Assessing the Convolutedness of Multivariate Physiological Time Series

Paolo Castiglioni, Giampiero Merati, and Andrea Faini

Abstract— **A feature of time-series variability that may reveal underlying complex dynamics is the degree of "convolutedness". For multivariate series of *m* components, convolutedness can be defined as the propensity of the trail of the time-series samples to fill the *m*-dimensional space. This work proposes different convolutedness indices and compare them on synthesized and real physiological signals.**

The indices are based on length *L* and planar extension *d* of the trail in *m* dimensions. The classical ones are: the *L/d* ratio, and the Mandelbrot's fractal dimension (FD) of a curve: $FD_M = log(L)/log(d)$. In this work we also consider a correction of the Katz's estimator of FD_M, i.e., $FD_{KC} = log(N)/(log(N)+log(d/L))$, with *N* the number of samples; and FD_{MC}, an estimator of FD_M based on FD_{KC} calculated over a shorter running window $N_w<N$ appropriately selected to reduce estimation bias.

Synthesized fractional Brownian motions indicated that all the indices increase with FD, but differ for other aspects, namely the dependence on *N*; the capacity to estimate FD; or to distinguish between true bivariate and degenerate bivariate time series. Application on real multivariate recordings of muscular activity before and after exercise-induced fatigue suggests that these indices can be profitably used to identify complex changes in the dynamics of physiological signals.

I. Introduction

The "complexity" analysis of time series often requires the calculation of quantities that describe specific features of variability associated to their nonlinear dynamics, or to their fractal structure, or to their predictability. For instance, complexity analysis may include measures of entropy, that characterizes the level of irregularity or unpredictability of the time series; estimations of Lyapunov's exponents, that quantify the rate of separation of close trajectories; or calculation of scale coefficients, that describe how the time series looks self-similar over different temporal scales. A feature of time-series variability common to these aspects of complexity and that appears particularly useful for characterizing multivariate time series is the degree of "convolutedness". For a multivariate time series of *m* components, the convolutedness can be defined in words as the propensity of the trail of its samples to fill the *m*-dimensional space where the time series is plotted.

P. Castiglioni is with IRCCS Fondazione Don C. Gnocchi, Milan Italy (e-mail: pcastiglioni@dongnocchi.it).

G. Merati is with University of Milan, Milan, Italy Milan, Italy. A. Faini is with IRCCS, Istituto Auxologico Italiano, Milan, Italy.

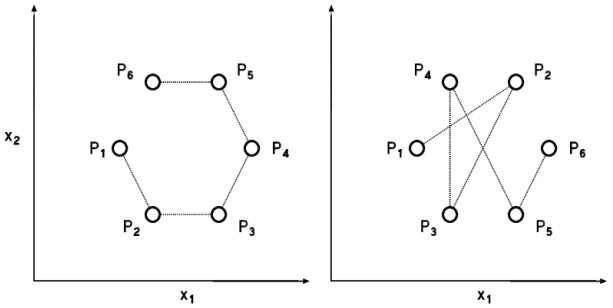

Figure 1. Two trails of bivariate time series with the same set of points but different degrees of convolutedness.

This work presents different indices for quantifying the degree of convolutedness of multivariate physiological time series. The first part illustrates the characteristics of these indices with synthesized signals. In particular, it shows the link with the concept of fractal dimension and highlights an error sometime committed in the quantification of the convolutedness of time series. The second part applies the indices on real multivariate recordings of muscular activity before and after a fatigue-inducing exercise. This is done to evaluate in a real physiological application the capability of convolutedness indices to detect changes in signal dynamics, as those possibly produced by fatigue on muscle activity.

II. Indices of Convolutedness

Let's consider *N* samples of a multivariate time series $\mathbf{X}(i)$ with *m* components, i.e., $\mathbf{X}(i)=[x_1(i),\ldots, x_m(i)]^T$ with $1\le i\le N$. Let's then construct the trail of $\mathbf{X}(i)$ in the *m*-dimensional Euclidean space by plotting the sequence of points P_i with coordinate $(x_1(i), \ldots, x_m(i))$. An index of convolutedness should describe how densely the trail covers the volume (or hypersurface in the *m*-dimensional space) that it occupies. It is important that any proposed index of convolutedness would not be defined only by the geometrical properties of the set of *N* points $\{P_i\}$, but that information on the temporal order of the data is maintained. For instance, an estimation of the geometrical fractal dimension of $\{P_i\}$ by a box-counting algorithm will provide the same "space filling" measure for identical sets of points, even if they are generated by time series that we want to characterize with different levels of convolutedness (see the example of figure 1). In this context, the relations between two parameters of the $\{P_i\}$ set appear important. One is the length, *L*, of the trajectory in the *m*-dimensional space, defined as:

$$L = \sum_{i=1}^{N-1} dist\left(P_i, P_{i+1}\right) \tag{1}$$

with $dist(P_i,P_j)$ the Euclidean distance between points *i* and *j*:

$$dist\left(P_i, P_j\right) = \sqrt{\sum_{k=1}^{m} \left(x_k(i) - x_k(j)\right)^2} \ . \qquad (2)$$

The second parameter is the planar extension of the trail, d, defined as the largest distance between any couple of points:

$$d = \max\left\{dist\left(P_i, P_j\right)\right\} \quad \forall i, j \ . \qquad (3)$$

The extension d is a measure of the volume occupied by the dataset $\{P_i\}$ while L is a geometric measure of length that preserves information on the temporal order in which the points are plotted. It should be underlined that the units of all the m components x_k should be the same for equations (1)-(3) to make sense. This rarely happens in real physiological applications. In these cases, each component x_k should be normalized before calculating L and d, for instance by dividing it by its standard deviation to get a dimensionless variable. For a monovariate time series $x(i)$, i.e. when m=1, the calculation of L and d is largely simplified because normalization is not needed, L is the sum of increments in absolute value, $L=\Sigma\ |x(i+1)-x(i)|$, and d is simply the range of the time series, d=max(x)-min(x).

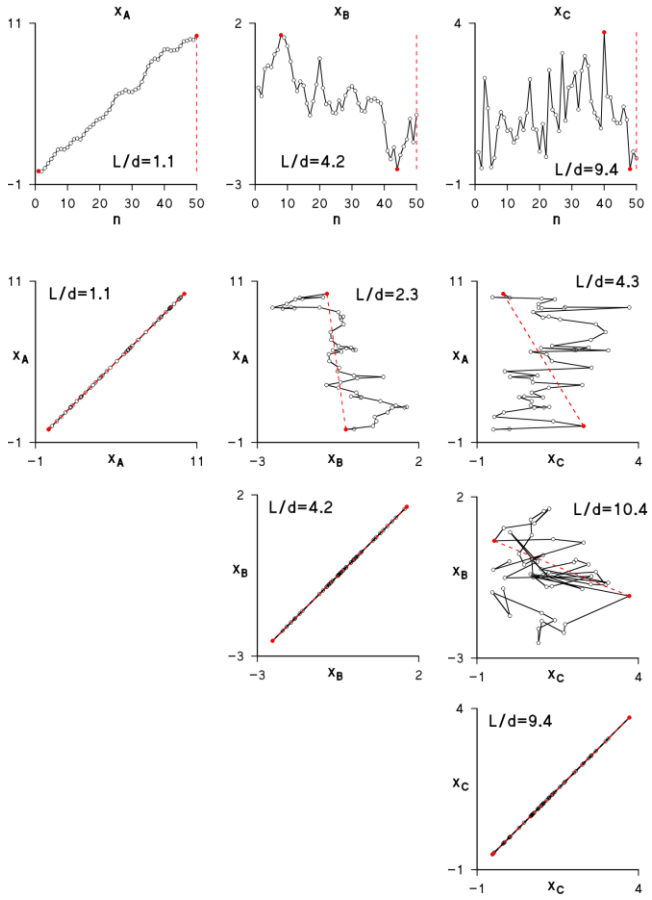

Figure 2. *Upper panels*: monovariate segments of fractional Brownian motions. *Mid and Lower panels*: Trails in the (x_i,x_j) space of bivariate time series derived by combining all couples of monovariate segments. Extension d (dashed red line) and L/d ratios are also shown.

A. Length/Extension Ratio, L/d

A simple index of convolutedness is the ratio between length and extension of the trails, L/d, a dimensionless number quantifying the "density" of the curve. Figure 2 shows an example of L/d calculation for three monovariate fractional Brownian motions, x_A, x_B and x_C, of N=50 samples with Hurst exponent H equal to 0.9 for x_A, 0.5 for x_B and 0.1 for x_C (the series were generated by the Matlab function wfbm(H,N)). Bivariate time series $\boldsymbol{X_{I,J}} = [x_I, x_J]^T$ were then obtained by combining all the possible couples of monovariate time series. For monovariate series, L/d increases when H decreases, as intuitively expected for increasingly irregular time series. An interesting case regards bivariate time series with identical components: $\boldsymbol{X_{A,A}}$, $\boldsymbol{X_{B,B}}$ and $\boldsymbol{X_{C,C}}$. They can be defined as "degenerate" bivariate series, because their trails in the bidimensional space can be seen as obtained from a monodimensional trail after a 45° rotation of the Cartesian reference. Actually, $\boldsymbol{X_{A,A}}$, $\boldsymbol{X_{B,B}}$ and $\boldsymbol{X_{C,C}}$, have trails that fills a monodimensional space even if plotted in a plane; moreover they have the same L/d ratio of their monovariate component (x_A, x_B and x_C) even if L and d of the bivariate series are larger.

B. Mandelbrot's winding of a curve, FD_M=log(L)/log(d)

To illustrate the concept of fractal dimension, Mandelbrot described the relation between the length of a winding river and the straight distance from its source to its mouth as: $length \propto distance^{FD}$ [1]. This leads to define $FD_M = log(L)/log(d)$ as index of convolutedness for any trail of multivariate time series in the m-dimensional space. Unlike the L/d index, FD_M requires normalization even for monovariate series to avoid that it depends on the units of measure.

C. Katz's proposal for the fractal dimension of waveforms

Mandelbrot's definition of fractal dimension of a curve was extended to monovariate time series by Katz [2]. Katz considered the graph of the time series $x(t)$ as a bidimensional curve in the (x,t) space, and the fractal dimension of this curve a measure of the time-series convolutedness. To make FD_M independent of the units of measure, L and d were divided by a, the average length of each "step" in the curve: $a=L/N$ [3]. This led to define the following index:

$$FD_K = \log(N) / \left(\log(N) + \log(d/L)\right). \qquad (4)$$

This measure of convolutedness is popular in biological sciences, but successive works showed that it suffers from important limitations in the estimation of the true fractal dimension FD of synthesized series [4,5]. These limitations make FD_K unable to properly describe the dynamics of physiological time series. Its poor performances were recently found to be due to a flaw in the definition of distance between points $P_i=(x(i), t_i)$ and $P_j=(x(j), t_j)$ that, in the Katz's method, is:

$$dist\left(P_i, P_j\right) = \sqrt{\left(x(i) - x(j)\right)^2 + \left(t_i - t_j\right)^2} \ . \qquad (5)$$

In fact, this equation erroneously sums together terms with different units [5].

D. Corrected Dimensions FD_{KC} and FD_{MC}

Katz's index in eq.(4) can be corrected by changing eq.(5) as $dist(P_i,P_j) = |x(i)-x(j)|$. It can be extended to multivariate series by calculating L and d according to eq.(1)-(3). Let's call FD_{KC} this correction of the FD_K index.

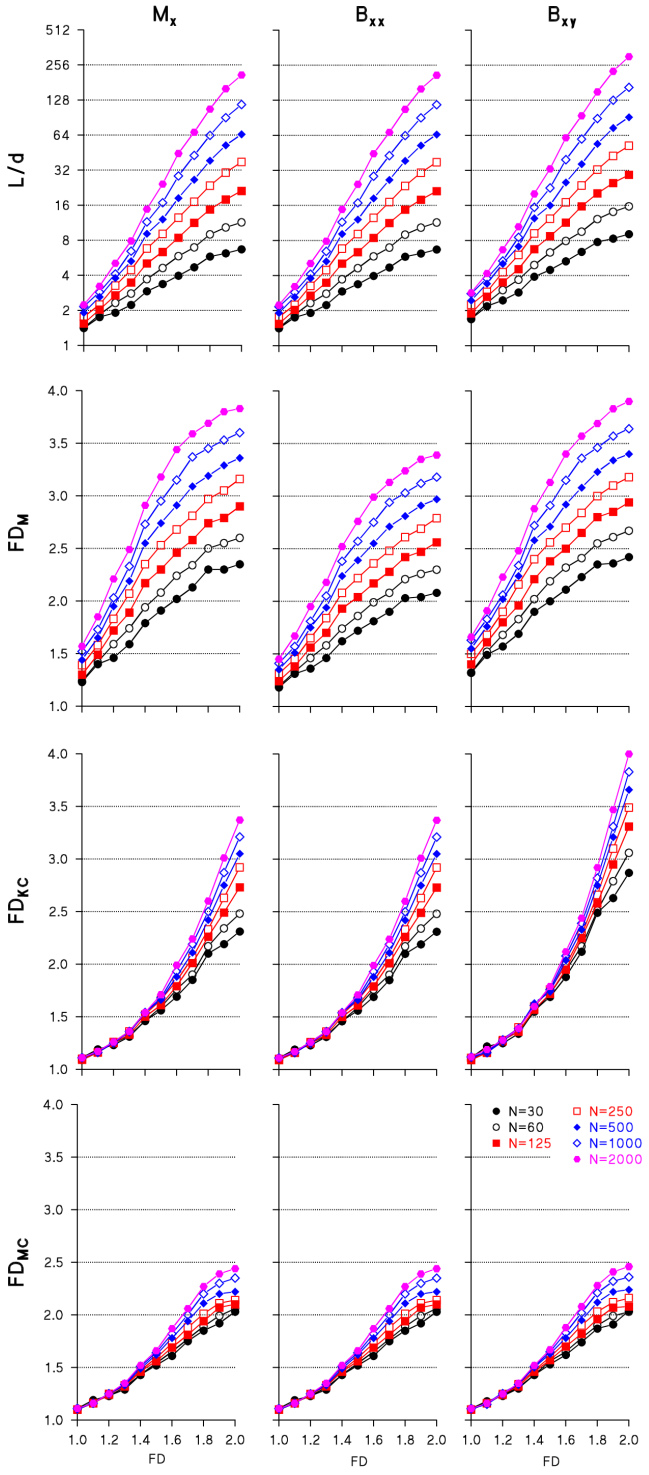

Figure 3. Convolutedness indices: means over 100 estimates for monovariate (M_x), degenerate bivariate (B_{xx}) and bivariate (B_{xy}) time series synthesized by fractional Brownian motions of length 30≤ N≤ 2000 and fractal dimension 1≤ FD≤ 2.

Simulations with deterministic and stochastic monovariate series showed that FD_{KC} correctly estimates the true FD when FD <1.5, but it tends to progressively overestimate FD when N increases and FD >1.5 [5]. To avoid overestimations, eq.(4) may be calculated over a running window of length $N_w<N$, with N_w the length of the data set with extension d equal to half the extension of the whole dataset [5]. Let's call FD_{MC} this further correction of the original Mandelbrot's FD_M.

III. Application on Fractional Brownian Motions

The performances of the presented indices of convolutedness (L/d, FD_M, FD_{KC} and FD_{MC}) were tested making use of synthesized fractional Brownian motions. Monovariate time series $M_x(i)=[x(i)]$, $1\le i\le N$, with N=30, 60, 125, 250, 500, 1000 and 2000 points were generated by the Matlab function wfbm(H,N). Eleven Hurst exponent H between 0 and 1, corresponding to FD=2-H, were considered. For each FD and for each size N, 100 monovariate series were generated. Then 100 bivariate series $\boldsymbol{B}_{xy}(i)=[x(i),y(i)]^T$ were generated for each FD combining couples of independent fractional Brownian motions synthesized with the same FD. Finally, 100 degenerate bivariate series $\boldsymbol{B}_{xx}(i)=[x(i),x(i)]^T$ were synthesized for each FD.

Results are shown in figure 3. All indices monotonically increased with the theoretic FD of the fractional Brownian motions generating the monovariate and bivariate time series. This is a desirable property because a larger FD is intuitively associated to a greater convolutedness. However, the indices also remarkably differed on other aspects of the quantification of convolutedness. Looking at the trails in the example of figure 2, one may expect that the convolutedness of a degenerate bivariate series is exactly the same of the corresponding monovariate series. This was true for all the proposed indices except FD_M, since $log(L)/log(d)$ was lower for B_{xx} compared to M_x. As to the relation between true bivariate series B_{xy} and degenerate bivariate series B_{xx}, the convolutedness was lower for degenerate series when quantified by L/d and FD_M, and for FD_{KC} when FD>1.5, but not for FD_{MC}. Regarding the use of these indices as FD estimators, FD_M diverged from FD when N increased. This occurred also for FD_{KC} but only when FD >1.5, while FD_{MC} was relatively close to the FD of synthesized series over the whole FD range even for large N.

IV. Application on Real Data

The proposed indices were calculated on real electromyogram (EMG) and mechanomyogram (MMG) signals. The aim was to evaluate whether convolutedness indices may indicate alterations in signal dynamics associated to muscle fatigue. It may be in fact hypothesized that muscle fatigue may change the dynamics of EMG and MMG signals during muscular contractions and that these fatigue-induced changes may be quantified differently by each index, thus allowing to separate different components of convolutedness.

EMG and MMG were recorded with electrodes and accelerometers placed on the biceps of the dominant arm in a healthy male volunteer during an isometric muscle contraction of the arm. The contraction force was maintained

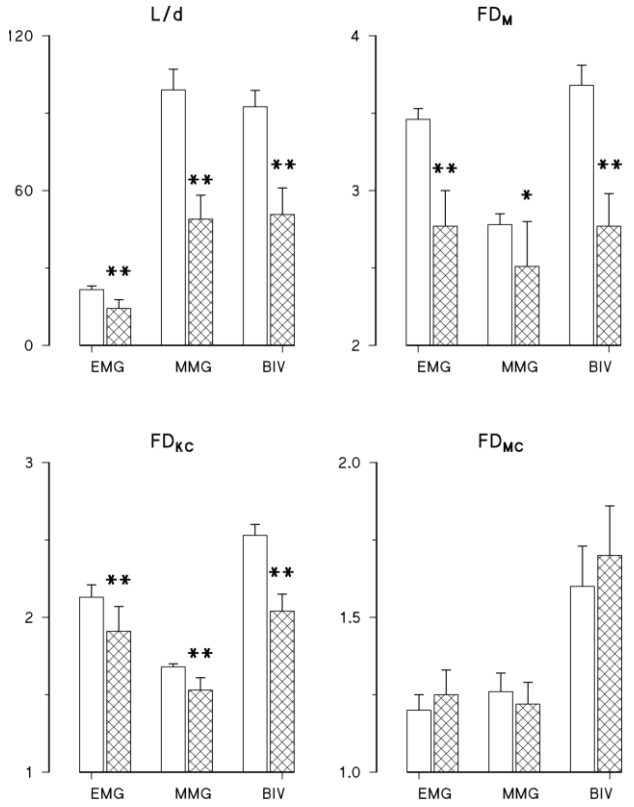

Figure 4. Indices of convolutedness (mean and standard deviation) in a volunteer before (open bar) and after (dashed bar) exercise, for mono- (EMG, MMG) and bi-variate (BIV) analysis; the * and ** indicate significant differences ($p<0.05$ and <0.01) between conditions.

close to 80% of the maximal volitional force measured immediately before the recording. Two recordings were performed, before and after an exercise protocol to induce muscle fatigue. Each recording lasted 25 s, with sampling rate of 2 KHz. A stable 6 s segment was selected at the end of each isometric contraction, before and after the fatigue-inducing exercise.

The indices of convolutedness were calculated over 6 consecutive, nonoverlapping windows of 1 s (N=2000 samples). The time series were normalized in each window by removing the mean and by dividing the time series by its standard deviation. The analysis was performed on monovariate EMG and MMG time series, and on the bivariate $[EMG, MMG]^T$ series. Mean and standard deviation of each index were calculated over the group of independent estimates in the 6 windows. The conditions before and after the fatigue-inducing exercise were statistically compared by Mann Whitney test.

After the fatigue inducing exercise the force produced during the stable isometric contraction decreased from 195 to 126 Newton. The associated changes in the convolutedness indices are shown in figure 4. All the indices decreased significantly after the fatigue inducing protocol, with the exception of FD_{MC} which did not vary substantially after exercise. This would suggests that the fatigue-inducing exercise preserves the fractal dimension of the time series while altering importantly other aspects of signals "irregularity".

V. Conclusion

The proposed indices are computationally simple and results on synthesized time series indicate that they can provide meaningful descriptions of convolutedness even for very short time series. This suggests their possible use also in real time monitoring of physiological signals. All indices are linked to the concept of fractal dimension, and simulations with fractional Brownian motions actually showed that all indices monotonically increases with the theoretic FD of the synthesized signals. However, among the proposed indices only FD_{MC} can be considered a reliable estimator of FD for physiological signals, whose FD values may widely range between 1 and 2. This would suggest that other features of time series variability may influence each of these indices. This is coherent with the general definition of "convolutedness", an aspect of time series variability in which different components of complexity (like signal entropy, deterministic chaos, self-similarity) may play a role.

Convolutedness analysis on real EMG and MMG series, although performed on a single volunteer, appears promising regarding its use as diagnostic tool for detecting alterations in the dynamics of physiological series. Moreover, the finding that exercise-induced fatigue did not alter the "purer" index of FD (i.e., FD_{MC}) but has dramatic effects on other convolutedness indices (like L/d, probably reflecting different complex components of variability) emphasizes the multifaceted nature of the dynamics of physiological time series.

References

[1] B. B. Mandelbrot, *The Fractal Geometry of Nature*, Ch.12, Freeman and Co., NewYork, 1983.

[2] M. J. Katz, "Fractals and the analysis of waveforms", *Comput. Biol. Med.* vol. 18, no. 3, pp. 145–156, 1988.

[3] M. J. Katz, E.B. George, "Fractals and the Analysis of Growth Paths", Bull. Math. Biol. vol.47, no.2, pp. 273-286, 1985.

[4] R. Esteller, G. Vachtsevanos, J. Echauz, B. Litt, A Comparison of Waveform Fractal Dimension Algorithms, *IEEE Trans. Circuits Syst.—I: Fundam. Theory Appl.* vol. 48, no. 2, pp. 177–183, 2001.

[5] P. Castiglioni, "What is wrong in Katz's method? Comments on: ''A note on fractal dimensions of biomedical waveforms'' *Comput Biol Med*., vol. 40, no.11-12, pp. 950-2, 2010.