

A New and Fast Index for the Quantification of Short Range Self-Similarity in RR Time Series

M A García-González, M Fernández-Chimeno, J Ramos-Castro

Group of Biomedical and Electronic Instrumentation, Department of Electronic Engineering,
Technical University of Catalonia (UPC), Barcelona, Spain

Abstract

We propose a new and very fast index (the frequency of sign changes of the mirrored differences or *fscmd*) with good correlation with the short range scaling exponent (α_1) estimated among scales 4 to 16 of the DFA. *fscmd* computes the relative number of sign changes of the difference of the RR time series minus their corresponding reversed RR time series after a moving average detrending procedure is applied using a window of 30 samples. Linear regression results with simulated time series with Fractional Brownian Noise and with actual time series using the Fantasia (FT), Normal Sinus Rhythm RR time series (NSR) and Congestive Heart Failure RR interval (CHF) databases after artifact correction show good agreement between *fscmd* and α_1 . Finally, Mann-Whitney Rank Sum tests applied to α_1 and *fscmd* when comparing NSR and CHF databases show very significant differences ($p < 0.001$) between groups for both indices.

1. Introduction

Detrended fluctuation analysis (DFA) aims to quantify the self affinity of time series and was first proposed to analyze RR time series by Peng et al [1]. The dynamics of the signal can be summarized by fitting scaling exponents in different range scales. The short range exponent (α_1) and large range exponent (α_2) have gained attention in heart rate variability studies for its ability of dealing with the nonstationarity of time series and for distinguishing between normal and pathological states [2].

One drawback of DFA is its computational cost when using long time series as is the case of 24 hour RR time series [3] so it is interesting to find alternative and fast indices with a high correlation with those obtained in the DFA. The aim of this work is to propose a very fast index with a high level of agreement with α_1 .

2. Materials and methods

2.1. Estimation of the short range exponent by DFA

Let's be $X = \{x(1), x(2), \dots, x(N)\}$ a time series with N samples. The first step of the DFA is to compute the mean value of the time series and integrate the time series without its mean:

$$y(j) = \sum_{i=1}^j (x(i) - \bar{x}) \quad (1)$$

Next, the integrated time series is divided in boxes of equal length n . For each box, a trend ($y_n(k)$) is obtained and the following statistic is computed:

$$F(n) = \sqrt{\frac{1}{n} \sum_{k=1}^n (y(k) - y_n(k))^2} \quad (2)$$

$F(n)$ is obtained for all the available boxes. Finally, the average of $F(n)$ is computed ($\overline{F(n)}$). By changing n , $\overline{F(n)}$ changes. The scaling exponent is obtained as the slope of a linear fitting of the log-log plot of n versus $\overline{F(n)}$ for a certain range of n . [1] defined α_1 in the range $4 \leq n \leq 16$ so this is the range that we have employed in this work. Moreover, we have used a linear detrending to compute $F(n)$ although other detrending methods are feasible.

In the case of a very long time series, the computation of the trend and $F(n)$ is performed in a very large number of boxes so the algorithm can be very time consuming, especially for short scales.

2.2. Computation of the proposed index

Let's define the reversed or mirrored time series as $X_r = \{x(N), x(N-1), \dots, x(1)\}$ so $x_r(n) = x(N-n+1)$ and the time series of the mirrored differences as:

$$d_m(n) = x(n) - x_r(n) = x(n) - x(N-n+1) \quad (3)$$

Then, the sign of $d_m(n)$ is computed. The proposed index (fscmd) is the number of sign changes normalized by $N-1$.

When applying the index to RR time series where a very slow component could be present (i.e. the night-day difference) it is advisable to detrend the time series prior to the computation of the index. In this work, all the RR time series have been detrended by using a bidirectional moving average filter with a window of 30 beats. The appendix contains the MATLAB code for the computation of the index (including the detrending procedure).

2.3. Simulations

In order to assess the agreement between α_1 and fscmd in purely self-similar time series, a fractional Gaussian noise generator proposed in [4] has been used with randomly generated Hurst exponents. The indices have been computed in time series with 30000, 3000 and 300 samples. For each length of the time series 100 realizations have been obtained.

2.4. Databases and preprocessing

For the sake of reproducibility we have used in this work only databases that are freely available online in www.physionet.org [5]. The Normal Sinus Rhythm RR time series database (NSR) has been employed to test the relationship between α_1 and fscmd in 24 hour recordings in healthy subjects. We have used all 54 recordings. The Fantasia database (FT) has also been employed in order to assess the agreement in a situation where all the subjects are in a resting situation. This database has a group of 20 young subjects and a group of 20 old subjects and the RR time series have a length of approximately two hours. Finally, the Congestive Heart Failure RR interval database (CHF) has been used in order to assess the agreement of both indices in a pathological situation and to compare with NSR. Previous to any analysis, the raw RR time series have been automatically corrected with the following algorithm:

If the change from one R-R interval to the next exceeds the inter-quartile range of the differentiated R-R time series ten-fold then it is decided that an artifact is present.

The artifact is then classified as a false positive (FP), false negative (FN) or ectopic beat (EB) artifact.

For a FP, as many consecutive R-R intervals as needed are added to obtain a corrected R-R beat whose value is close to the mean of the previous ten R-R intervals.

A FN is split in as many R-R intervals with equal value as needed to be close to the mean of the previous

ten R-R intervals.

An EB is substituted with two equal R-R intervals corresponding to the mean of the two R-R intervals involved in the EB.

After correction, α_1 and fscmd have been sequentially computed for each time series of the three databases in the same computer. In the case of the NSR, the time of computation for α_1 and fscmd has been also stored in order to compare the speed of processing.

2.5. Statistical Analysis

Statistical analysis were performed with SigmaStat© 3.10. The agreement of both indices was assessed with linear regression. The intercept, slope, R^2 and significance (p) of the analysis of variance of the linear regression were obtained for the simulated signals and the three databases. In the linear regression, α_1 was considered the dependent variable so it can be estimated by multiplying fscmd by the slope and then adding the intercept. Finally two Mann-Whitney Rank Sum tests were performed in order to assess if both α_1 and fscmd have significant differences when comparing the NSR and the CHF. The T statistic (sum of the ranks for the smaller group) and significance of the test (p) were used to compare the performance of both indices.

3. Results

Figures 1 and 2 show the results for the linear regression in simulated time series and RR time series respectively. Table 1 show the intercept and slope of the linear regression and the R^2 of the linear fitting. All the linear regression were very significant ($p < 0.001$).

Table 1. Results of the linear regression for the simulated time series and RR time series.

	Intercept	Slope	R^2
Simulated 30000 beats	1.593	-2.205	0.995
Simulated 3000 beats	1.577	-2.176	0.986
Simulated 300 beats	1.467	-1.841	0.834
NSR	1.657	-2.241	0.793
FT	1.605	-2.245	0.822
CHF	1.568	-1.923	0.937

Figure 3 show the results for the comparison of NSR and CHF using both indices using Mann-Whitney Rank Sum tests. Finally, the time expended in the computation of α_1 is 461.3 ± 44.5 (mean \pm SD) times that of the time needed to compute fscmd in the NSR.

4. Discussion

As seen from figure 1 and table 1 there is a significant relationship between α_1 and fscmd in self-similar time series especially when the length of the time series is large. Nevertheless, when using actual RR time series the relationship is still significant but at a lesser extent. Moreover, while the slope and intercept are similar in FT and NSR, the results for CHF show a lower intercept and slope. This can be due to the fact that actual RR time series can be approximated by a mix of several long-range correlation processes, quantification noise and pseudo periodic signals (i.e. respiratory sinus arrhythmia). In CHF, the mix seems to be different from that of FT or NSR. Figure 3 shows significant differences between NSR and CHF with both indices. Nevertheless, the sum of ranks (T) shows a higher significance when using fscmd instead of α_1 .

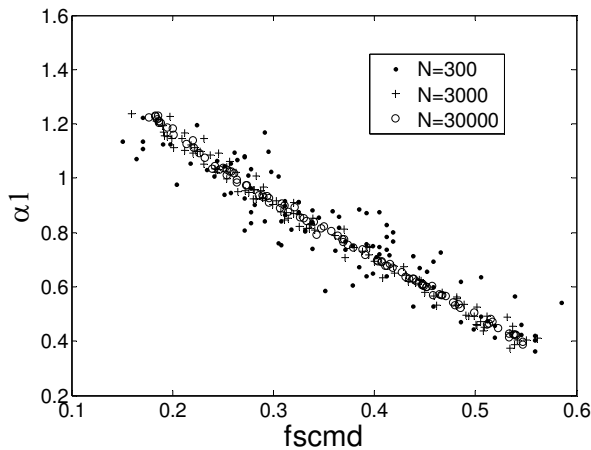


Figure 1. Comparison of indices for simulated self-similar time series for different time series lengths.

By definition, fscmd is computed from a time series that is generated as the point-to-point differences of the original time series and its reversed version. The sign of the differences could be completely different if the last samples of the original series are not considered or are missing. We have assessed the error by comparing fscmd obtained with the whole time series and the averaged fscmd obtained by computing the index when the last samples (discarding from one up to ten samples) are removed in the NSR. The relative error was $0.41\% \pm 0.38\%$ (mean \pm SD) so the effect of obtaining the index from a time series of length N instead of its subset of length N-k its quite negligible.

α_1 and fscmd are linearly related as seen in figure 1 and 2 so fscmd can be used as an alternative way to estimate α_1 . Nevertheless, table 1 shows that this linear relationship depends on the database. By joining the

results of NSR, FT and CHF the relationship between both indices can be written as:

$$\alpha_1 \Big|_{fscmd}^{predicted} = 1.594 - 2.068 \cdot fscmd \quad (4)$$

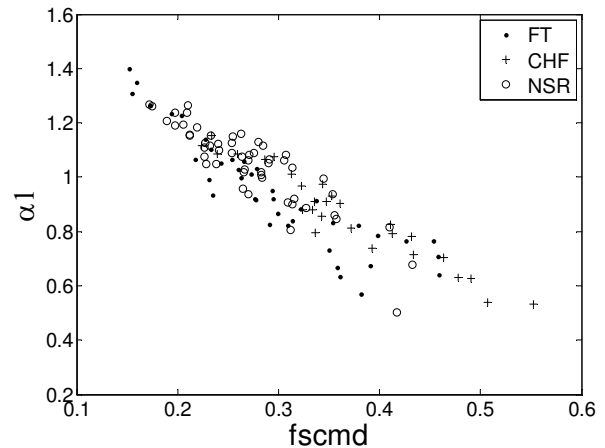


Figure 2. Comparison of indices for actual RR time series with different lengths. CHF and NSR correspond approximately to 24 hour recordings while FT uses approximately 2 hour recordings.

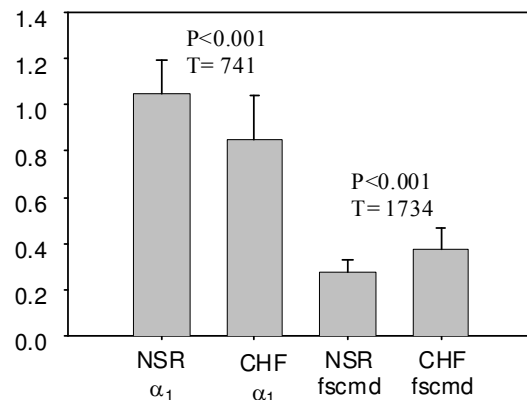


Figure 3. Mean and standard deviation of the indices for NSR and CHF and results of Mann-Whitney Rank Sum tests.

Figure 4 shows the Bland-Altman when comparing this estimation of the short range exponent and that obtained by DFA. The absolute difference between the estimations of the short range exponent is 0.058 ± 0.049 (mean \pm SD). So fscmd can be used as an alternative way to estimate α_1 . Nevertheless, further studies are required in order to assess if the relationship between both indices remains similar in other pathological or physiological situations.

5. Conclusions

The frequency of the sign changes of the mirrored differences (fscmd) is proposed as an alternative way to assess the self-affinity of RR time series in the short range. This index has a very significant linear regression with α_1 but it is computed nearly 500 times faster. When comparing RR time series from a normal group and from patients with congestive heart failure, there are very significant differences between both groups using fscmd or α_1 .

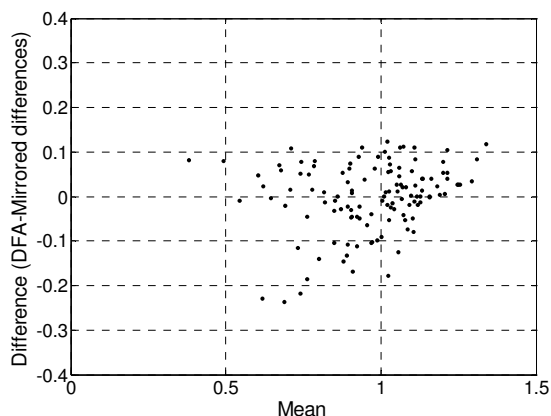


Figure 4. Bland-Altman diagram for the comparison of short range exponent estimated by DFA or by equation (4).

Acknowledgments

This work was supported by MEC project PSI2008-06417-C03

References

- [1] Peng CK, Havlin S, Stanley HE, Goldberger AL. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos* 1995; 5: 82-87.
- [2] Kleiger RE, Stein PK, Bigger JT. Heart rate variability: Measurement and clinical utility. *Annals of Noninvasive Electrocardiology* 2005; 10: 88-101.
- [3] Mielniczuk J, Wojdylo P. Estimation of Hurst exponent revisited. *Computational Statistics & Data Analysis* 2007; 51: 4510-4525.
- [4] Paxson V. Fast, approximate synthesis of fractional gaussian noise for generating self-similar network traffic. *Computer Communication Review* 1997; 27: 5-18.
- [5] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 2000; 101: e215-e220.

Appendix

The following function computes fscmd in MATLAB.

```
function fscmd=fscmdcomput(rr)
    rrd=rr-filtfilt(ones(size(1:30))/30,1,rr);
    rrr=rot90(rot90(rrd));
    d=sign(rrd-rrr);
    k=find(d==0);
    d(k)=-ones(size(k));
    fscmd=length(find(diff(d)))/(length(d)-1);
```

Address for correspondence.

PhD Miguel Ángel García González
 C/Jordi Girona 1-3, Edifici C-4, Campus Nord UPC
 08034 Barcelona, Spain
magarcia@eel.upc.edu