CinC Challenge 2002 Undertaken by Non-stationary and Fractal Techniques

JC Echeverría[1,2], MS Woolfson[1], JA Crowe[1], BR Hayes-Gill[1]

[1]School of Electrical and Electronic Engineering, University of Nottingham, UK
[2]Electrical Engineering Department, Universidad Autónoma Metropolitana-Izt, México City, México

Abstract

By using a local recursive least squares approach, named the $\alpha\beta$ filter, to improve Detrended Fluctuation Analysis, we addressed the identification of real and synthetic Heart Rate Variability (HRV) data proposed as part of the Computers in Cardiology Challenge 2002. This approach provided power-law patterns that were used to correctly classify the majority (46/50) of the Challenge dataset. In addition, we show that this approach revealed patterns in real HRV data that do not necessarily follow a uniform power-law. Consequently, this paper promotes familiarisation with scaling information that has not been reported previously for healthy HRV data.

1. Introduction

Either by a linear spectral characterisation or by quantifying the non-linear properties, the analysis of Heart Rate Variability (HRV) can be seriously affected by the non-stationarity nature of these data [1]. Yet stationary conditions are difficult to achieve even in short-term records under physiologically stable or autonomic controlled circumstances. Accordingly, Detrended Fluctuation Analysis (DFA) [2] has been applied. This is a fractal method that has advantages because it permits the detection of long-range correlations embedded in non-stationary time series. DFA has confirmed, for healthy HRV data, long-range correlations over a wide range of time scales [3]. These results suggest that HRV data, lacking characteristic scales and having long-range correlation, reveal a healthy organising principle that seems to break down in several pathological states [3].

Essentially, the DFA explores the power-law relationship of the average root-mean-squared fluctuations as a function of time scales to provide the fractal exponent, or slope covering the short or long-term ranges. Rather than finding this fractal exponent to coarsely characterise those ranges, we have adopted a recursive least squares method, the $\alpha\beta$ filter, to recover the behaviour patterns in the power-law as a function of the time scales. We have used simulated and real data to evaluate this incorporation that we believe confers advantages to DFA by showing continuous variations in the power-law. Thus, these variations reveal more clearly abnormal physiological conditions.

It is in this context, that the second event of the Computers in Cardiology (CinC) Challenge 2002 [4] presented an opportunity for us to obtain a further evaluation of the DFA improvement by using the $\alpha\beta$ filter. Here, we present the results of addressing that Challenge by this approach. Briefly, the aim of the Challenge was the identification of either real or synthetic data from a set of 50 long-term records to promote the understanding of mechanisms underlying HRV as well as the exploration of novel analytic methods.

2. DFA improvement by $\alpha\beta$ filter

Basically, the DFA method [2] explores a power-law relationship after removing non-stationary trends from the original HRV series. Initially, the series is integrated and divided into windows, or boxes, of equal number of n intervals. In each window, the local trend is obtained by a least squares line fit. This trend is locally subtracted from the integrated series. The average root-mean-square fluctuation, $F_m(n)$, is then calculated.

The previous procedure is repeated for all window sizes or time scales n. Subsequently, the relationship on a double-log graph between these fluctuations $F_m(n)$ and time scales n can be explored. Normally, by assuming a linear model ($F_m(n) \sim n^{\zeta}$) the scaling, or fractal, exponent ζ is estimated by the slope of the log-log plot covering short or long-term range. Rather than finding this exponent only for predefined ranges, we have adopted the $\alpha\beta$ filter to recursively estimate a least-squares fitting for tracking the evolution of the gradient (*i.e.* of the power-law) as a function of log time scales. This filter has been used to characterise the operation of an induction motor or for tracking targets [5, 6]. A description of how the $\alpha\beta$ filter can be used to improve the DFA method as applied to HRV data (equations 1–5) is presented next.

Let $G_m(n)$ be the log of the average root-mean-squared fluctuations, $F_m(n)$, produced by DFA at the window size n, with n being the number of RR intervals (or the equivalent physiological time scale). Define $m_e(n)$ as the required estimate of the gradient at the log window size n.

Since a logarithmic representation of the window size n is employed, it is convenient to use an initial interpolation of $G_m(n)$ so that it is parameterised by a uniformly sampled variable in the logarithmic representation of the window size domain; using k to denote the discrete elements or samples of this new variable. Then, it is possible to predict the value of the root-mean-squared fluctuations at k according to:

$$G_p(k) = G_e(k-1) + m_e(k-1)\delta \quad (1)$$

where $G_p(k)$ is the predicted log root-mean-squared fluctuations, $G_e(k-1)$ is the estimate of log root-mean-squared fluctuations at $k-1$, and δ is the uniform separation between successive elements of the parametrized variable. The original piece of information, *i.e.* the log root-mean-squared fluctuations $G_m(k)$, produced by the DFA method at k, is combined with the previous prediction $G_p(k)$ according to:

$$G_e(k) = [1-\alpha(k)]G_p(k) + \alpha(k)G_m(k) \quad (2)$$

in order to obtain the estimate of the log root-mean-squared fluctuations at k. Here, $\alpha(k)$ is a smoothing coefficient at k. Finally, the desired estimate of the gradient at k, $m_e(k)$, can be obtained from:

$$m_e(k) = m_e(k-1) + \frac{\beta(k)}{\delta}[G_m(k) - G_p(k)] \quad (3)$$

where $\beta(k)$ is the second smoothing coefficient of this approach. The smoothing coefficients $\alpha(k)$ and $\beta(k)$ are given by:

$$\alpha(k) = \frac{2(2k-1)}{k(k+1)} \quad (4)$$

and

$$\beta(k) = \frac{6}{k(k+1)} \quad (5)$$

To take into account any possible deviation from linearity in the log domain, these coefficients are prevented from going to zero for $k > Q$ (we have found the value $Q = 500$ as an appropriate balance to avoid either the gradient not being adequately tracked or the noise artefacts becoming significant).

3. Data to be analysed. CinC challenge 2002

As part of the second event of the PhysioNet/CinC 2002 a dataset consisting of 50 time series of RR intervals was posted to promote the classification of this set into real and synthetic data [4].

The only information provided for this dataset was that approximately half of the series were obtained from long-term ambulatory ECG recordings of subjects between the ages of the 20 and 50 who have no known cardiac abnormality. The remaining part of the dataset corresponded to synthetic HRV data (produced using generators submitted by participants in the first event of the challenge as well as by generators provided by the organisers). Each series contained between 20 and 24 hours of RR intervals and real data may also have included isolated ectopic beats.

4. Classification criteria

We used the improved DFA analysis (*i.e* DFA plus $\alpha\beta$) to obtain the power-law pattern for each series of the Challenge dataset. Prior to the application of this technique the pre-processing procedure suggested in [7] was used to remove unqualified sinus beats. The resulting power-law patterns were studied under the following assumptions to classify them as either real or synthetic.

For real HRV data, scaling exponents, or slopes, near to unity in the long-range have been reported [2] which support the hypothesis that healthy HRV presents a type of *1/f* behaviour or uniform power law. By contrast, for short-range, the scaling exponents are larger than unity. These have been attributed to the smooth fluctuations that are associated with respiration [2]. In addition, we have studied power-law patterns from healthy HRV data by using the improved DFA described above. The resulting patterns are generally in accordance with those observations, showing deviations above unity in short-range and on occasions following a uniform power-law, or constant gradient, in long-range. However, since this approach provides more information it is also possible to find deviations from a uniform power-law that cannot be quantified or detected by a single exponent.

5. Results

Using the above criteria it was possible to identify 20 entries on the dataset that did not present normal power-law behaviour in either short or long-range so providing an initial identification of synthetic data. Furthermore, by comparing their power-law patterns it was possible to unambiguously associate these entries in pairs. Hence, these results were in accordance with the Challenge specifications describing that each generator of synthetic data was used to produce two series.

As an example of this synthetic group, Figure 1 presents entries 02 (left) and 29 (right) of the Challenge dataset. In the top part of the figure are included the original series whilst in the bottom are depicted the corresponding

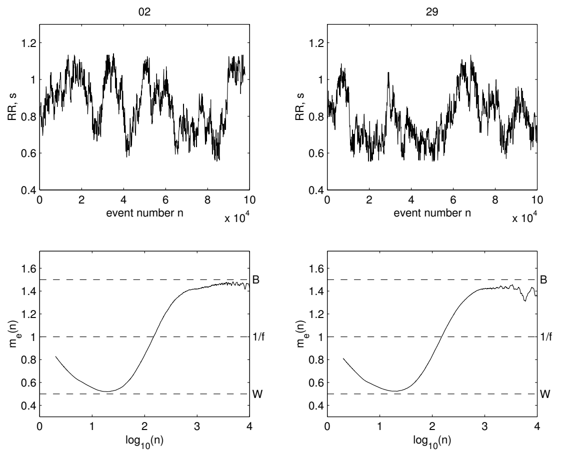

Figure 1. Power-law patterns of synthetic HRV data

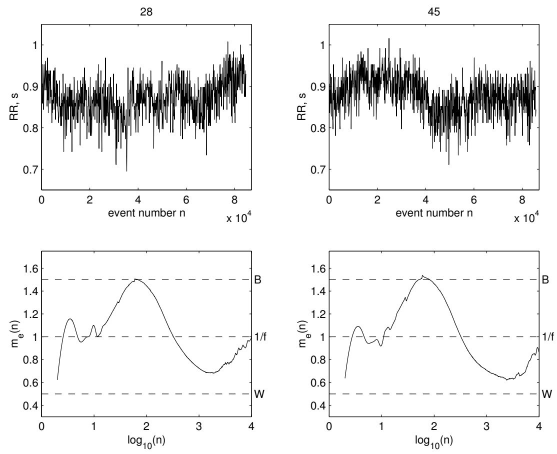

Figure 2. Power-law patterns of synthetic HRV data

power-law patterns. Also shown as dashed lines are the theoretically expected values for white noise (0.5), 1/f noise (1), and Brownian noise (1.5). Interestingly, it is not possible to relate these series by only studying their temporal evolution; however, the power-law patterns clearly reveal the same statistical behaviour for both series. In the long-range it is possible to appreciate that these series involve Brownian like behaviour whilst for the short-range a white noise appearance can be postulated. These two conditions, not showing either smooth oscillations in the short-range or persistent long-term correlations, are unrealistic characteristics of healthy HRV data.

Figure 2 presents another pair of presumed synthetic data, entries 28 (left) and 45 (right). Here again, it is possible to appreciate different temporal evolutions yet almost identical power-law patterns. These patterns do not present long-range correlations or clear deviations above unity in short-range. Additionally, it is also possible to appreciate a strong Brownian behaviour for the intermediate range and some notches that can be associated with the existence of perfectly regular oscillations. These again are dubious characteristics of real HRV data.

By contrast, Figure 3 presents examples of six power-law patterns that were considered real because they present long-range plateaus near unity (i.e. they follow a uniform power-law) and they show clear deviations above one in short-range. This means that in these patterns it is possible to appreciate the involvement of smooth oscillations perhaps associated with short-term autonomic responses and the underlying existence of persistent long-term correlations. These patterns were obtained from entries 06, 08, 24, 41, 47 and 50.

In addition to the 20 initial entries classified as synthetic, one more pair having very similar power-law patterns was also found. For this reason, these patterns were classified as synthetic in spite of presenting more realistic features.

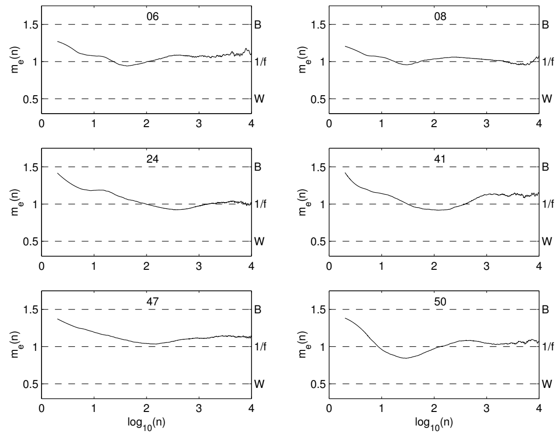

Figure 3. Power-law patterns of real HRV data

As a result, 22 entries were selected as synthetic whilst the remaining 28 apparently showing normal behaviour were identified as real. Using this classification we achieved a Challenge score of 88 points indicating that the majority of the dataset (46/50) was correctly classified. Yet four entries still remain to be properly re-classified. Figure 4 presents two possible candidates. These power-law patterns, obtained from entries 10 and 37, were initially classified as real due to existence of long-range correlations. However, owing to the intrinsic deviations of DFA for small window sizes, these kind of patterns that asymptotically approach [8] a uniform power-law and do not show major deviations in short-range, could also be produced by data generated using conventional *1/f*$^{\gamma}$ noise synthesis methods [9].

Finally to promote a further familiarisation with the power-law patterns that can be found in healthy HRV data, Figure 5 presents six additional results obtained from entries 01, 07, 13, 18, 21 and 40 of the Challenge dataset. Since strictly similar pairs were not possible to be found

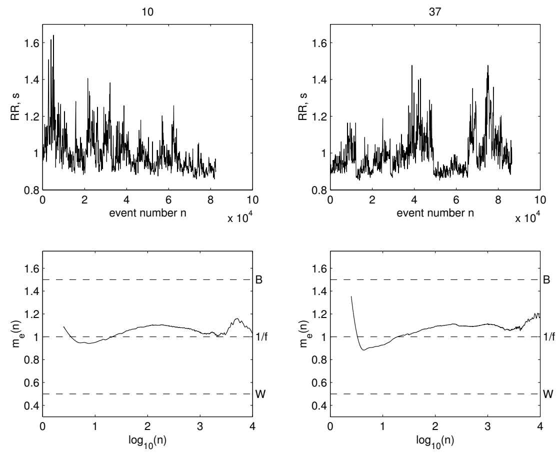

Figure 4. Possible candidates for re-classification, *1/f*$^\gamma$ noise ?

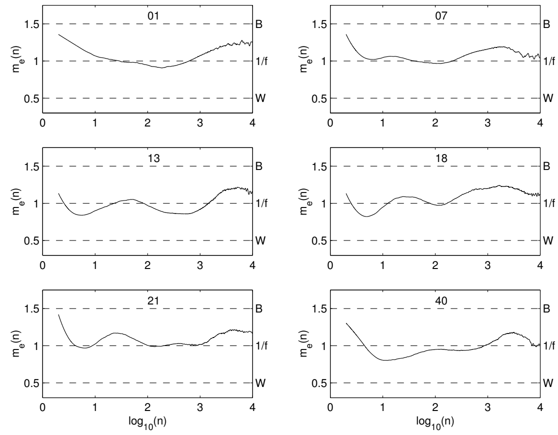

Figure 5. Power-law patterns of presumed real HRV data

for these patterns, we assume that they belong to real data. Interestingly, three patterns (13,18 and 40) seem to lack deviations associated with smooth oscillations in short-range and all patterns involve long-range deviations from a uniform power-law. These deviations could be explained by the existence of characteristic scales in the data. Consequently, HRV data obtained from subjects with no cardiac abnormalities do not always appear to present a uniform power-law as has been suggested before [3]. Furthermore, it is reasonable to speculate that either these kind of deviations are not necessary abnormal or that these subjects presented sub-clinical or undetected manifestations.

6. Conclusions

By studying only power-law patterns it was possible to classify correctly the majority of the CinC Challenge dataset and to gain an insight into the statistical properties of the generators used to create synthetic data. These patterns were obtained as a result of improving DFA by a local recursive least squares method named the $\alpha\beta$ filter. In addition, it was shown that this improvement reveals patterns in real data that do not necessarily follow a uniform power-law. Consequently, this paper promotes familiarisation with scaling information that has not been reported previously for healthy HRV data.

Acknowledgements

The contribution of the Mexican Council for Science and Technology (CONACyT) for the financial support of JC Echeverria is gratefully acknowledged.

References

[1] European, society of cardiology, the North American society of pacing. Heart rate variability, standards of measurements, physiological interpretation, and clinical use. European Heart Journal 1996;17:354.
[2] Peng CK, Havlin S, Stanley HE, Goldberger AL. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos 1995;5(1):82.
[3] Walleczek J (ed.). Self-organized biological dynamics and nonlinear control. Cambridge, UK: Cambridge University Press, 2000; 66.
[4] Goldberger A, Amaral LAN, Glass L, Hausdorff JM, Ivanov P, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. Physiobank, physiotoolkit and physionet. Circulation 2000; 101:e215.
[5] Woolfson M, Ferrah A, Bradley KJ, Asher GM. A novel method to track the frequency of a sinusoid embedded in sinusoids of known frequencies using recursive phase compensation. Applied Signal Processing 1997;4:212.
[6] Munu M, Harrison I, Woolfson M. Comparison of the kalman, $\alpha\beta$ and $\alpha\beta\gamma$ filters for the tracking of targets using radar with variable update time. Applied Signal Processing 1995;2:24.
[7] Ho K, Moody GB, Peng CK, Mietus JE, Larson MG, Levy D, Goldberger AL. Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics. Circulation 1997;96:842.
[8] Kantelhardt J, Koscielny-Bunde E, Rego HHA, Havlin S, Bunde A. Detecting long-range correlations with detrended fluctuation analysis. Physica A 2001;295:441.
[9] Heinz-Otto P, Saupe D (eds.). The science of fractal images. New York, USA: Springer-Verlag, 1988; 71.

Address for correspondence:

JC Echeverria
UAM-I, Ingenieria Electrica
San Rafael Atlixco 186, 09340, Mexico DF, Mexico
jcea@xanum.uam.mx