X_t^1, X_t^1) & \rho_s(X_t^1, X_t^2) & \dots & \rho_s(X_t^1, X_t^h) \\ \rho_s(X_t^2, X_t^1) & \rho_s(X_t^2, X_t^2) & \dots & \rho_s(X_t^2, X_t^h) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_s(X_t^h, X_t^1) & \rho_s(X_t^h, X_t^2) & \dots & \rho_s(X_t^h, X_t^h) \end{bmatrix}, \tag{11}$$

where $\rho_s(X_t^i, X_t^j)$ is the Spearman index between the *i-th* and *j-th* component of the stochastic process, as in Equation 5. Let then

$$\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ ...\ \mathbf{x}_h] = \begin{bmatrix} x_{1,1}(t) & x_{1,2}(t) & \dots & x_{1,h}(t) \\ x_{2,1}(t) & x_{2,2}(t) & \dots & x_{2,h}(t) \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1}(t) & x_{n,2}(t) & \dots & x_{n,h}(t) \end{bmatrix}_{t \in I}, \tag{12}$$

be a multivariate functional dataset, composed by $n$ realizations of the stochastic process $X_t$, where the vectors

$$\mathbf{x}_i = (x_{1,i}(t), x_{2,i}(t), ..., x_{n,i}(t))'_{t \in I}, \quad i = 1, ..., h$$

represent the functional samples containing the realizations of a specific component of the process. To avoid hard notations, the vectors are represented neglecting the dependence on time. The sample Spearman Matrix $\widehat{SM}(\mathbf{X})$ is given by

$$\widehat{SM}(\mathbf{X}) = \begin{bmatrix} \hat{\rho}_s(\mathbf{x}_1, \mathbf{x}_1) & \hat{\rho}_s(\mathbf{x}_1, \mathbf{x}_2) & \dots & \hat{\rho}_s(\mathbf{x}_1, \mathbf{x}_h) \\ \hat{\rho}_s(\mathbf{x}_2, \mathbf{x}_1) & \hat{\rho}_s(\mathbf{x}_2, \mathbf{x}_2) & \dots & \hat{\rho}_s(\mathbf{x}_2, \mathbf{x}_h) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_s(\mathbf{x}_h, \mathbf{x}_1) & \hat{\rho}_s(\mathbf{x}_h, \mathbf{x}_2) & \dots & \hat{\rho}_s(\mathbf{x}_h, \mathbf{x}_h) \end{bmatrix}_{t \in I}, \tag{13}$$

where $\hat{\rho}_s(\mathbf{x}_i, \mathbf{x}_j)$ is the sample Spearman index computed on the bivariate functional dataset $[\mathbf{x}_i, \mathbf{x}_j]$, as in Definition 3. It can be immediately noticed that $\widehat{SM}(\mathbf{X})$ is an easy to handle and easy to interpret mathematical object and its cross diagonal elements give a quick and effective overview of the pattern of dependence among components of a multivariate functional dataset. The great advantage with respect to the variance-covariance operator is the fact that the dependence among components is described through scalar indexes that may be tested in a suitable inferential context. Since $SM(\mathbf{X}_t)$ and $\widehat{SM}(\mathbf{X})$ are symmetric, in the following we will show only their upper triangular part.

# 3 | THE CASE STUDY

In this section, we apply the techniques previously described to a real case study. The aim is to compare the Spearman Matrix arising from the 8-variate electrocardiographic signals (ECG hereafter) of a population of healthy people with the one arising

from signals of people affected by Left Bundle Brunch Block (LBBB hereafter), a kind of Acute Myocardial Infarction. We want to investigate if the pattern of dependence between the components of the multivariate signals presents remarkable differences in the two cases.

## 3.1 | The dataset

Our data consist in a multivariate functional dataset containing the ECG traces of a population of healthy people and one composed by individuals affected by an heart disease called Left Bundle Brunch Block (LBBB). Each statistical unit (patient) is characterized by the 8-variate functional datum of his/her electrocardiogram, which describes his/her heart dynamics on the eight leads I, II, V1, V2, V3, V4, V5 and V6. The data are from PROMETEO (PROgetto sull'area Milanese Elettrocardiogrammi Teletrasferiti dall'Extra Ospedaliero) database. PROMETEO project has been started in 2008 with the aim of spreading the intensive use of ECGs as prehospital diagnostic tool. Each file contained in PROMETEO database can be associated to three sub-files, called *Details*, *Rhythm* and *Median*. For the aims of the analysis, only the last one is necessary. The *Median* file depicts a reference beat (obtained throught an authomatic filtering procedure applied to the Rhythm file) lasting 1.2 seconds on a grid of 1200 points. It provides, among others, 8 curves (one for each ECG lead) for each patient, representing patient's "Median" beat for that lead. This representative heartbeat is a trace of a single cardiac cycle (heartbeat), i.e., of a P wave, a QRS complex, a T wave, and a U wave. Actually PROMETEO database contains 6734 curves; among these, 1633 are healthy (i.e., not affected by cardiovascular diseases detectable through the ECG), whereas 5101 are affected by different heart diseases. See[14,15,16,17] for further details on the dataset and its use for statistical applications.

In what follows we will focus just on one of the most common disease, that is easily detectable observing the ECG signal. It is a kind of Myocardial Infarction named Left Bundle Brunch Block (LBBB). In the PROMETEO dataset, 314 people are affected by this pathology. After suitable preprocessing and robustification (see[17] for more details) of the dataset, the final sample available for the analyses is composed by 1564 Physiological curves and 205 LBBB curves, discretized on a uniformly time grid T of 1024 points. Each patient is represented by his/her discretized multivariate signal, i.e., for $i = 1, ..., n$, $\Phi_i(t)$: $T \subset \mathbb{R} \rightarrow \mathbb{R}^8$. All the curves of the available sample are registered and denoised (see[14] for further details on wavelet denoising and landmarks registration adopted for preprocessing data).

To fix the notation, we assume that the ECG signals of physiological and pathological patients are realizations of two different multivariate stochastic processes, $X_t = (X_t^1, X_t^2, ..., X_t^8)$ and $Y_t = (Y_t^1, Y_t^2, ..., Y_t^8)$, respectively. For the two processes, we require the same continuity assumptions introduced in Subsection 2.3. For the analysis, we construct two different multivariate functional dataset from the available sample: the first is denoted with $\mathbf{X}$ and collects $n_x = 200$ randomly chosen ECG signals from the population of the physiological (healthy) patients. In other words, $\mathbf{X}$ is a dataset $200 \times 8$ discretized functions, where the *i-th* row contains the multivariate curve (ECG) associated to the *i-th* selected patient. The second functional dataset is denoted with $\mathbf{Y}$ and contains the multivariate curves of $n_y = 200$ randomly chosen patients affected by LBBB. Notice that, without loss of generality, we are considering in order to ease the computations two populations of data with the same number of realizations.

Figures 1 and 2 show the ECG signals selected in the datasets $\mathbf{X}$ and $\mathbf{Y}$, respectively.

## 3.2 | Comparison between the Spearman matrices of the two populations

As we said before, we aim to study the pattern of dependence among leads in the two populations of healthy people and patients affected by LBBB and pointing out possibly significant differences. This aim is supported by the following argument: it is likely, clinically speaking, that the presence of the disease might affect the way the leads depend on each other. In fact, the LBBB patients have a region of the heart that is damaged, and this modifies the heart dynamics. So, we believe that the relation of dependence among leads may change due to the presence of the disease. Tables 1 and 2 show the Spearman matrices for physiological ($\widehat{SM}(\mathbf{X})$) and pathological ($\widehat{SM}(\mathbf{Y})$) ECGs, respectively (see[8]). The entries coloured in yellow represent those for which there isn't statistical evidence of being different from zero, based on the statistical procedure explained in the following, and indicate that the corresponding pairs of leads can be assumed independent. Their detection is performed observing the confidence intervals, based on bootstrap iterations contained in the matrices $\widehat{SM}(\mathbf{X})_{0.95}$ and $\widehat{SM}(\mathbf{Y})_{0.95}$ (reported in Tables 3 and 4, respectively): if an interval contains zero, the hypothesis of independence between the corresponding pair of leads is not rejected and so the component of the Spearman Matrix is highlighted to indicate a non significant dependence. We decide to highlight the independent pairs of leads instead of the dependent ones in order to point out, in a easier way, the dissimilarities between the patterns. The two matrices provide an effective insight on the way in which the leads of the ECG signals depend on

each other and give us the possibility to compare the pattern of dependence in the two populations. Firstly, we notice a similarity: the upper diagonals of the matrices are composed, except for one case, by high and significantly different from zero entries. This means that in both cases the dynamics of the heart on a lead is strictly related to the dynamics on the following one.
However, we notice remarkable differences. For instance, the pattern of dependence of physiological signals is more connected, whereas the one of LBBBs is more sparse, due to the presence of several pairs of independent leads. Moreover, it seems that V2 is particularly affected by the disease. In fact, in healthy patients, it depends on all the other leads, but the same does not hold in the pathological patients, being V2 correlated with only 3 leads.
What we observe can be interpreted in terms of heart dynamics: in physiological patients, the heart dynamics is more regular and expresses coordinated behaviours in all the components of the ECGs, whereas it becomes more chaotic and characterized by disjointed behaviours when the pathology is present.
Another difference can be noticed comparing the two matrices: in the case of physiological signals, the entries that are significatively different from zero are positive, indicating that there is agreement between the grades of the leads. The same does not happen for the LBBB signals, where we notice that the entries associated to the pairs V1-I, V1-V5 and V1-V6 are negative. Hence, it seems that the disease is able to change the natural relation of dependence among some leads of the ECG.

## 3.3 | Testing the equality of two Spearman matrices

A test that compares the patterns of dependence of two populations of multivariate functional data can be formulated exploiting the notion of Spearman Matrix, as follows.

Suppose to have two stochastic processes

$$\mathbf{X}_t = (X_t^1, X_t^2, ..., X_t^h), \qquad \mathbf{Y}_t = (Y_t^1, Y_t^2, ..., Y_t^h),$$

with $h > 2$ and assume that the same continuity assumptions of Section 2.3 hold. Assume also to have two multivariate functional datasets sampled

$$\mathbf{X} = \begin{bmatrix} x_{1,1}(t) & x_{1,2}(t) & \dots & x_{1,h}(t) \\ x_{2,1}(t) & x_{2,2}(t) & \dots & x_{2,h}(t) \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_1,1}(t) & x_{n_1,2}(t) & \dots & x_{n_1,h}(t) \end{bmatrix}_{t\in T}, \qquad \mathbf{Y} = \begin{bmatrix} y_{1,1}(t) & y_{1,2}(t) & \dots & y_{1,h}(t) \\ y_{2,1}(t) & y_{2,2}(t) & \dots & y_{2,h}(t) \\ \vdots & \vdots & \ddots & \vdots \\ y_{n_2,1}(t) & y_{n_2,2}(t) & \dots & y_{n_2,h}(t) \end{bmatrix}_{t\in T},$$

from $\mathbf{X}_t$ and $\mathbf{Y}_t$, respectively. We want to perform the test:

$$H_0 : SM(\mathbf{X}_t) = SM(\mathbf{Y}_t) \quad vs \quad H_1 : SM(\mathbf{X}_t) \neq SM(\mathbf{Y}_t).$$

The two matrices that we want to test enjoy an important property. In fact, according to (5), we have

$$(SM(\mathbf{X}_t))_{i,j} = \rho_s(X_t^i, X_t^j) = \rho_p((IL - grade(X_t^i), IL - grade(X_t^j))),$$

for $i, j = 1, ..., h$, where $\rho_p$ indicates the Pearson correlation coefficient.
Hence, the equality of two Spearman matrices can be tested exploiting the multivariate realizations defined by the sample Inferior (or Superior) Length grades of the multivariate curves.
In literature, the problem of testing the equality of two correlation matrices has been studied extensively and several tests in this direction have been proposed. The typical approach is to assume a parametric model for the two populations of multivariate data and determine a test statistic that is able to capture the possible deviation from the null hypothesis (e.g. Jennrich's[18] and Larntz-Perlman's[19] statistics).
Since these methods require questionable assumptions on the grades of the curves, we move on to a non parametric approach presenting a permutation-based testing procedure. Our proposal has a great advantage with respect to the tests mentioned above since it does not require any distributional assumption on data that may restrict its application. In this section we will discuss the newly proposed test statistic and its associated algorithm for the two-sample problem (see[20]).

**Definition 6.** Let $R^{h\times h}$ be a vector space of symmetric matrices of size $h \times h$ and $f$: $R^{h\times h} \to R^{m\times n}$. The angle between two symmetric matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ in $R^{h\times h}$ with respect to $f$ is defined as $arccos(\mathbf{M}_1,\mathbf{M}_2)$ and

$$cos(\mathbf{M}_1, \mathbf{M}_2) = \frac{\langle f(\mathbf{M}_1), f(\mathbf{M}_2)\rangle}{\|f(\mathbf{M}_1)\| \|f(\mathbf{M}_2)\|}, \tag{14}$$

$$\widehat{SM}(\mathbf{X}) = \begin{array}{c} \\ I \\ II \\ V1 \\ V2 \\ V3 \\ V4 \\ V5 \\ V6 \end{array} \begin{array}{c} \begin{array}{cccccccc} I & II & V1 & V2 & V3 & V4 & V5 & V6 \end{array} \\ \begin{pmatrix} 1 & 0.382 & \text{-0.069} & 0.303 & 0.327 & 0.386 & 0.439 & 0.456 \\ & 1 & 0.036 & 0.202 & 0.500 & 0.596 & 0.605 & 0.611 \\ & & 1 & 0.674 & 0.372 & 0.146 & \text{-0.001} & \text{-0.046} \\ & & & 1 & 0.635 & 0.475 & 0.376 & 0.300 \\ & & & & 1 & 0.830 & 0.678 & 0.496 \\ & & & & & 1 & 0.869 & 0.662 \\ & & & & & & 1 & 0.811 \\ & & & & & & & 1 \end{pmatrix} \end{array}$$

**TABLE 1** Spearman Matrix for the population of the physiological signals. The non significant components (highlighted in yellow) are detected according to the confidence intervals of Table 3 .

$$\widehat{SM}(\mathbf{Y}) = \begin{array}{c} \\ I \\ II \\ V1 \\ V2 \\ V3 \\ V4 \\ V5 \\ V6 \end{array} \begin{array}{c} \begin{array}{cccccccc} I & II & V1 & V2 & V3 & V4 & V5 & V6 \end{array} \\ \begin{pmatrix} 1 & 0.459 & -0.392 & \text{-0.052} & \text{-0.016} & 0.346 & 0.607 & 0.653 \\ & 1 & \text{-0.095} & 0.036 & 0.198 & 0.471 & 0.599 & 0.582 \\ & & 1 & 0.750 & 0.560 & 0.123 & -0.220 & -0.370 \\ & & & 1 & 0.734 & 0.363 & 0.010 & \text{-0.150} \\ & & & & 1 & 0.688 & 0.246 & \text{-0.036} \\ & & & & & 1 & 0.727 & 0.451 \\ & & & & & & 1 & 0.843 \\ & & & & & & & 1 \end{pmatrix} \end{array}$$

**TABLE 2** Spearman Matrix for the population of the LBBB signals. The non significant components (highlighted in yellow) are detected according to the confidence intervals of Table 4 .

$$\begin{array}{c} \\ I \\ II \\ V1 \\ V2 \\ V3 \\ V4 \\ V5 \\ V6 \end{array} \begin{array}{c} \begin{array}{cccccccc} I & II & V1 & V2 & V3 & V4 & V5 & V6 \end{array} \\ \begin{pmatrix} 1 & (0.259, 0.506) & \text{(-0.207,0.073)} & (0.19, 0.416) & (0.183, 0.448) & (0.25, 0.499) & (0.316, 0.55) & (0.319, 0.564) \\ & 1 & \text{(-0.105,0.187)} & (0.063, 0.339) & (0.397, 0.59) & (0.493, 0.686) & (0.506, 0.691) & (0.514, 0.692) \\ & & 1 & (0.554, 0.742) & (0.255, 0.485) & (0.002, 0.28) & \text{(-0.138,0.144)} & \text{(-0.188,0.1)} \\ & & & 1 & (0.538, 0.713) & (0.36, 0.573) & (0.261, 0.483) & (0.165, 0.424) \\ & & & & 1 & (0.778, 0.867) & (0.588, 0.745) & (0.373, 0.591) \\ & & & & & 1 & (0.809, 0.902) & (0.551, 0.741) \\ & & & & & & 1 & (0.741, 0.869) \\ & & & & & & & 1 \end{pmatrix} \end{array}$$

**TABLE 3** Matrix of confidence intervals of coverage probability 0.95 for the components of the Spearman Matrix associated to the population of the physiological signals where each interval is computed using $B = 1000$ bootstrap iterations. The intervals containing zero are highlighted in yellow.

$$\begin{array}{c} \\ I \\ II \\ V1 \\ V2 \\ V3 \\ V4 \\ V5 \\ V6 \end{array} \begin{array}{c} \begin{array}{cccccccc} I & II & V1 & V2 & V3 & V4 & V5 & V6 \end{array} \\ \begin{pmatrix} 1 & (0.308, 0.573) & (-0.508, -0.261) & \text{(-0.205,0.082)} & \text{(-0.147,0.123)} & (0.23, 0.46) & (0.501, 0.686) & (0.547, 0.732) \\ & 1 & \text{(-0.239,0.064)} & \text{(-0.125,0.183)} & (0.049, 0.333) & (0.352, 0.569) & (0.499, 0.69) & (0.459, 0.676) \\ & & 1 & (0.661, 0.807) & (0.461, 0.653) & \text{(-0.021,0.257)} & (-0.355, -0.074) & (-0.503, -0.225) \\ & & & 1 & (0.654, 0.796) & (0.234, 0.49) & \text{(-0.162,0.16)} & \text{(-0.302,0.012)} \\ & & & & 1 & (0.58, 0.77) & (0.086, 0.375) & \text{(-0.179,0.111)} \\ & & & & & 1 & (0.638, 0.795) & (0.333, 0.561) \\ & & & & & & 1 & (0.777, 0.884) \\ & & & & & & & 1 \end{pmatrix} \end{array}$$

**TABLE 4** Matrix of confidence intervals of coverage probability 0.95 for the components of the Spearman Matrix associated to the population of the pathological signals where each interval is computed using $B = 1000$ bootstrap iterations. The intervals containing zero are highlighted in yellow.

where $\langle \cdot, \cdot \rangle$ is an inner product and $\| \cdot \|$ is the corresponding norm in $R^{m \times n}$.

Since a $h \times h$ symmetric matrix is completely determined by its lower triangular elements together with the symmetry, we construct the mapping $f$ by applying the half-vectorization operator on $\mathbf{M}_1$ and $\mathbf{M}_2$ directly. The cosine between $\mathbf{M}_1$ and $\mathbf{M}_2$ can be obtained from Equation (14) as follows:

$$cos(\mathbf{M}_1, \mathbf{M}_2) = \frac{vech(\mathbf{M}_1)^T vech(\mathbf{M}_2)}{\|vech(\mathbf{M}_1)\| \|vech(\mathbf{M}_2)\|}. \tag{15}$$

For correlation matrices, the computation can be simplified by introducing a modified half-vectorization operator $vech^*(\cdot)$ from $vech(\cdot)$ by excluding the diagonal elements of the matrix. Suppose $\mathbf{M}_1$ and $\mathbf{M}_2$ are two correlation matrices, the cosine can be computed by Equation (14) as

$$cos(\mathbf{M}_1, \mathbf{M}_2) = \frac{vech^*(\mathbf{M}_1)^T vech^*(\mathbf{M}_2)}{\|vech^*(\mathbf{M}_1)\| \|vech^*(\mathbf{M}_2)\|}. \tag{16}$$

As mentioned in[20] the half-vectorization operator, $vech(\cdot)$, for covariance matrices and the modified half-vectorization operator, $vech^*(\cdot)$, for correlation matrices completely remove the redundancy in symmetric matrices and are very easy to compute. Moreover, these two operators show better statistical power in pilot simulation studies. The cosine value computed from Equation (14) measures the similarity between two symmetric matrices. When this value is one, the two matrices are identical. As proposed in[20] we consider the following test: let $\mathbf{X}_{n_1 \times p}$ and $\mathbf{Y}_{n_2 \times p}$ be two moultivariate functional dataset with sample Spearman correlation matrices $\mathbf{S}_1$ and $\mathbf{S}_2$, respectively. We consider the following test statistic

$$1 - \frac{vech^*(\mathbf{S}_1)^T vech^*(\mathbf{S}_2)}{\|vech^*(\mathbf{S}_1)\| \|vech^*(\mathbf{S}_2)\|}, \tag{17}$$

We propose Algorithm 1 to compute p-values from the distribution of this test statistic under the null hypothesis of equality. In this algorithm, two data matrices $\mathbf{X}_{n_1 \times h}$ and $\mathbf{Y}_{n_2 \times h}$ are stacked to form a new data matrix $\mathbf{D}_{n \times h}$ ($n = n_1 + n_2$). In each permutation the rows of $\mathbf{D}_{n \times h}$ are randomly permuted to generate a permuted data matrix $\mathbf{D}^*_{n \times h}$, which is then split into two data matrices $\mathbf{X}^*_{n_1 \times h}$ and $\mathbf{Y}^*_{n_2 \times h}$ to compute the test statistic (17). Algorithm 1 assumes that $X_{1i}$ in $\mathbf{X}$ and $Y_{1i}$ in $\mathbf{Y}$ have the same distribution for all $i$'s under the null hypothesis. One advantage of our proposed two-sample test is that $X_{1i}$ and $Y_{1j}$, for $i \neq j$, need not to have the same distribution. The rationale behind Algorithm 1 is that the cosine value between $\mathbf{S}^*_1$ and $\mathbf{S}^*_2$ is similar to that of $\mathbf{S}_1$ and $\mathbf{S}_2$ under the null hypothesis and the permutations provide a good control of the type I error. Under the alternative, the repeated random-mixing rows of $\mathbf{X}_{n_1 \times h}$ and $\mathbf{Y}_{n_2 \times h}$ produce $\mathbf{S}^*_1$ and $\mathbf{S}^*_2$ such that the cosine value between the two is bigger than that of $\mathbf{S}_1$ and $\mathbf{S}_2$, therefore the test statistic (17) has good power to reject the null at a pre-determined significance level.

**Algorithm 1:** Test for the equality of two Spearman matrices.

$B \leftarrow$ the number of permutations
$T(i) \leftarrow 0$, $i$=1,...,$B$
$\mathbf{S}_1 \leftarrow$ Compute the sample Spearman matrix from $\mathbf{X}_{n_1 \times h}$
$\mathbf{S}_2 \leftarrow$ Compute the sample Spearman matrix from $\mathbf{Y}_{n_2 \times h}$
$T_0 \leftarrow$ Compute (17)
$\mathbf{D}_{(n_1+n_2) \times h} \leftarrow$ stack $\mathbf{X}_{n_1 \times h}$ and $\mathbf{Y}_{n_2 \times h}$
For $i$=1 to $i$=$B$
  $\mathbf{D}^*_{(n_1+n_2) \times h} \leftarrow$ randomly shuffle the rows of $\mathbf{D}_{(n_1+n_2) \times h}$
  $\mathbf{X}^*_{n_1 \times h} \leftarrow$ the first $n_1$ rows of $\mathbf{D}^*_{(n_1+n_2) \times h}$
  $\mathbf{Y}^*_{n_2 \times h} \leftarrow$ the remaining $n_2$ rows of $\mathbf{D}^*_{(n_1+n_2) \times h}$
  $\mathbf{S}^*_1 \leftarrow$ Compute the sample Spearman matrix from $\mathbf{X}^*_{n_1 \times h}$
  $\mathbf{S}^*_2 \leftarrow$ Compute the sample Spearman matrix from $\mathbf{Y}^*_{n_2 \times h}$
  $T(i) \leftarrow$ Compute (17)
End For
Report $p$-value = (#$(T(i) \geq T_0)$+1)/($B$+1)

## 3.4 | Application to the case study

Now we are ready to apply the two-sample Anderson-Darling test to the case of the Spearman matrices of physiological and LBBB patients. In Figure 3 we report the histogram of $B = 1000$ permutational replications of T under $H_0$ and a dashed line denoting the observed value $T_0(\mathbf{X}, \mathbf{Y})$. As you can see, the line is drawn on the right hand side of the replications, indicating that the observed value is not likely under the null hypothesis (the p-value of the test is approximately 0.000999001). Therefore, the test gives strong evidence to reject $H_0$ and to state that the Spearman matrices of physiological and pathological signals are different, coherently with the results presented in Section 3.2.
If we consider other distances between matrices such as the one induced by the *one, infinity* and *Frobenius norm*, the result does not change confirming strong evidence in favour of the dissimilarity of the two Spearman matrices.

# 4 | CONCLUSIONS

In this work, we consider the notion of Spearman index in the infinite dimensional framework to quantify the dependence among two families of functional data. We studied also its properties to prove that it is a consistent estimator of the index of the original processes. Starting from this definition, we build the Spearman Matrix, a new mathematical object that mimics the correlation matrix of multivariate statistics and that provides an effective insight of the pattern of dependence among the components of a multivariate functional dataset. This is a new tool in the literature of functional data analysis that can be adopted for different applications. In both methodological and applicative parts of this work, we present interesting results.

The principal result of the methodological part is the definition of robust and innovative tools to investigate dependence in the functional setting. We consider the Spearman index in the multivariate framework through the notion of grade for a stochastic process. Subsequently we focus on the bivariate case defining the Spearman index for bivariate functional data based on the Pearson correlation coefficient between grades associated with the two stochastic processes. We demonstrated the consistency of functional Spearman estimator through the notion of UB-statistic, obtaining an asymptotic result in the functional field for the Spearman's coefficient. Finally, we define the Spearman Matrix for $h$-variate functional data, an easy to handle and easy to interpret mathematical object with its cross diagonal elements that give us a quick and effective overview of the pattern of dependence among components of a multivariate functional dataset. This matrix will be of great importance in the case study.

In the applicative part of our work, we moved to the analysis of a real dataset. We compared the Spearman Matrix arising from the 8-variate electrocardiographic signals of a population of healthy people with the one arising from signals of people affected by Left Bundle Brunch Block (LBBB). The aim was to verify if the pattern of dependence in the two cases are different due to the presence of the disease. From a visual comparison between the sample Spearman matrices of the two populations, we observed that the ECG signals of the physiological patients are characterized by a coordinated pattern in which most of the components depend on the others. The some does not happen in the case of LBBB curves, where we noticed several pairs of independent leads. Moreover, in pathological curves some components of the Spearman Matrix change sign with respect to the physiological case. These basic observations induced us to believe that the pathology changes the relations between the leads of the electrocardiographic signal. A statistical confirmation of this conjecture was provided using quantitative tools. We tested the hypothesis that physiological and pathological signals have the same Spearman Matrix, adapting to our framework a non-parametric test which checks the equality of Spearman correlation matrices arising from different populations of multivariate data. This procedure is an alternative way to perform the test which avoids strong assumptions on the grade of the multivariate curves. The test proposed is a permutational test procedure, based on the notion of the generalized cosine measure between two symmetric matrices, and it gave us strong evidence to reject the null hypothesis and to consider the Spearman matrices of the two populations different. Therefore, the synergy between the inferential framework built for the Spearman index and the test for comparing the Spearman matrices of different multivariate functional datasets enabled us to show, in a rigorous way, that LBBB affects the heart dynamics changing the way in which the leads of the electrocardiographic signal depend on each other.

# References

1. Ramsay J, Silverman B. *Functional Data Analysis*. 0172-7397Springer-Verlag New York . 2005.

2. Kokoszka P, Reimherr M. *Introduction to Functional Data Analysis*. Chapman and Hall/CRC . 2017.

3. Ferraty F, View P. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in StatisticsSpringer-Verlag New York . 2006.

4. Li R, Chow M. Evaluation of reproducibility for paired functional data. *Journal of Multivariate Analysis* 2005(923): 81-101.

5. Valencia D, Lillo R, Romo J. A Kendall correlation coefficient between functional data. *Advances in Data Analysis and Classification* 2019(13): 1083-1103.

6. Valencia D, Lillo R, Romo J. *Dependence for functions: Spearman coefficient*. Technical Report, Department of Statistics, UC3M - University Carlos III of Madrid, Spain . 2016.

7. R Core Team . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; Vienna, Austria: 2021.

8. Ieva F, Paganoni AM, Romo J, Tarabelloni N. roahd Package: Robust Analysis of High Dimensional Data. *The R Journal* 2019; 11(2): 291–307.

9. Spearman C. The proof and measurement of association between two things. *American Journal of Psychology* 1904(15): 72–101.

10. Martin-Barragan B, Lillo R, Romo J. *Functional Boxplots based on Half-regions*. Submitted . 2012.

11. Lopez-Pintado S, Romo J. *A half-region depth for functional data*. Computational Statistics and Data Analysis 55, pp 1679-1695 . 2011.

12. Borovskikh . *U-statistics in Banach space*. VSP BV, Oud-Beijerland . 1996.

13. Schwabik S, Guoju Y. *Topics in Banach space integration*. World Scientific Publishing, Singapore . 2005.

14. Ieva F, Paganoni A, Pigoli D, Vitelli V. Multivariate functional clustering for the analysis of ECG curves morphology. *Journal of the Royal Statistical Society – Series C* 2013; 62: 401–418.

15. Tarabelloni N, Ieva F, Paganoni A, Biasi R. Use of depth measure for multivariate functional data in disease prediction: an application to electrocardiograph signals. *International Journal of Biostatistics* 2015; 11: 189—201.

16. Ieva F, Paganoni A. Risk Prediction for Myocardial Infarction via Generalized Functional Regression Models. *Statistical Methods in Medical Research* 2016; 25: 1648–1660.

17. Ieva F, Paganoni A. Component-wise outlier detection methods for robustifying multivariate functional samples. *Statistical Papers* 2020(61).

18. Jennrich R. *An asymptotic $\chi^2$ test for the equality of two correlation matrices*. Journal of the American Statistical Association. 65, pp 904-912 . 1965.

19. Larntz K, Perlman M. *A simple test for the equality of correlation matrices*. Technical Report n. 63, Department of Statistics, University of Washington . 1985.

20. Wu L, Weng C, Wang X, Wang K, Liu X. Test of Covariance and Correlation Matrices. 2018.

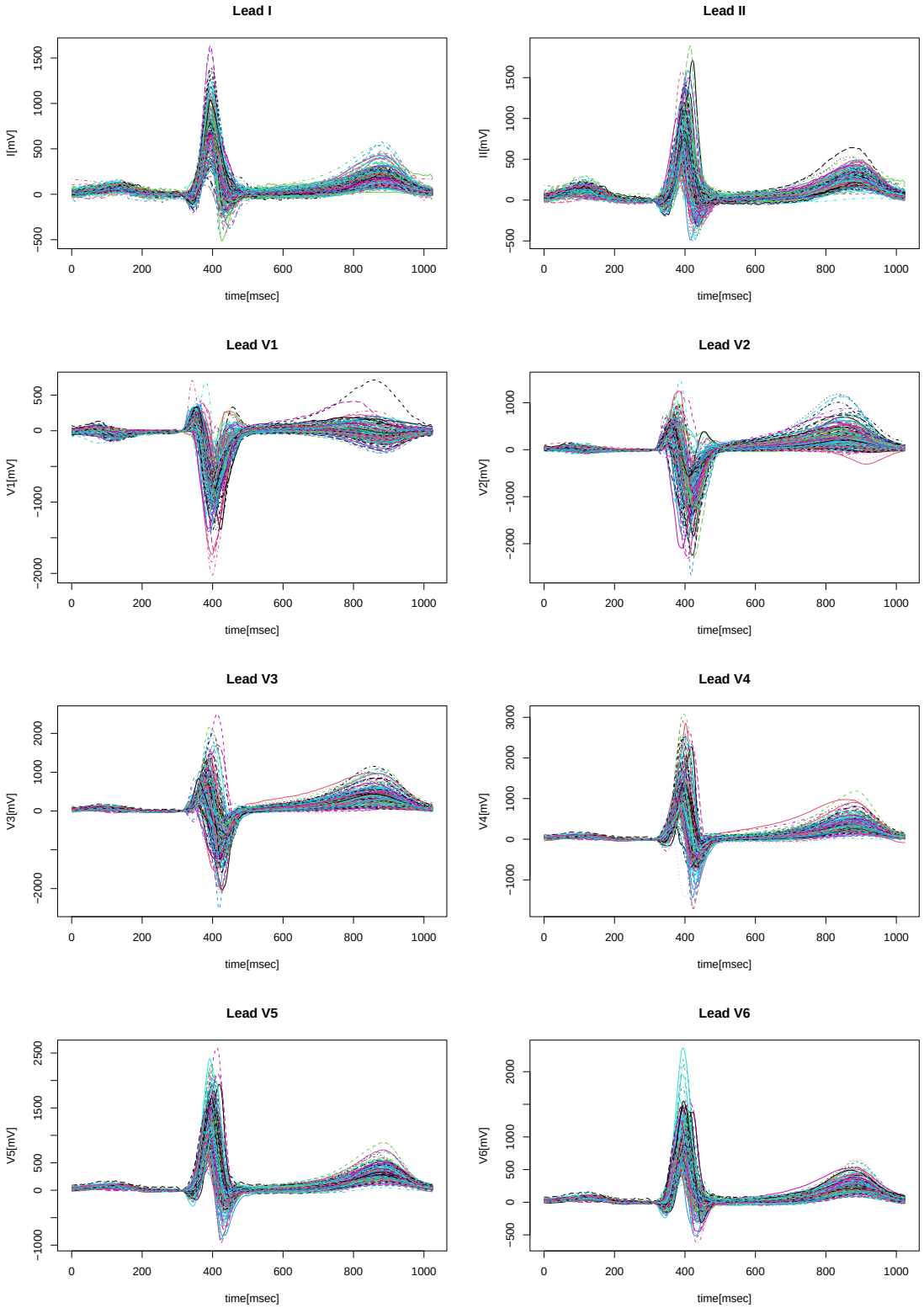


**FIGURE 1** Registered and denoised ECG signals of the $n_x = 200$ physiological patients used for the analysis.

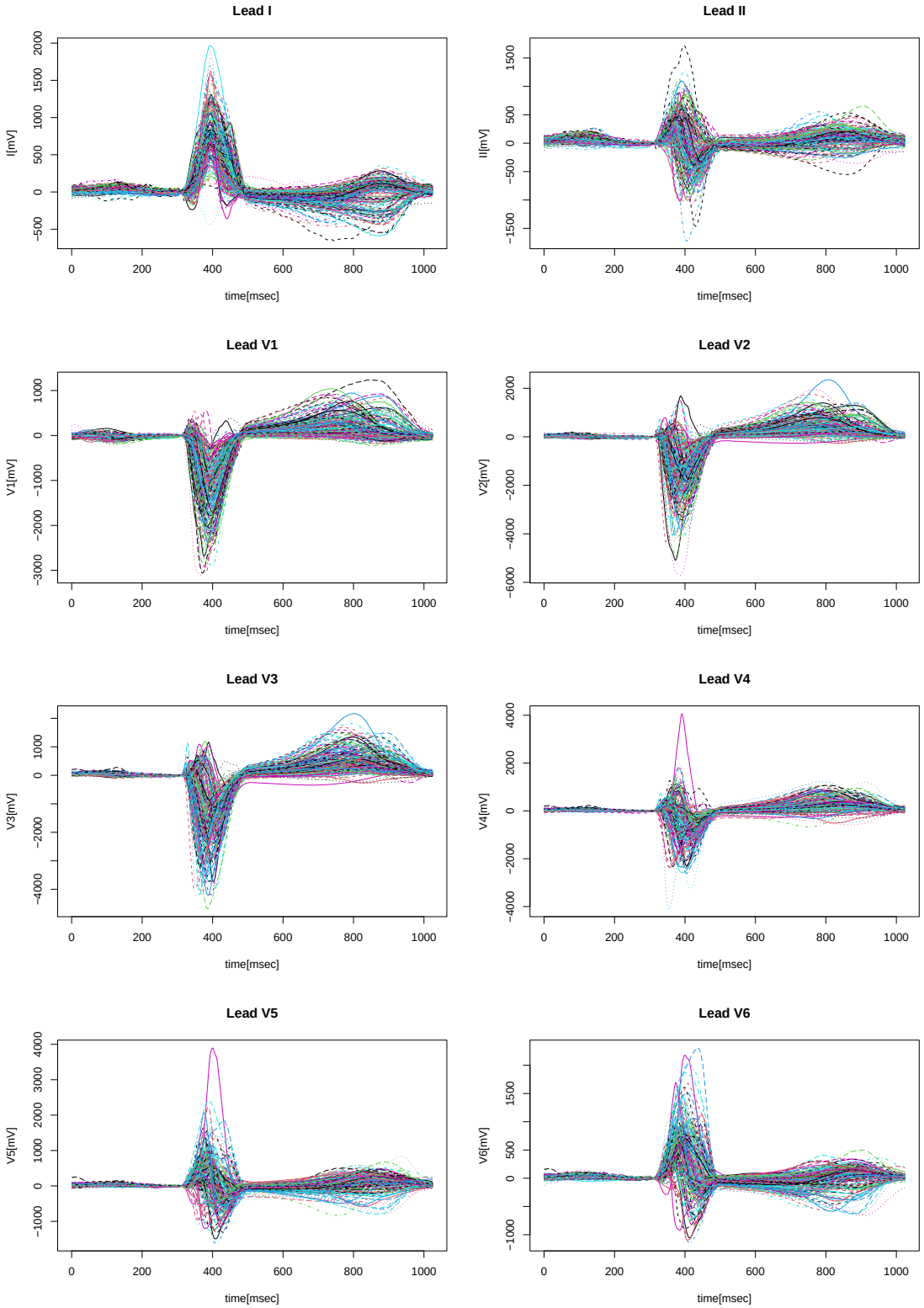


**FIGURE 2** Registered and denoised ECG signals of the $n_y = 200$ LBBB patients used for the analysis.

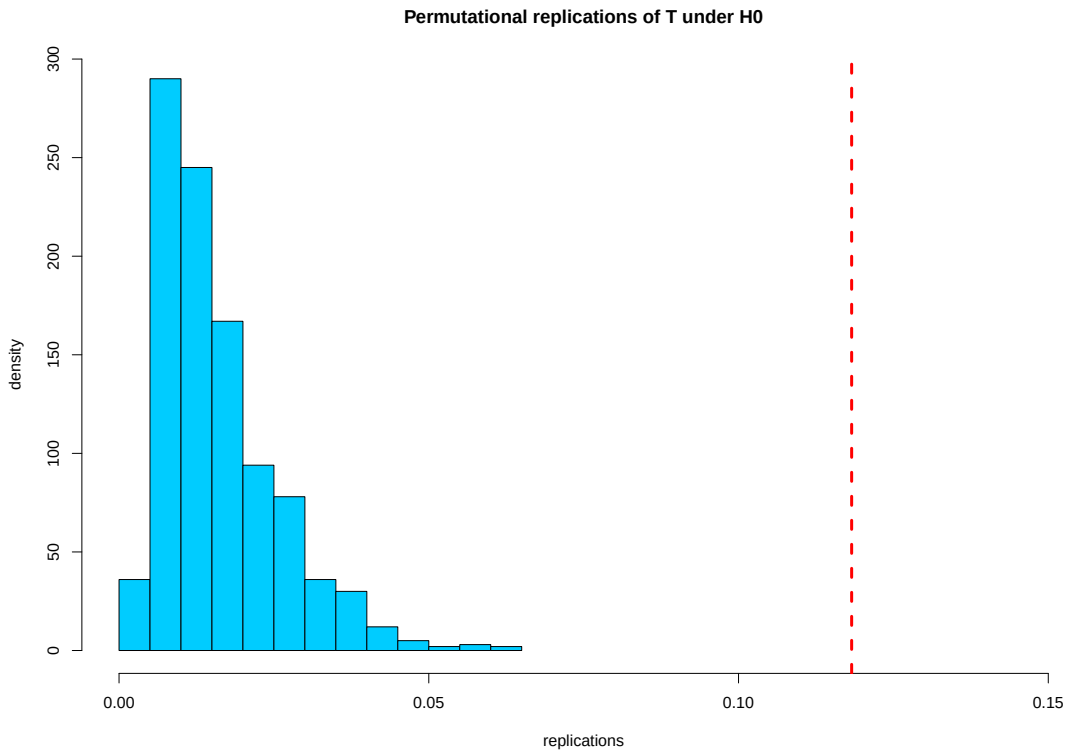


**FIGURE 3** Histogram of $B = 1000$ permutational replications of T under $H_0$ computed using the notion of the generalized cosine measure between two symmetric matrices. The dashed line is drawn at the observed value $T_0(\mathbf{X}, \mathbf{Y}) = 0.12$.