XLI CILAMCE
XLI Ibero-Latin American Congress on Computational Methods in Engineering
16-19 November, 2020 FOZ DO IGUAÇU PARANÁ - BRAZIL

Using Multivariate Energy Distance on Frequency-based Damage Detection

Horacio V. Duarte[1], Lázaro V. Donadon[1]

[1] *Dept. de Engenharia Mecânica, Universidade Federal de Minas Gerais*
Av. Antônio Carlos 6627, CEP , Minas Gerais, Brasil
hvduarte@ufmg.br, lazaro@ufmg.br

Abstract. The Frequency-based Damage Detection Method using FRFs (Frequency Response Functions) in a high-frequency range can identify a structural failure of about 0.16% and 0.34% of the total mass. These results demonstrate the effectiveness and reliability of the method to track the problem in its initial stage. The Robust Singular Value Decomposition algorithm (RSVD) is employed in Frequency-based Damage Detection Method to find out a damaged subset from a set composed by reference and damaged data. Although effective, this procedure has some drawbacks, it is time consuming and the method performance is linked to optimum singular value basis reduction. The main objective of this work is to select an algorithm to overcome these less acceptable features. In order to achieve this goal there are a short review of the main methods employed in data classification, parameter extraction. The selection criteria is to pick out those that present simple algorithms and that are based on metric distances. From those review the Multivariate Energy Distance Correlation was selected. Using experimental data this Energy statistic method are compared with Robust Singular Value Decomposition (RSVD).

Keywords: Frequency-based Damage Detection Method, Multivariate Distance Correlation, Energy Distance, Bootstrap

1 Introduction

The main objective of this work is to find out a procedure to measure the distance between two FRF data sets or two matrices. The idea is using a metric distance that ensures a unique value to each one in sense that one could compare them. Being effective the procedures must satisfy all metric axioms.

A recent proposed procedure is the Energy Covariance and Correlation Distance introduced by Székely, Rizzo and Bakirov [4]. Energy Distance is a distance between probability distributions (Rizzo[5]). It applies to random vectors in arbitrary dimensions, and the methodology requires only the mild assumption of finite first moments (Rizzo [5]). This method has been made a significant impact and there are many routines for different areas as reported in [5] mainly, but not only, in times series analysis where identify temporal dependence structure on time series is an important issue (Zhou [6], Pitsillou [7]).

For this work, beyond simplicity, the main aspect of the Energy Correlation Distance is the property of the coefficients $\mathcal{R}_n^2(X,Y)$. For variables X and Y with finite first moments (finite mean), all possibilities for $\mathcal{R}_n^2(X,Y)$ lies on the range $0 \leq \mathcal{R}_n^2(X,Y) \leq 1$. $\mathcal{R}_n^2(X,Y) = 0$ only if X and Y are linearly independent and $\mathcal{R}_n^2(X,Y) = 1$ only if there is a linear dependence between X and Y. This is important because $\mathcal{R}_n^2(X,Y) = 0$ does not hold for linear independent sets for general metric spaces , only for separable ones (Lyons [2]). The Energy Distance $\mathcal{R}_n^2(X,Y)$ satisfies all axioms of a metric distance, (Székely [4]), and are valid not only for Euclidean metric although it is used in following revision. The authors also highlights that Distance correlation can be applied as an index of dependence, (Székely [4]), more important, without requiring distributional assumptions (Székely [8]).

2 The Energy Distances

As stated the algorithm Coefficient for Covariance Correlation Distances are very simple. From classical Euclidean distance matrix a_{ij} defined as $a_{ij} = |x_i - x_j| = \left[\sum_{k=1}^{p}(x_{ik} - x_{jk})^2\right]^{\frac{1}{2}}$, where $x_k \in \mathbb{R}^p$, $x_k \subset X^{n,p}$. Similarly $y_l \subset Y^{n,q}$ defines a Euclidean distance matrix $b_{ij} = |y_i - y_j|$. The bi-centered matrix A_{ij} is computed from Euclidean matrix, $A_{ij} = a_{ij} - \bar{a}_i - \bar{a}_j - \bar{a}$, where

$$\bar{a}_i = \bar{a}_{i,} = \frac{1}{n}\sum_{k=1}^{n} a_{i,k} \qquad \bar{a}_j = \bar{a}_{,j} = \frac{1}{n}\sum_{k=1}^{n} a_{k,j} \qquad \bar{a} = \frac{1}{n^2}\sum_{i,j=1}^{n} a_{i,j} \tag{1}$$

B_{ij} is computed using similar procedure $B_{ij} = b_{ij} - \bar{b}_i - \bar{b}_j - \bar{b}$. The Distance Covariance, $\mathcal{V}_n^2(X,Y)$, and the Distance Correlation coefficient $\mathcal{R}_n^2(X,Y)$ between $X_{n,p}$ and $Y_{n,q}$ is defined as[4]:

$$\mathcal{V}_n^2(X,Y) = \frac{1}{n^2}\sum_{i,j=1}^{n} A_{ij}B_{ij} \qquad \text{and} \qquad \mathcal{R}_n^2(X,Y)\begin{cases} = \dfrac{\mathcal{V}_n^2(X,Y)}{\sqrt{\mathcal{V}_n^2(X,X)\mathcal{V}_n^2(Y,Y)}} & \text{if} \quad \mathcal{V}_n(X,X)\times\mathcal{V}_n(Y,Y) \neq 0 \\ = \quad 0 & \text{if} \quad \mathcal{V}_n(X,X)\times\mathcal{V}_n(Y,Y) = 0 \end{cases} \tag{2}$$

$0 \leq \mathcal{V}_n^2(X,Y) < \infty$ and $0 \leq \mathcal{R}_n^2(X,Y) \leq 1$ both are equal 0 only if X and Y are independent. Also the empirical statistical coefficients $\mathcal{V}_n^2(X,Y)$ and $\mathcal{R}_n^2(X,Y)$ converge almost surely to the population coefficients: $n \to \infty$, $\lim_{n\to\infty}\mathcal{V}_n^2(X,Y) \to \mathcal{V}^2(X,Y)$ and $\lim_{n\to\infty}\mathcal{R}_n^2(X,Y) \to \mathcal{R}^2(X,Y)$.

The Székely paper [9] work starts with the observation that the bias of the $\mathcal{R}_n^2(X,Y)$ statistic increases with dimension. For $X_{n,p}$ and $Y_{n,q}$ when $p,q \to \infty$ then $\mathcal{R}_n(X,Y) \to 1$ and $\mathcal{V}_n(X,Y) \to \infty$. Székely [9] propose the bias corrected $A_{i,j}^*$ and $B_{i,j}^*$ matrices using the previous defined matrices $a_{i,j}$, $b_{i,j}$, $A_{i,j}$ and $B_{i,j}$

$$A_{i,j}^*\begin{cases} = \dfrac{n}{n-1}\left(A_{i,j} - \dfrac{a_{i,j}}{n}\right), & i \neq j \\ = \dfrac{n}{n-1}\left(\bar{a}_i - \bar{a}\right), & i = j \end{cases} \quad \text{and} \quad B_{i,j}^*\begin{cases} = \dfrac{n}{n-1}\left(B_{i,j} - \dfrac{b_{i,j}}{n}\right), & i \neq j \\ = \dfrac{n}{n-1}\left(\bar{b}_i - \bar{b}\right), & i = j \end{cases} \tag{3}$$

The Unbiased Distance Covariance statistics $\mathcal{V}_n^*(X,Y)$ and Unbiased Distance Correlation coefficient $\mathcal{R}_n^*(X,Y)$ are

$$\mathcal{V}_n^*(X,Y) = \frac{1}{n(n-3)}\left[\sum_{i,j=1}^{n} A_{i,j}^*B_{i,j}^* - \frac{n}{n-2}\sum_{i=1}^{n} A_{i,i}^*B_{i,i}^*\right] \qquad \text{and} \qquad \mathcal{R}_n^*(X,Y) = \frac{\mathcal{V}_n^*(X,Y)}{\sqrt{\mathcal{V}_n^*(X,X)\mathcal{V}_n^*(Y,Y)}} \tag{4}$$

if the Distance Covariances $\mathcal{V}_n^*(X,X) = 0$ or $\mathcal{V}_n^*(Y,Y) = 0$ then $\mathcal{R}_n^*(X,Y) = 0$. Also the empirical statistical coefficients $\mathcal{V}_n^*(X,Y)$ and $\mathcal{R}_n^*(X,Y)$ converge almost surely to the population coefficients $\mathcal{V}^*(X,Y)$ and $\mathcal{R}^*(X,Y)$ as $n \to \infty$.

2.1 Bootstrapping

The Bootstrap main objective is to perform a empirical test to estimate some statistic inferences and not to verify the underlying distribution from sample. Usually this procedure is done by permutation or resampling tests. Since Efron initial work [10], there has been developing a solid theoretical Bootstrap basis (Chernick and LaBudde [11]). The method is also very simple to program and there are many implementations (Davison and Kuonen [12]) in many software packages as in R software (Canty and Ripley [13]).

Bootstrap will be used to stablish the $\mathcal{R}^*$ and $\mathcal{R}^2$ Confidence Limits, CL. The Basic Confidence Limits is computed by $(\mathcal{R}_0^* - (T^\star_{(R+1)(1-\alpha)} - \mathcal{R}_0^*)\ ,\ \mathcal{R}_0^* - (T^\star_{(R+1)(\alpha)} - \mathcal{R}_0^*))$. $\mathcal{R}_0^*$ is the Distance Correlation from original set. The $T^\star_{(R+1)(1-\alpha)}, T^\star_{(R+1)(\alpha)}$ is the percentile interval for the basic Bootstrap confidence limits $(1-2\alpha)$, usually 95%. The percentile are taken from sorted Bootstrap statistic $T^\star_{(1)}, \ldots, T^\star_{(R)}$, R is the Bootstrap iterations number. $T^\star$ is the Distance Correlation Bootstrap computed, $(\mathcal{R}_0^* - T^\star)$ is named bias. The Percentile Confidence Limits uses the percentile limits only and is computed using $(\ T^\star_{(R+1)(1-\alpha)}\ ,\ T^\star_{(R+1)(\alpha)})$.

The Normal Confidence Limit should be used only if the statistic $T^\star$ follows a Gaussian distribution: $(\mathcal{R}_0^* - (T^\star - v(T^\star)z_{(1-\alpha)} - \mathcal{R}_0^*)\ ,\ \mathcal{R}_0^* - (T^\star - v(T^\star)z_{(\alpha)} - \mathcal{R}_0^*))$. The therm $v(T^\star)$ is the variance of Bootstrap statistic $T_r^\star$, and $z_{(\alpha)}$ is the percentile of normal distribution.

3 Results and Analysis

The first results presented are from an artificial dataset created to test the routine in a different situation that it is intended to use. Independent sets, linearly dependent sets and partial dependent sets are used from data designed to produce these results. Following, the experimental data from a metallic laminate beam are presented. These data was previously published Duarte [14] and this set was used because there are a poor RSVD failure identification for the presented frequency range.

3.1 Comparing Designed Sets for Algorithm Testing

Comparing Orthogonal Sets. In this section there are results comparing two datasets one from a matrix $X_{n,p}$ with $n = 270$ rows and $p = 10$ columns where each column is a random vector generated by a mean 1 exponential distribution. The second matrix $Y_{n,p}$ has equal dimensions and each column was a vector generated by random Gaussian distribution mean 1 and standard error sd=2, $\mathcal{N}(1, 2)$. Generating independent sets is only necessary to use dissimilar distributions sets, but different distributions types underlining these independent sets are also employed.

As the Energy Distance is intended at first to compare data distribution, for this two sets or two matrices, $X_{n,p}$ and $Y_{n,p}$, the Energy Distance should be zero as there are not only different distributions but also different distributions types. In Fig. (1) the Bootstrap results for the statistic $\mathcal{R}^*_n(X, Y)$, in the left-rand side the histogram and at right side the quantile plot comparing the Bootstrap quantile distribution versus quantile for normal distribution. This right side plot $Q_{\mathcal{R}^2_n} \times Q_{N(0,1)}$ shows that the Bootstrap distribution for coefficient deviates from normal at the beginning and is remarkable far from normal at the end quartiles.

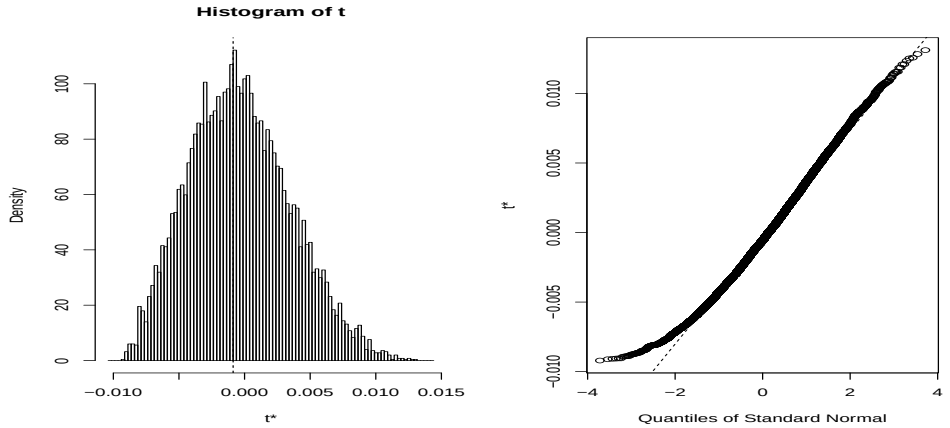

Figure 1. Energy Method, Bootstrap density histogram for $T^\star = \mathcal{R}^*_n$, and quantile plot $Q_{\mathcal{R}^*_n} \times Q_{N(0,1)}$ for $\mathcal{R}^*_n$ distribution versus normal distribution.

On Tab. 1 the Bootstrap statistic inference for $\mathcal{R}^2_n$ and $\mathcal{R}^*_n$ coefficients. For each one there are the result for the original data, T_o, where T_o represents $\mathcal{R}^2_o$ or $\mathcal{R}^*_o$, the Bootstrap bias related to initial result are $(T_o - T^\star)$, the standard error, the Bootstrap Confidence Limit computed using different procedures.

Table 1. Bootstrap statistic computed from orthogonal dataset, T_0, bias, standard error, sd, two-sided 95% Confidence Limits, CL, for coefficients $\mathcal{R}^2_n(X, Y)$ and $\mathcal{R}^*_n(X, Y)$. $n = 270$ rows and $p = 10$.

Coef.	T_0	bias $(T_0 - T^\star)$	sd	normal Conf. L.	basic Conf. L.	percentile Conf. L.
$\mathcal{R}^2_n$	0.1907321621	-0.0239433488	0.014130006	(0.1870, 0.2424)	(0.1904, 0.2437)	(0.1377, 0.1910)
$\mathcal{R}^*_n$	-0.0008710860	0.0005302428	0.003839384	(-0.0089, 0.0061)	(-0.0094, 0.0053)	(-0.0071, 0.0077)

As stated the expected result for the correlation coefficients are zero. The 95% two-sided Confidence Limit indicates a better performance for Unbiased $\mathcal{R}^*_n$ compared to original $\mathcal{R}^2_n$ coefficient. For Unbiased $\mathcal{R}^*_n$ the confidence interval is very small and includes zero indicating independence between the X and Y sets. The Confidence Limits for $\mathcal{R}^2_n$ are very large and do not include zero.

The poor performance for the $\mathcal{R}^2_n$ coefficient was not expected. The theoretical interval for $\mathcal{R}^2_n$ is

$0 \leq \mathcal{R}_n^2 \leq 1$ so the Confidence Limit interval could include zero, at least numerically in its inferior limit or nearest it, and this expected behaviour did not happen. As the vector dimension n is not small and p is not large it is a not expected result.

Comparing Linearly Dependent Sets. A vector $U_s = \mathcal{N}(1, 2)$ has a random Gaussian distribution, mean 1 and standard error sd=2. The reference set $X = [X_1, \ldots, X_p]$ is generated by adding to this U_s vector a random normal vector with a small variance ($\mathcal{N}(0.0, 0.1)$) acting as a noise, $X_i = U_s + \mathcal{N}_i(0, 0.1)$. The set Y to be tested follows the similar procedure, $Y_i = U_s + \mathcal{N}_i(0, 0.1)$, $Y = [Y_1, \ldots, Y_p]$. The expected Energy Distances should be close to 1 or linearly dependent.

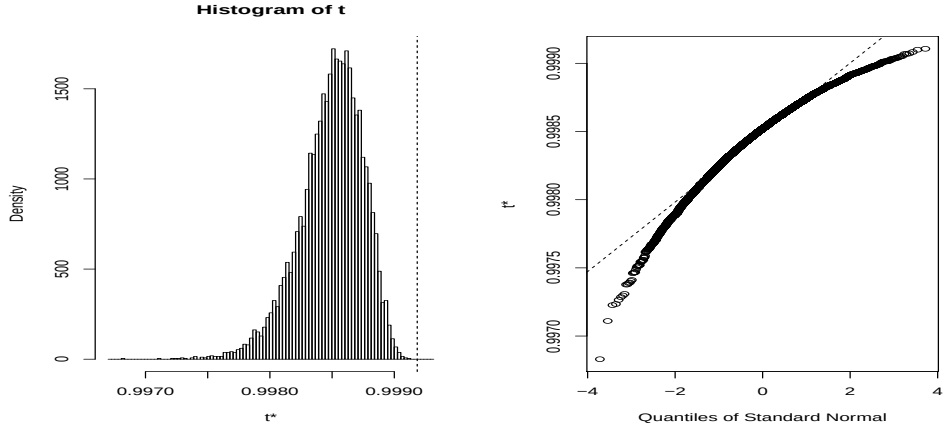

Figure 2. Energy Method, Bootstrap density histogram for $T^\star = \mathcal{R}_n^*$, and quantile plot $Q_{\mathcal{R}_n^*} \times Q_{N(0,1)}$ for $\mathcal{R}_n^*$ distribution versus normal distribution.

The histograms and quantile plot $Q_{T^\star} \times Q_{N(0,1)}$ for $\mathcal{R}_n^*$ coefficient are in Fig. Fig.(2), the Confidence Limits on Tab.(2). All results for linear dependent X and Y are close to 1 and with tiny differences between coefficients. From $\mathcal{R}_n^*(X, Y)$ results it is possible to conclude that the algorithm presents results as predicted by the the Energy Distance theory. In the following sections only the $\mathcal{R}_n^*(X, Y)$ Energy Unbiased Distance is employed, the coefficient $\mathcal{R}_n^2$ does not presented a good performance for linear independent case.

Table 2. Bootstrap statistic computed from original linear dependent dataset, T_0, bias, two-sided 95% Confidence Limits, for coefficients $\mathcal{R}_n^2(X, Y)$ and $\mathcal{R}_n^*(X, Y)$. $n = 270$ rows and $p = 10$

Dist. Coef.	T_0	bias ($T_0 - T^\star$)	basic Conf. Limit	percentile Conf. Limit
$\mathcal{R}_n^2$	0.9995981	-0.0003411565	(0.9997, 1.0002)	(0.9990, 0.9995)
$\mathcal{R}_n^*$	0.9991843	-0.0006920689	(0.9995, 1.0005)	(0.9979, 0.9989)

Comparing Partial Dependent Datasets for $0 \leq \mathcal{R}_n^* \leq 1$**.** On Fig.(3) left panel there are three Bootstrap Density Histograms for the following Distances $\mathcal{R}_n^*(Y, Y)$, $\mathcal{R}_n^*(X, Y)$ and $\mathcal{R}_n^*(X, X)$. This density histogram are obtained from random sets X and Y both composed by a linear combination of a reference vector U_s. This reference vector is a random normal vector $U_s = \mathcal{N}(1, 1)$, mean=1 and sd= 1. The X set is a combination of U_s and a random Gaussian vector $\mathcal{N}_i(0.0\,,\, 0.1)$ (mean=0 and sd=0.1), $X_i = U_s + \mathcal{N}_i(0.0\,,\, 0.1)$. The Y set is also a combination of vector U_s, $Y_i = U_s + \mathcal{N}_i(0.5, 0.25)$. The dimensions used for $X_{n,p}$ and $Y_{n,p}$ are $n = 270$ and $p = 10$.

On the left side of Fig.(3), the first histogram is for $\mathcal{R}_n^*(Y, Y)$ Distance Coefficient, the largest data dispersion histogram in blue color, the latest and tallest is for distance coefficient $\mathcal{R}_n^*(X, X)$, green color. Between them there is the $\mathcal{R}_n^*(X, Y)$ Distance Coefficient comparing X and Y sets, salmon color. The $\mathcal{R}_n^*(X, Y)$ Distance Coefficient appears between $\mathcal{R}_n^*(Y, Y)$ and $\mathcal{R}_n^*(X, X)$ because the reference vector $\mathcal{N}(1, 1)$ is the significant structure behind the X and Y sets, and the difference between them depends on the standard deviation in vectors added to the base.

On the right side panel of Fig.(3) are histograms using another datasets. There is also a random reference set $U_s = \mathcal{N}(2, 4)$ common to both sets, a common random vector for X set $X_s = \mathcal{N}_X(1, 1)$ and $X_i = U_s + X_s + \mathcal{N}_i(1, 0.5)$, where for $X_{n,p}$ $n = 270$ and $p = 10$. The $Y_{n,p}$ is $Y_i = U_s + Y_s + E_p(1.5)$ where $E(1.5)$ is the random exponential with mean $(1/1.5)$ and $Y_s = \mathcal{N}_Y(-1, 0.5)$ is a Gaussian vector.

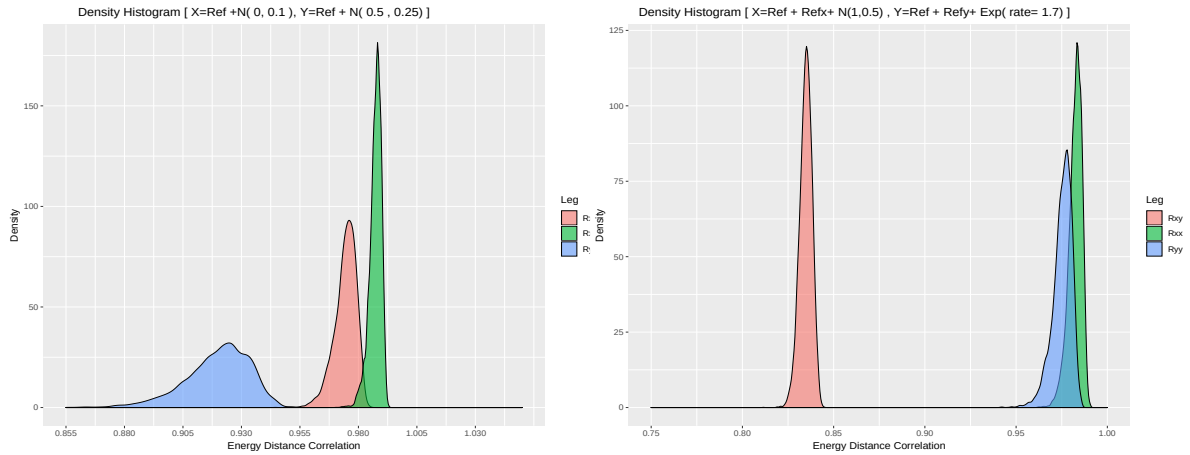

Figure 3. Density histogram for $\mathcal{R}_n^*(X,Y)$ in salmon, $\mathcal{R}_n^*(X,X)$ green and $\mathcal{R}_n^*(Y,Y)$ in blue color.

The $\mathcal{R}_n^*(X,Y)$ Distance Coefficient behaviour in the right side of the Fig.(3) is different from the previous case, the $\mathcal{R}_n^*(X,Y)$ is not only related to the dispersion inside the sets. In this case there is a common structure U_s vector, but there is also a different distributions present in each set, $\mathcal{N}_X(1,1)$ for X and $\mathcal{N}_Y(-1,0.5)$ for Y set. So there is a specific structure in each set and this difference is indicated by the $\mathcal{R}_n^*(X,Y)$ Distance Coefficient.

3.2 Using Energy Distance in Experimental Beam Data

The results presented in this section are obtained from Duarte [14], the FRF dataset are from a metal laminate beam, material characteristics, procedures and measurement details can be found in reference [14]. The healthy structure has 60×10^{-3} kg and the damaged structure is the same structure that had an added mass with 1.2×10^{-3} kg, which represents about 2% of the structure's mass. For the frequency range used here, 2 to 10 kHz, the RSVD method presented poor positive failure identification.

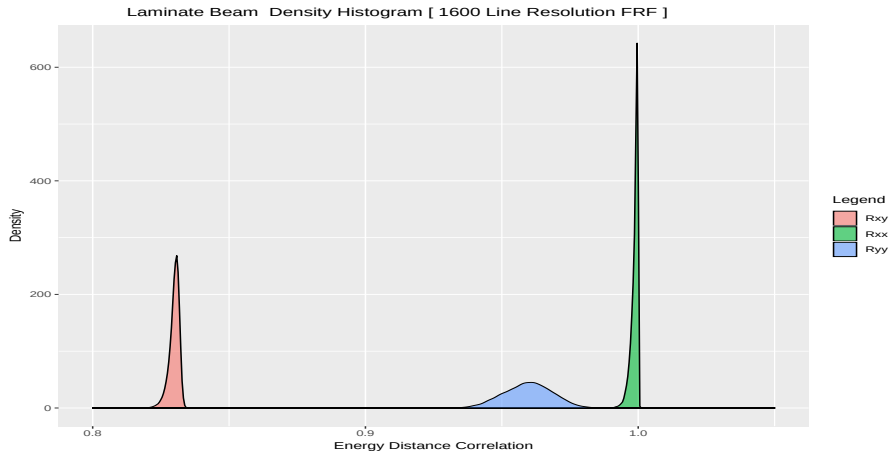

Figure 4. Density Histogram for Laminate beam, frequency range from 2 to 10 kHz (1600 line resolution FRFs). First left salmon histogram, $\mathcal{R}_n^*(X,Y)$, at center $\mathcal{R}_n^*(Y,Y)$ and $\mathcal{R}_n^*(X,X)$, rightmost.

Beam Results for 1600 line resolution FRFs. For this resolution the reference Dataset, $X_{n,p}$, is composed by $n = 401$ rows and $p = 20$ columns, each column is a FRF. The matrix to be tested, $Y_{n,p}$, has same dimensions. Using Unbiased Distance Correlation $\mathcal{R}_n^*(X,Y)$, $\mathcal{R}_n^*(X,X)$ and $\mathcal{R}_n^*(Y,Y)$ the Bootstrap results are on Fig.(4), and on Tab.(3). On Fig.(4) one can see that there is a significant difference between the reference set, X, and the set to be tested, Y. On Fig.(4) and Tab.(3) the Confidence Limits for Unbiased Energy Distance $\mathcal{R}_n^*(X,Y)$ did not overlap the Confidence Limits for $\mathcal{R}_n^*(X,X)$ or $\mathcal{R}_n^*(Y,Y)$. As the X and Y data are taken from the same position on structure so there is a clear failure in Y dataset if X is the reference. The Bootstrap $\mathcal{R}_n^*(X,Y)$ Distance between sets indicates that there

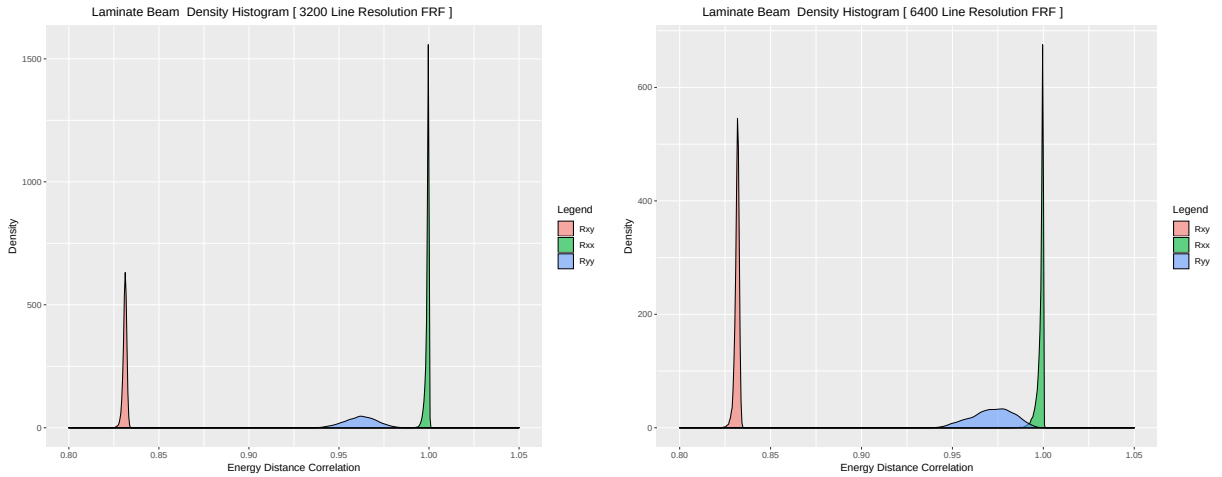

Figure 5. Density Histogram for Laminate beam 2 to 10 kHz (3200 and 6400 line resolution FRFs). In both plots the histogram sequence is: $\mathcal{R}_n^*(X,Y)$, $\mathcal{R}_n^*(Y,Y)$ and $\mathcal{R}_n^*(X,X)$.

is significant structures that are not common.

Table 3. Beam 1600 line resolution FRF Bootstrap statistic, T_0, bias, basic and percentile two-sided 95% Confidence Limits, CL, for coefficients $\mathcal{R}_n^*(X,Y)$, $\mathcal{R}_n^*(X,X)$ and $\mathcal{R}_n^*(Y,Y)$.

Dist. Coef.	T_0	bias $(T_0 - T^\star)$	basic Conf. Lim	percentile Conf. Lim
$\mathcal{R}_n^*(X,Y)$	0.8307100	-0.0008191982	(0.8288357 0.8362371)	(0.8288357 0.8362371)
$\mathcal{R}_n^*(X,X)$	0.9992636	-0.0005576273	(0.9986100 1.0033463)	(0.9951810 0.9999172)
$\mathcal{R}_n^*(Y,Y)$	0.9608105	-0.0012700205	(0.9457125 0.9792952)	(0.9423257 0.9759085)

It is possible to implement a Bootstrap hypothesis testing, but there is a so significant gap between $\mathcal{R}_n^*(X,Y)$ and $\mathcal{R}_n^*(X,X)$ or $\mathcal{R}_n^*(X,Y)$ and $\mathcal{R}_n^*(Y,Y)$ Confidence Limits that there is no sense to implement it. This gap between Bootstrap Distances Confidence Limits indicates that there is relevant statistical differences between X and Y sets that can not explained only by the data dispersion present in the measurement.

Beam Distances for 3200 and 6400 line resolution FRFs. There are results for datasets $X_{n,p}$ and $Y_{n,q}$ using 3200 line resolution FRF where $n = 801$ rows and for 6400 line resolution FRF dataset this with $n = 1601$ rows. Using the Unbiased Distance Correlation $\mathcal{R}_n^*(X,Y)$, $\mathcal{R}_n^*(X,X)$ and $\mathcal{R}_n^*(Y,Y)$ the Bootstrap results are on Fig.(5), and on Tab.(4).

Table 4. Beam 3200 and 6400 line resolution FRF Bootstrap statistic. T_0, bias, basic and percentile two-sided 95% Confidence Limits, for coefficients $\mathcal{R}_n^*(X,Y)$, $\mathcal{R}_n^*(X,X)$ and $\mathcal{R}_n^*(Y,Y)$.

Dist. Coef.	T_0	bias $(T_0 - T^\star)$	basic Conf. Lim	percentile CL
$\mathcal{R}_n^*(X,Y)^{(3200)}$	0.8318232	-0.0007293904	(0.8308066 0.8351339)	(0.8285124 0.8328397)
$\mathcal{R}_n^*(X,X)^{(3200)}$	0.9994893	-0.0003399808	(0.9990535 1.0025498)	(0.9964288 0.9999251)
$\mathcal{R}_n^*(Y,Y)^{(3200)}$	0.9633568	-0.0012471797	(0.9484269 0.9818788)	(0.9448347 0.9782866)
$\mathcal{R}_n^*(X,Y)^{(6400)}$	0.8324080	-0.0008629327	(0.8314283 0.8363740)	(0.8284420 0.8333877)
$\mathcal{R}_n^*(X,X)^{(6400)}$	0.9989766	-0.0003745203	(0.9980177 1.0044378)	(.9935153 0.9999354)
$\mathcal{R}_n^*(Y,Y)^{(6400)}$	0.9743316	-0.0023613615	(0.9576356 0.9994949)	(0.9491683 0.9910276)

In these results there are no significant differences from the previous data for 1600 line resolution FRF case, there are very close results for all Bootstrap statistics. The only remarkable difference could be seen on the upper Confidence Limit Distance $\mathcal{R}_n^*(Y,Y)$ for $Y_{n,q}$ 6400 line resolution set where the upper Bootstrap scattering goes toward 1 and match with $X_{n,p}$ upper Confidence Limit for Distance $\mathcal{R}_n^*(X,X)$, right side panel on Fig.(5). This could be interpreted as a improvement of the Y set measurement with the FRF resolution. Despite that, there is no significant change on Confidence Limit for Distance between $Y_{n,q}$ and $X_{n,p}$ datasets, $\mathcal{R}_n^*(X,Y)$, or its Bootstrap mean value. Its an empirical confirmation that the

structural change in measurement and not the measurement dispel that indicates the damage.

4 Closing Remarks

The clear and undeniable statistical results for the experimental data using Unbiased Correlation Energy Distance $\mathcal{R}_n^*$ multivariate procedure is the contribution of this work. The procedure is simple, fast and precise. There is no identification problem using $\mathcal{R}_n^*$ on available data. There is a clear difference between the health and damaged datasets. There is no dependence on the operator skills to identify operational parameters. So There is no identifications problems found in RSVD method that are consequence of the correct choice of significant singular values. This procedure will improve the Frequency-based Damage Detection method reliability and will allow to create new procedures.

It is important to highlight some important and practical aspects of the outlined method. The Unbiased Energy Distance satisfies all axioms of a metric distance so it can be used as a direct tool to compare different signals. There is no restriction regarding data distribution type, using Bootstrap there is no concern about distributional assumptions to determine the statistic inference for Correlation or Covariance Distance. Determining distribution type and appropriate procedures are time consuming in data analysis. So it can be extended to other Structure Health Monitoring methods where comparing multivariate data is important.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] G. J. Székely and M. L. Rizzo. Partial distance correlation with methods for dissimilarities. *The Annals of Statistics*, 42(6):2382–2412, 2014. doi: 10.1214/14-AOS1255.

[2] R. Lyons. Distance covariance in metric spaces. *Ann. Probab.*, 41(5):3284–3305, 2013. doi: 10.1214/12-AOP803.

[3] D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *The Annals of Statistics*, 41:2263–2291, 2013. doi: 10.1214/13-AOS1140.

[4] G. J. Székely, M. L. Rizzo, and N. K. Bakirov. Measuring and testing independence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794, 2007. doi: 10.1214/009053607000000505.

[5] M. L. Rizzo and G. J. Székely. Energy distance. *WIREs Computational Statistics*, 8:27–38, 2016. doi: 10.1002/wics.1375.

[6] Z. Zhou. Measuring nonlinear dependence in time-series, a distance correlation approach. *Journal of Time Series Analysis*, 33(3):438–457, 2012. doi: 10.1111/j.1467-9892.2011.00780.x.

[7] D. Edelmann, K. Fokianos, and M. Pitsillou. An updated literature review of distance correlation and its applications to time series. *International Statistical Review*, 87(2):237–262, 2019. doi: 10.1111/insr.12294.

[8] G. J. Székely and M. L. Rizzo. Brownian distance covariance. *The Annals of Applied Statistics*, 3 (4):1236–1265, 2009. doi: 10.1214/09-AOAS312.

[9] G. J. Székely and M. L. Rizzo. The distance correlation t-test of independence in high dimension. *Journal of Multivariate Analysis*, 117:193–213, 2013. doi: 10.1016/j.jmva.2013.02.012.

[10] B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979.

[11] A. C. Davison and D.V. Hinkley. *An Introduction to Bootstrap methods with applications to R.* John Wiley & Sons, 2011.

[12] A. C. Davison and D. Kuonen. An introduction to the bootstrap with applications in r. *Statistical Computing & Statistical Graphics Newsletter*, 13(1):6–11, 2002.

[13] A. Canty and B. Ripley. Bootstrap functions (originally by angelo canty for s) version 1.3-25. 26 April 2020 `https://cran.r-project.org/web/packages/boot/index.html`, 2020.

[14] H. V. Duarte, L. V. Donadon, R. P. Ribeiro, and C. M. Possatti jr. Analysis of rsvd algorithm using high frequency frfs in damage detection. In *Proceedings of the XXXVII Iberian Latin-American Congress on Computational Methods in Engineering, ABMEC, Brasilia, DF, Brasil*, pages 1–10, 6-9 November 2016.