Leave One Out Detrended Fluctuation Analysis

R.N. Vargas[a,*], A.C. Paschoarelli Veiga[a] and R.R. Linhares[b]

[a] *Federal University of Uberlândia, Electrical Engineering College, 38408-100, Uberlândia, Brasil*
[b] *Federal University of Uberlândia, Faculty of Mathematics, 38408-100, Uberlândia, Brasil*

Received: 29.07.2022 & Accepted: 13.12.2022

Doi: 10.12693/APhysPolA.143.47
*e-mail: regisnv@yahoo.com.br

A relevant issue in time series analysis is the estimation of long-range dependence, that is, how much future values of a time series depend on current values. One of the ways to verify this dependence is by estimating the Hurst exponent using methods such as detrended fluctuation analysis. Here, we propose a new methodology to estimate the Hurst exponent, named leave one out detrended fluctuation analysis. Furthermore, based on this new estimator for the Hurst exponent, we propose the noise reduction by the leave one out detrended fluctuation analysis method. We apply this new denoising method to electrocardiogram noise reduction. The results presented in this work show that this new methodology outperforms the SureShrink and universal noise reduction methods.

topics: detrended fluctuation analysis (DFA), denoising, electrocardiogram (ECG), wavelets

1. Introduction

This paper addresses two issues related to time series analysis and digital signal processing research, i.e., long-range dependency estimation and *electrocardiogram* (ECG) signal denoising. The *detrended fluctuation analysis* (DFA), proposed by Peng et al. [1], is a method capable of estimating the Hurst exponent, which describes the long-range correlation between time series. Several research areas apply the DFA method, such as arrhythmia detection [2], seismic trace analysis [3], and econophysics [4].

There are also several studies related to the scale range that aim to verify the variation of the scale range according to the signal characteristics, such as the length of the time series and the consideration of uncorrelated data. This approach goes beyond the scope of our paper; see [5] for details.

The second issue addressed by this paper is the ECG signal denoising. The quality of the ECG signal is directly related to the accuracy of the cardiac diagnoses. For this reason, there is a frequent proposition of new and better methods to reduce noise in electrocardiogram signals. In recent years, we can highlight methods that use fractional wavelet transform [6], empirical mode decomposition and discrete wavelet transform [7]. We can also highlight Chatterjee et al., see [8] for details, who discuss the workflow and design principles followed by the recent ECG denoising methods and classify the state-of-the-art methods into different categories for mutual comparison and development of modern methodologies to reduce ECG noise.

The discrete wavelet transform is a very effective tool when we want to mitigate the noise present in the signal; for this reason, it has been applied not only in ECG denoising, but also, e.g., in seismic signal denoising [9] and noise reduction in images [10]. Here, we combine the discrete wavelet transform and DFA to propose a new ECG denoising method.

DFA estimates the Hurst exponent by a sequence of steps. Here we propose a modification in one of these steps, thus presenting a new estimation method. We have denominated this new method as *leave one out detrended fluctuation analysis* (LOO-DFA).

Additionally, we propose an ECG signal noise reduction method based on the LOO-DFA method. We named this new ECG signal denoising method *noise reduction by leave one out detrended fluctuation analysis* (NR-LOO-DFA). We compare the NR-LOO-DFA results with the *Universal* and *SureShrink* wavelet shrinkage methods [11–13].

The paper is organized as follows. We present basic concepts and methodologies in Sect. 2. Section 3 demonstrates the simulation results, and Sect. 4 shows the conclusions.

2. Methodology

Here, we approach the basic concepts of the LOO-DFA and NR-LOO-DFA methods.

2.1. Basic concepts

This subsection introduces two fundamental concepts for the LOO-DFA and NR-LOO-DFA methods, namely the detrended fluctuation analysis (DFA) and the discrete wavelet transform (DWT).

2.1.1. Detrended fluctuation analysis

The detrended fluctuation analysis (DFA) method, proposed by Peng et al. [1], is an approach that estimates the time series long-memory. Given the time series $\{X_t\}_{t=1}^{n}$, DFA consists of five steps.

(i) For each $t \in \{1, 2, \ldots, n\}$, we calculate

$$Y_t = \sum_{j=1}^{t} X_j. \tag{1}$$

(ii) We divide the time series $\{Y_t\}_{t=1}^{n}$ into $\lfloor \frac{n}{l} \rfloor$ nonoverlapping blocks, each containing l observations, where $\lfloor \cdot \rfloor$ indicates the integer part function.

(iii) For each block, one fits a least-square line to the data.

(iv) In each block we calculate

$$Z_t^l = Y_t - Y_t^l, \tag{2}$$

where Y_t^l denotes the adjusted fit on each block.

(v) Let $g(n)$ be the maximum block size. For each $l \in \{4, 5, \ldots, g(n)\}$, we calculate the *root mean square fluctuation*

$$\mathrm{F}(l) = \sqrt{\frac{1}{\widetilde{n}} \sum_{t=1}^{\widetilde{n}} (Z_t^l)^2}, \quad \text{where} \quad \widetilde{n} = l \left\lfloor \frac{n}{l} \right\rfloor. \tag{3}$$

If we apply the natural logarithm to each $F(l)$ and each l, that is, $\ln(F(l))$ versus $\ln(l)$, we will verify the linear relationship between these two quantities. The linear coefficient obtained from the linear regression of these points is an estimate for Hurst's exponent.

2.1.2. Discrete wavelet transform

The wavelet theory consists of the approximation of functions by a linear combination of functions called wavelets. The mother and father wavelets, respectively given by $\psi(\cdot)$ and $\phi(\cdot)$, are real functions with respect to ψ and ϕ; $\phi \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$, $\int_{\mathbb{R}} \mathrm{d}t\, \psi(t) = 0$ and $\int_{\mathbb{R}} \mathrm{d}t\, \phi(t) = 1$.

Usually, the mother wavelet is bounded and centered at the origin, and $\psi(\cdot) \to 0$ when $\mid t \mid \to \infty$. Considering $j, k \in \mathbb{Z}$, these functions relate to each other with the equations $\psi(t) = \sqrt{2} \sum_k h_k \phi(2t - k)$ and $\phi(t) = \sqrt{2} \sum_k g_k \phi(2t - k)$, where g_k and h_k are the respective low-pass filter and high-pass filter coefficients satisfying $h_k = (-1)^k g_{1-k}$.

From there on, it is possible to build a wavelet sequence given by

$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\, \psi(2^j t - k), \tag{4}$$

$$\phi_{j,k}(t) = 2^{\frac{j}{2}}\, \phi(2^j t - k). \tag{5}$$

Definition 1 (DWT) If $\boldsymbol{y} = (y_0, y_1, \ldots, y_{N-1})'$ is a signal with length N, such that $N = 2^J$, $J \in \mathbb{N}$. Then, DWT of $\boldsymbol{y}$, according to its mother wavelet $\psi(\cdot)$, is given by

$$d_{j,k} = \sum_{t=0}^{N-1} y_t \psi_{j,k}(t), \tag{6}$$

where $j = 0, 1, 2, \ldots, J-1$ and $k = 0, 1, 2, \ldots, 2^j - 1$.

Definition 1 formalizes a construction that maps the data from the time domain to the wavelet domain.

In practice, we calculate DWT using the pyramid algorithm given by Meyer [14] instead of the definition 1. This algorithm consists of an iterative application of high-pass and low-pass filters that return a detail wavelet coefficients set $\{d_{j,k}\}$. To minimize or remove a noise signal, we need to decide which one of these detail wavelet coefficients should have their magnitude reduced or eliminated. Then, we need to apply the inverse discrete wavelet transform (IDWT) in these previously processed coefficients.

2.2. LOO-DFA methodology

The LOO-DFA approach proposes a way to minimize the impact caused by random data errors by removing a single data at a time, aiming to prevent overfitting and allowing greater generalization for the estimators. To apply the LOO-DFA method, we changed the last step of DFA (see Sect. 2.1.1). In addition to the usual linear regression of $\ln(F(l))$ over $\ln(l)$ with $l \in \{4, 5, \ldots, g(n)\}$, we calculate another $g(n) - 3$ linear regressions, leaving out one l per time. We denote: α_1 — slope of linear regression considering all $l \in \{4, 5, \ldots, g(n)\}$ values, α_2 — slope of linear regression considering $l \in \{5, 6, 7, \ldots, g(n)\}$, α_3 — slope of linear regression considering $l \in \{4, 6, 7, \ldots, g(n)\}$, and so on, up to $\alpha_{g(n)-2}$, which is the slope of linear regression considering $l \in \{4, 5, 6, \ldots, g(n) - 1\}$. Then, we obtain the Hurst exponent (denoted by $\hat{H}$) by the following equation

$$\hat{H} = \operatorname*{argmin}_{s \in \mathbb{R}} \left[\sum_{j=1}^{g(n)-2} |\alpha_j - s| \right]. \tag{7}$$

2.3. NR-LOO-DFA methodology

To build the NR-LOO-DFA method, we need the pre-processing stage detailed below.

2.3.1. Pre-processing stage

The pre-processing related to the NR-LOO-DFA method consists of three hundred iterations, each of which comprehends the following steps.

(i) Randomly choose a size signal $N \in \mathbb{N}$.

(ii) Generate a random N-length additive white Gaussian noise (AWGN) $\boldsymbol{e}$. Then, the noise $\boldsymbol{e}$ is applied in the clean ECG synthetic signal $\boldsymbol{x}$ to obtain the noisy signal $\boldsymbol{y}$.

(iii) Apply DWT in the noisy signal $\boldsymbol{y}$ (see Sect. 2.1.2) and estimate the standard deviation of the noise $\hat{\sigma}$ with

$$\hat{\sigma} = \text{median}\big(\{d_{J-1,k} : k = 0, 1, \ldots, 2^{J-1} - 1\}\big). \tag{8}$$

See [11–13] for details.

(iv) Let $\lambda \in \mathbb{R}$ be the threshold. Let $\eta_S(d_{j,k}, \lambda)$ be the soft-thresholding function. Let $\hat{\boldsymbol{x}}$ be the signal obtained when we apply IDWT with $\eta_S(d_{j,k}, \lambda)$ instead of $d_{j,k}$. In the fourth step, we find the threshold $\hat{\lambda}$ that satisfies

$$\eta_S(d_{j,k}, \lambda) = \begin{cases} \text{sgn}(d_{j,k})\,(|d_{j,k}|{-}\lambda), & \text{if } |d_{j,k}| > \lambda, \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

$$\hat{\lambda} = \underset{\lambda \in \mathbb{R}}{\text{argmin}} \Big[\frac{1}{N}\sum_{j=1}^{N}(x_j - \hat{x}_j)^2\Big]. \tag{10}$$

(v) Let $\hat{\boldsymbol{x}}_{\hat{\lambda}}$ be the signal obtained when we apply IDWT over $\eta_S(d_{j,k}, \hat{\lambda})$. In the fifth step, we estimate the Hurst exponent $H_{\hat{\boldsymbol{e}}}$ using the LOO-DFA method (see Sect. 2.2 for details) for the noise estimate $\hat{\boldsymbol{e}}$. It is given as

$$H_{\hat{\boldsymbol{e}}} = \text{LOO-DFA}(\hat{\boldsymbol{e}}), \quad \text{where} \quad \hat{\boldsymbol{e}} = \boldsymbol{y} - \hat{\boldsymbol{x}}_{\hat{\lambda}}. \tag{11}$$

This gives us one training set row composed of N, $\hat{\sigma}$, $H_{\hat{\boldsymbol{e}}}$. Repeating this procedure, we obtain a three-hundred-size training set (Fig. 1 illustrates this process). It is possible to correlate the values N and $\hat{\sigma}$ with the value $H_{\hat{\boldsymbol{e}}}$, as shown below as

$$H_{\hat{\boldsymbol{e}}} = A + B\,\hat{\sigma} + C\,N, \tag{12}$$

where $A = 6.292\,10^{-1}$, $B = -1.038\,10^{-1}$, and $C = 7.093\,10^{-7}$. We use this equation in the NR-LOO-DFA method.

We kept the set with three hundred rows because we noticed that the parameters A, B, and C in (12) present changes from the sixth decimal point as more data is added, which in our view indicated their convergence.

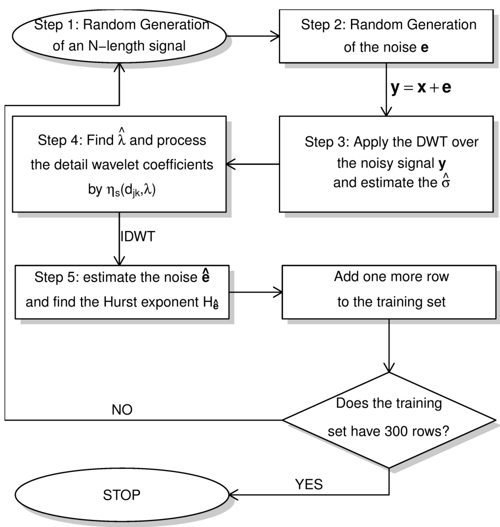

Fig. 1. Flowchart illustrating a training set building.

2.3.2. NR-LOO-DFA

Let $\boldsymbol{y}$ be an N-length noisy signal obtained from the addition of the additive white Gaussian noise $\boldsymbol{e}$ to the clean signal $\boldsymbol{x}$. The method NR-LOO-DFA consists of the following steps.

First, we apply DWT in the noisy signal $\boldsymbol{y}$ (see Sect. 2.1.2) and estimate the standard deviation of the noise $\hat{\sigma}$ by (8). Then, we find the value of $H_{\hat{\boldsymbol{e}}}$ by (12).

The value $H_{\hat{\boldsymbol{e}}}$ is the expected Hurst exponent value for the noise estimate. We need to do the same evaluation as in the current denoising process. The current noise estimate $\hat{\boldsymbol{e}}_c$ is given by

$$\hat{\boldsymbol{e}}_c = \boldsymbol{y} - \hat{\boldsymbol{x}}_c, \tag{13}$$

where $\hat{\boldsymbol{x}}_c$ is the clean signal estimate for the current denoising process, obtained when we apply IDWT with $\eta_S(d_{j,k}, \lambda)$ instead of $d_{j,k}$.

In the second step, we estimate the Hurst exponent $H_{\hat{\boldsymbol{e}}_c}$ by the LOO-DFA method.

In the third step, we find $\hat{\lambda}_c$ using

$$\hat{\lambda}_c = \underset{\lambda \in \mathbb{R}}{\text{argmin}} \Big| H_{\hat{\boldsymbol{e}}} - H_{\hat{\boldsymbol{e}}_c} \Big|. \tag{14}$$

Thus, $\hat{\lambda}_c$ guarantees the best approximation between the two estimates for Hurst's exponents $H_{\hat{\boldsymbol{e}}}$ and $H_{\hat{\boldsymbol{e}}_c}$. We interpret it as an agreement between the value expected for Hurst's exponent, $H_{\hat{\boldsymbol{e}}}$, and the value we estimate for this same exponent in the current problem, $H_{\hat{\boldsymbol{e}}_c}$. The $\hat{\lambda}_c$ value that approximates these values as close as possible will be the chosen threshold, then we obtain the clean signal estimate $\hat{\boldsymbol{x}}$ by

$$\hat{\boldsymbol{x}} = \text{IDWT}(\eta_S(d_{j,k}, \hat{\lambda}_c)). \tag{15}$$

3. Simulation results

This section presents simulations results for the methods LOO-DFA and NR-LOO-DFA.

3.1. LOO-DFA simulation results

In this subsection, we compare the performance of the LOO-DFA method with the traditional DFA method. To make this comparison we use a process known as Monte Carlo simulation.

Monte Carlo simulation is an approach that seeks to find a populational parameter based on a significant sample. In our case, we will estimate the bias parameter.

To investigate the LOO-DFA performance compared to the traditional DFA method, we generate time series from fractional Gaussian noise (FGN) models, with 100 replications and length $n \in \{1000, 5000, 10000, 15000\}$. Here, we adopt the largest window as $g(n) = \lfloor \ln(n)^2 \rfloor$; see [15] for details. For each method, we calculate the empirical values of the mean, the bias, the mean square error (MSE), and the variance (Var) values. Tables I–IV present the experiment results.

TABLE I

New leave one out DFA compared with the traditional DFA, fractional Gaussian noise with length $n = 1000$.

Largest window $g(n)$	Hurst exponent	Leave one out DFA				Traditional DFA			
		Mean	Bias	MSE	Var	Mean	Bias	MSE	Var
	0.10	0.169632179	0.069632179	0.004972196	0.000124804	0.170018356	0.070018356	0.00502968	0.000128393
	0.15	0.210162786	0.060162786	0.003812454	0.000194841	0.210965187	0.060965187	0.003916466	0.000201729
	0.20	0.251784569	0.051784569	0.002953657	0.000274763	0.253060705	0.053060705	0.003098374	0.000285794
	0.25	0.294424126	0.044476698	0.002330928	0.000361036	0.29620223	0.046264597	0.002507832	0.000376955
	0.30	0.338003008	0.039011171	0.001891823	0.000452115	0.340289927	0.041183502	0.002090619	0.000472062
	0.35	0.38242132	0.035140492	0.001590635	0.000544942	0.385228294	0.037729622	0.001803823	0.000568475
	0.40	0.427600344	0.032154136	0.001391903	0.000636489	0.430927267	0.03509653	0.001613928	0.000664073
	0.45	0.473454447	0.029876695	0.001268712	0.00072586	0.477302938	0.033039273	0.001495101	0.000757223
$\ln(n)^2$	0.50	0.519906437	0.028410436	0.00119976	0.00081161	0.524277938	0.031685127	0.001427694	0.000846743
	0.55	0.566883982	0.027603171	0.00116945	0.000893314	0.571781586	0.030888562	0.001396953	0.000931834
	0.60	0.614323813	0.027363692	0.001165876	0.000970409	0.619749898	0.030435598	0.001391914	0.001011975
	0.65	0.66216876	0.027665811	0.001180181	0.001042527	0.668125763	0.030426111	0.001404421	0.001086745
	0.70	0.710370808	0.028187062	0.001204843	0.001108373	0.716860117	0.03071319	0.001428151	0.001155442
	0.75	0.758898978	0.028668347	0.001234383	0.00116686	0.76591772	0.03116731	0.001457324	0.001216112
	0.80	0.807762807	0.029024615	0.001260448	0.001212309	0.815305769	0.031663722	0.00148407	0.001262428
	0.85	0.858320267	0.03381749	0.001714681	0.001662075	0.86670731	0.036578122	0.002004313	0.001742604
	0.90	0.907033676	0.034126624	0.001746118	0.001713783	0.915950668	0.036879297	0.002024431	0.001787886
	0.95	0.955951777	0.034286532	0.001775008	0.001757156	0.965404237	0.037136398	0.002041381	0.001822314

TABLE II

New leave one out DFA compared with the traditional DFA, fractional Gaussian noise with length $n = 5000$.

Largest window $g(n)$	Hurst exponent	Leave one out DFA				Traditional DFA			
		Mean	Bias	MSE	Var	Mean	Bias	MSE	Var
	0.10	0.154747391	0.054747391	0.003021767	0.000024737	0.155052562	0.055052562	0.003055754	0.000025222
	0.15	0.197836942	0.047836942	0.00232913	0.000041168	0.198414266	0.048414266	0.002385542	0.000042021
	0.20	0.241813488	0.041813488	0.001807914	0.000060147	0.2427076	0.0427076	0.001884834	0.000061510
	0.25	0.286604752	0.036604752	0.001420081	0.000080983	0.287817753	0.037817753	0.001512283	0.000082930
	0.30	0.332109084	0.032109084	0.001133167	0.000103206	0.333639791	0.033639791	0.00123627	0.000105692
	0.35	0.37822876	0.02822876	0.000921974	0.000126375	0.380079144	0.030079144	0.001032797	0.000129335
	0.40	0.424883476	0.024957713	0.000767749	0.000150062	0.427051575	0.027091137	0.000883746	0.000153494
	0.45	0.471999317	0.022516378	0.000656162	0.000173931	0.474482787	0.024678811	0.000775497	0.000177869
$\ln(n)^2$	0.50	0.51951038	0.020968895	0.000576475	0.000197798	0.522307784	0.023207751	0.000697827	0.000202212
	0.55	0.567360045	0.019830949	0.000520595	0.000221439	0.570470095	0.022181496	0.000643073	0.000226311
	0.60	0.615498728	0.01913246	0.000482446	0.000244683	0.618920906	0.021427067	0.000605487	0.000249986
	0.65	0.663885883	0.018560489	0.000457576	0.000267433	0.667618136	0.02096944	0.000580764	0.000273097
	0.70	0.712484178	0.018082363	0.000442528	0.000289569	0.716525347	0.020652982	0.000565711	0.000295579
	0.75	0.761261516	0.017780921	0.000434951	0.000311242	0.76561019	0.020403563	0.000558076	0.000317574
	0.80	0.81018874	0.017699325	0.000433657	0.000333178	0.814841588	0.020229914	0.0005567	0.000339825
	0.85	0.857220669	0.016386888	0.000416016	0.000367554	0.862187326	0.018176942	0.000519017	0.000374229
	0.90	0.906746355	0.016337252	0.000417358	0.000375601	0.912028652	0.01817276	0.000522979	0.000382112
	0.95	0.956412874	0.016254587	0.000419451	0.000382148	0.962012757	0.018241546	0.000529024	0.000388604

Figures 2–5 show the bias for the individual executions, the DFA method in the horizontal axis, and the LOO-DFA method in the vertical axis. We can see that most of the points are below the identity line (gray line), indicating a lower bias value for the LOO-DFA method. For $n = 1000$, $n = 5000$, $n = 10000$, and $n = 15000$, 77.61%, 86.88%, 90%, and 93.5% of the executions, respectively, showed a DFA bias greater than the LOO-DFA bias.

We can see that the LOO-DFA method presents the best performance compared to the traditional DFA method. We can resort to the ANOVA test to ensure a significant difference between method performances; see [16] and [17] for details.

The results of this analysis show that there is a significant difference between the LOO-DFA bias when compared to the DFA bias. Table V shows a p value less than 0.05, which means a difference between the groups.

TABLE III

New leave one out DFA compared with the traditional DFA, fractional Gaussian noise with length $n = 10000$.

Largest window $g(n)$	Hurst exponent	Leave one out DFA				Traditional DFA			
		Mean	Bias	MSE	Var	Mean	Bias	MSE	Var
	0.10	0.14847571	0.04847571	0.002366881	0.000017158	0.148754658	0.048754658	0.002394289	0.0000174471
	0.15	0.191958353	0.041958353	0.001789245	0.000029032	0.19247952	0.04247952	0.001833765	0.0000295508
	0.20	0.236371835	0.036371835	0.001365115	0.000042631	0.237164892	0.037164892	0.001424273	0.0000434782
	0.25	0.281618247	0.031618247	0.001056593	0.000057454	0.282687201	0.032687201	0.001126496	0.000058630
	0.30	0.327591805	0.027591805	0.000833591	0.000073013	0.328934400	0.028934400	0.000910999	0.000074545
	0.35	0.374191569	0.024214199	0.000673336	0.000088994	0.375806297	0.025806297	0.000755923	0.000090867
	0.40	0.421329026	0.021588285	0.000559018	0.000105143	0.423214337	0.023412759	0.000645148	0.000107316
	0.45	0.468927727	0.019474147	0.000478240	0.000121193	0.471080994	0.021532355	0.000566844	0.000123672
$\ln(n)^2$	0.50	0.516919489	0.017786878	0.000421916	0.000137017	0.519338920	0.020052287	0.000512361	0.000139764
	0.55	0.565245525	0.016486940	0.000383378	0.000152477	0.567929941	0.018905322	0.000475388	0.000155460
	0.60	0.613855530	0.015677813	0.000357762	0.000167461	0.616803899	0.018052666	0.000451323	0.000170659
	0.65	0.662706797	0.015230659	0.000341516	0.000181872	0.665917243	0.017547238	0.000436788	0.000185283
	0.70	0.711759932	0.014981687	0.000331962	0.000195623	0.715230870	0.017273016	0.000429238	0.000199251
	0.75	0.760975126	0.014852554	0.000326944	0.000208577	0.764705720	0.017174421	0.000426559	0.000212425
	0.80	0.810301426	0.014761404	0.000324294	0.000220378	0.814291072	0.017141093	0.000426434	0.000224443
	0.85	0.858209069	0.012727241	0.000270590	0.000205254	0.862441356	0.014866938	0.000361170	0.000208467
	0.90	0.907835716	0.012782651	0.000276950	0.000217728	0.912319327	0.015000363	0.000370455	0.000220898
	0.95	0.957574223	0.012987327	0.000285562	0.000230498	0.962307902	0.015179477	0.000382819	0.000233671

TABLE IV

New leave one out DFA compared with the traditional DFA, fractional Gaussian noise with length $n = 15000$.

Largest window $g(n)$	Hurst exponent	Leave one out DFA				Traditional DFA			
		Mean	Bias	MSE	Var	Mean	Bias	MSE	Var
	0.10	0.145836997	0.045836997	0.002113188	0.000012280	0.146075016	0.046075016	0.002135204	0.000012421
	0.15	0.18966781	0.03966781	0.001594057	0.000020729	0.190120623	0.040120623	0.001630431	0.000020977
	0.20	0.234391315	0.034391315	0.001212722	0.000030262	0.235085422	0.035085422	0.001261316	0.000030635
	0.25	0.279914771	0.029914771	0.000934940	0.000040451	0.280846270	0.030846270	0.000992020	0.000040937
	0.30	0.326123363	0.026123363	0.000732872	0.000050951	0.327292611	0.027292611	0.000795927	0.000051556
	0.35	0.372920694	0.022920694	0.000586262	0.000061519	0.374326500	0.024326500	0.000653417	0.000062261
	0.40	0.420219405	0.020276411	0.000480132	0.000072028	0.421862123	0.021896263	0.000550112	0.000072888
	0.45	0.467946019	0.018107776	0.000403580	0.000082344	0.469824989	0.019934439	0.000475521	0.000083324
$\ln(n)^2$	0.50	0.516036796	0.016342081	0.000348635	0.000092380	0.518150945	0.018390687	0.000422013	0.000093491
	0.55	0.564436984	0.014900823	0.000309490	0.000102085	0.566785122	0.017142207	0.000384043	0.000103336
	0.60	0.613100436	0.013826350	0.000281944	0.000111437	0.615680923	0.016148801	0.000357585	0.000112822
	0.65	0.661987278	0.013022499	0.000262907	0.000120416	0.664799145	0.015468195	0.000339717	0.000121922
	0.70	0.711065201	0.012454252	0.000250120	0.000128971	0.714107423	0.014960195	0.000328313	0.000130600
	0.75	0.760309094	0.012033125	0.000241937	0.000137030	0.763580484	0.014633589	0.000321817	0.000138775
	0.80	0.809703994	0.011805027	0.000237091	0.000144367	0.813202831	0.014455108	0.000319073	0.000146221
	0.85	0.857428248	0.010611263	0.000194738	0.000140969	0.861205610	0.013107460	0.000268281	0.000144157
	0.90	0.907061755	0.010702423	0.000197844	0.000149470	0.911068167	0.013178433	0.000273977	0.000153003
	0.95	0.956798307	0.010903995	0.000202965	0.000158332	0.961033938	0.013319697	0.000282301	0.000162175

TABLE V

ANOVA considering LOO-DFA and DFA bias. Here df stands for degrees of freedom.

Source of variation	Sum of squares	df	Mean squares	F	Sig
Between the groups	0.014	1	0.013958	41.89	9.99×10^{-11}
Within the groups	4.798	14398	0.000333		
Total	4.812	14399			

TABLE VI

Electrocardiogram denoising results.

Standard deviation of noise	Mean squared error average			Mean squared error variance		
	SureShirink	Universal	NR-LOO-DFA	SureShirink	Universal	NR-LOO-DFA
0.025	0.000271374	0.001306042	0.000254775	0.000000000759	0.000000008046	0.000000000402
0.050	0.000945719	0.003674543	0.000900410	0.000000008110	0.000000081021	0.000000004751
0.075	0.001879179	0.006591046	0.001865236	0.000000037633	0.000000250311	0.000000023306

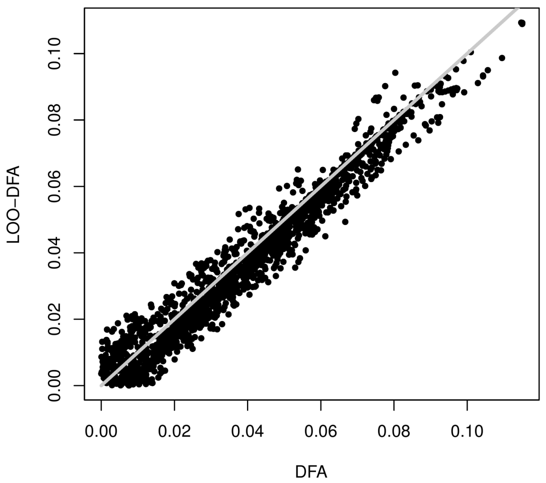

Fig. 2. Bias value for each execution, $n = 1000$. Here, we consider the bias for all Hurst exponents that generate the consolidated data in Table I, that is, 18000 points representing the 100 replications for each of the 18 estimates for the Hurst exponent (which range from 0.10 to 0.95, considering a step of 0.05).

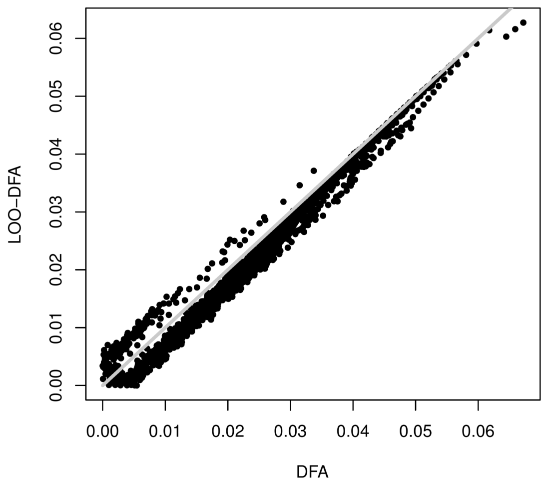

Fig. 4. Bias value for each execution, $n = 10000$. Here, we consider the bias for all Hurst exponents that generate the consolidated data in Table III, that is, 18000 points representing the 100 replications for each of the 18 estimates for the Hurst exponent (which range from 0.10 to 0.95, considering a step of 0.05).

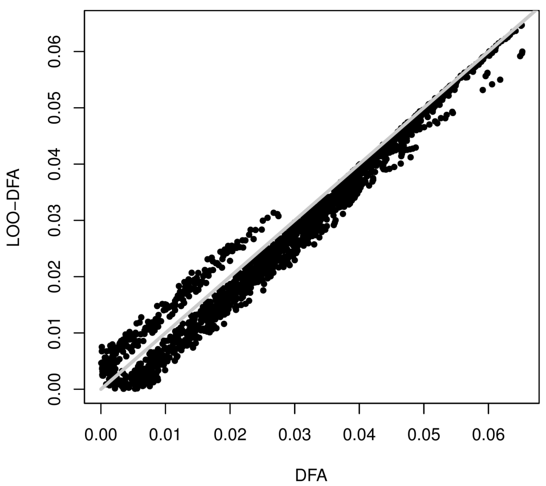

Fig. 3. Bias value for each execution, $n = 5000$. Here, we consider the bias for all Hurst exponents that generate the consolidated data in Table II, that is, 18000 points representing the 100 replications for each of the 18 estimates for the Hurst exponent (which range from 0.10 to 0.95, considering a step of 0.05).

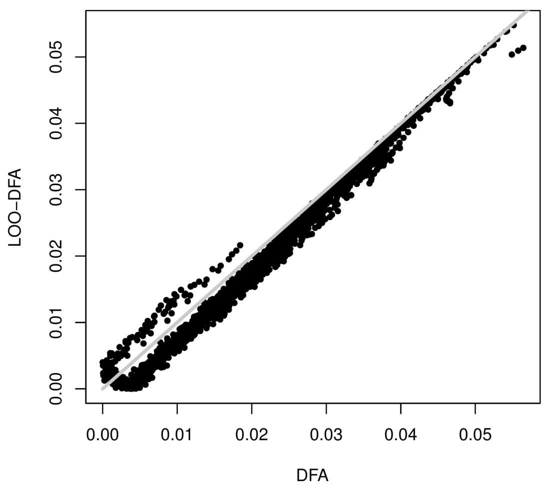

Fig. 5. Bias value for each execution, $n = 15000$. Here, we consider the bias for all Hurst exponents that generate the consolidated data in Table IV, that is, 18000 points representing the 100 replications for each of the 18 estimates for the Hurst exponent (which range from 0.10 to 0.95, considering a step of 0.05).

TABLE VII

ANOVA considering NR-LOO-DFA, SureShrink and Universal mean squared errors. Here df stands for degrees of freedom.

Source of variation	Sum of squares	df	Mean squares	F	Sig
Between the groups	0.0016	2	0.001	423	2.6×10^{-130}
Within the groups	0.0017	897	1.9×10^{-6}		
Total	0.0033	899			

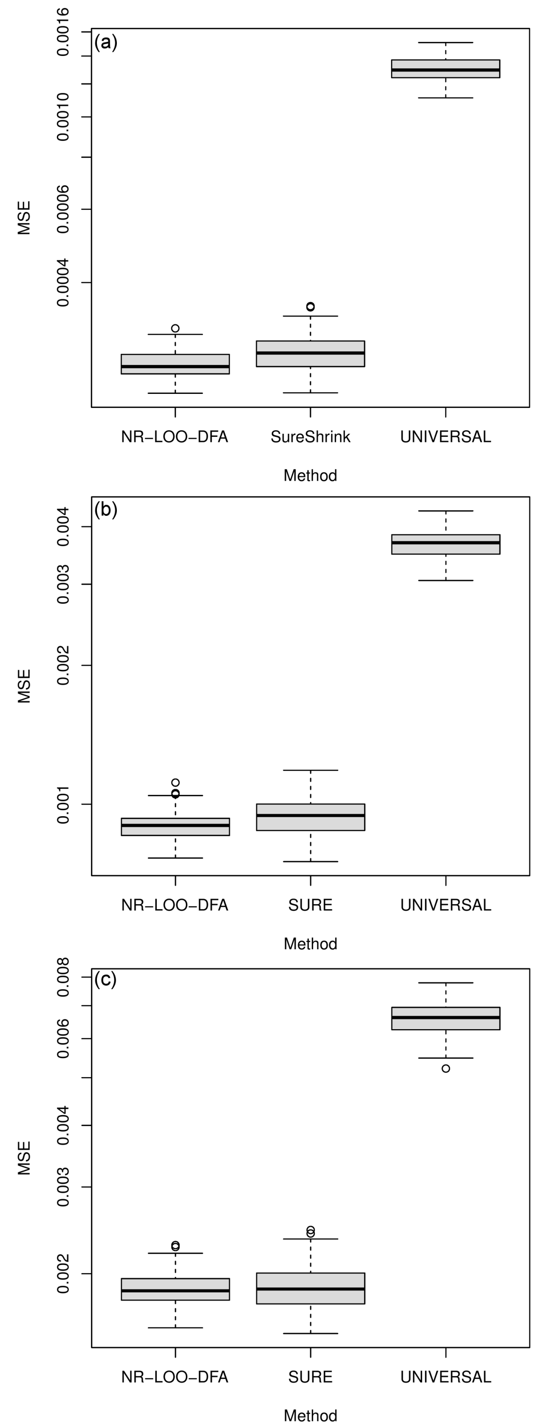

Fig. 6. Box plot considering a noice standard deviation equal to (a) 0.025, (b) 0.050 and (c) 0.075 (see Table VI).

3.2. NR-LOO-DFA simulation results

This subsection presents the simulation results for the method NR-LOO-DFA. This simulation uses synthetic ECG signals corrupted by additive white Gaussian noises (AWGN) with standard deviations of 0.025, 0.050, and 0.075. We compare NR-LOO-DFA with the SureShrink and Universal wavelet shrinkage methods. Table VI shows the simulation results taking into account the average and variance of the mean squared error of the denoising processes. For each standard deviation, we generate one hundred AWGN noises. Figure 6a, 6b and 6c shows the box plot for each method considering the standard deviations 0.025, 0.050, and 0.075,

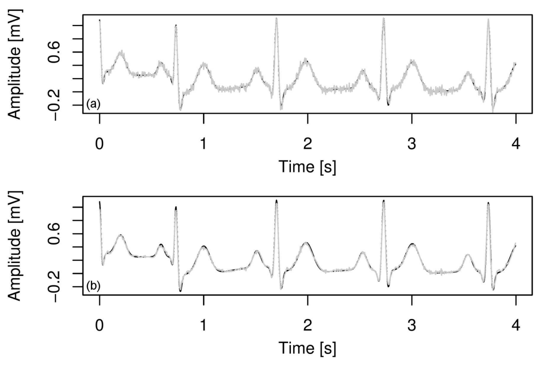

Fig. 7. (a) Simulation of a noisy signal (gray) and a clean signal (black). (b) Noisy signal smoothed by the NR-LOO-DFA method with a synthetic ECG signal.

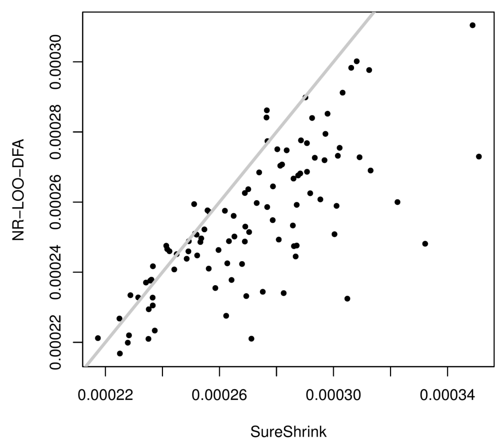

Fig. 8. MSE value for each execution, $sd = 0.025$.

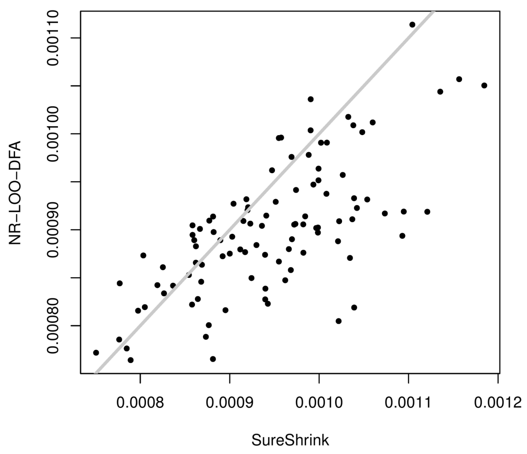

Fig. 9. MSE value for each execution, $sd = 0.050$.

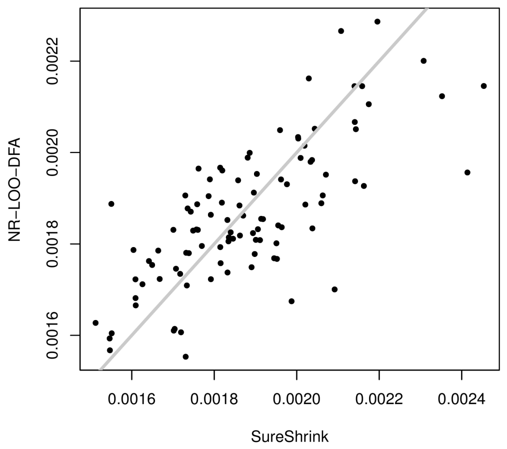

Fig. 10. MSE value for each execution, $sd = 0.075$.

respectively. Figure 7 shows an example of ECG noise reduction using the NR-LOO-DFA method. The analysis of variance presented in Table VII shows significant differences between the groups since the p value is less than 0.05.

Figures 8–10 show MSE for individual executions, the SureShrink method on the horizontal axis, and the NR-LOO-DFA method on the vertical axis. We can see that most of the points are below the identity line (gray line), indicating a lower MSE value for the NR-LOO-DFA method. For $sd = 0.025$, $sd = 0.050$, and $sd = 0.075$, respectively 82%, 70%, and 51% of executions showed SureShrink MSE bigger than NR-LOO-DFA MSE.

4. Conclusions

Here, we presented a new estimation method for Hurst's exponent (i.e., LOO-DFA) and a new ECG denoising method (i.e., NR-LOO-DFA).

Variance analysis and individual performance verification show that the LOO-DFA method presents a better performance compared to the traditional DFA method.

Regarding the ECG denoising context, we observed, by Monte Carlo simulation, that the NR-LOO-DFA method outperforms the SureShrink and Universal wavelet shrinkage methods, which are two methods already established in the denoising literature.

We also verified this best performance of the NR-LOO-DFA method using boxplot graphics and individual performance verification in each execution.

These results reinforce our point of view about the possibility of using these methods in research areas of the long-range dependence analysis and the ECG signal denoising.

References

[1] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, *Phys. Rev. E* **49**, 1685 (1994).

[2] R.N. Vargas, A.C.P. Veiga, R.R. Linhares, *Model Assist. Stat. Appl.* **16**, 189 (2021).

[3] F. Li, B. Zhang, S. Verma, K.J. Marfurtet, *Explor. Geophys.* **49**, 450 (2018).

[4] D. Grech, Z. Mazur, *Phys. A Stat. Mech. App.* **336**, 133 (2004).

[5] D. Grech, Z. Mazur, *Phys. A Stat. Mech. App.* **392**, 2384 (2013).

[6] I. Houamed, L. Saidi, F. Srairi, *Res. Biomed. Eng.* **36**, 349 (2020).

[7] A.K. Dwivedi, H. Ranjan, A. Menon, P. Periasamy, *Circuits, Syst. Signal Process.* **40**, 827 (2021).

[8] S. Chatterjee, R.S. Thakur, R.N. Yadav, L. Gupta, D.K. Raghuvanshi, *IET Signal Process.* **14**, 569 (2020).

[9] J.J. Galiana-Merino, J.L. Rosa-Herranz, S. Rosa-Cintas, J.J. Martinez-Espla, *Compt. Phys. Commun.* **184**, 162 (2013).

[10] B.M. Ferzo, F.M. Mustafa, in: *2020 Int. Conf. on Computer Science and Software Engineering (CSASE)*, IEEE, 2020, p. 106.

[11] D.L. Donoho, I.M. Johnstone, *C. R. séances Acad. Sci. Sér. 1 Math.* **319**, 1317 (1994).

[12] D.L. Donoho, J.M. Johnstone, *Biometrika* **81**, 425 (1994).

[13] D.L. Donoho, I.M. Johnstone, *J. Am. Stat. Assoc.* **90**, 1200 (1995).

[14] Y. Meyer, *Wavelets: Algorithms and Applications*, Siam, Philadelphia (PA) 1993.

[15] R.R. Linhares, *Int. J. Cardiol.* **224**, 27 (2016).

[16] S. Ari, M.K. Das, A. Chacko, *Compt. Biol. Med.* **43**, 649 (2013).

[17] Z. Wang, Z. Wang, F. Wan, C.M. Wong, L. Zhang, *Compt. Biol. Med.* **77**, 195 (2016).