

Trend Strips: a New Tool to Analyze RR Time Series

Antonio Carlos da Silva Filho¹, Fatima M H S P Silva², Julio Cesar Crescencio³,
Lourenço Gallo Junior³

¹Centro Universitário de Franca, Uni-FACEF, Franca, Brazil

²Department of Physics and Chemistry, School of Pharmaceutical Sciences of Ribeirao Preto,
University of Sao Paulo, Ribeirao Preto, Brazil

³Division of Cardiology, Medical School of Ribeirao Preto, University of Sao Paulo, Ribeirao Preto,
Brazil

Abstract

There are many different tools and parameter that can be used to analyze RR time series, as the Correlation Dimension, the Lyapunov Exponents, the Higher Reconstruction Step. Here we address the problem of discriminating between two populations: one with some cardiac disease and another with healthy people, in a non invasive way, through RR time series. The mathematical tool used to analyze the series is called Trend Strips (TS) and was proposed recently in the scientific literature. TS are a sequence of "0" and "1", where "1" means that the value in some position is greater than the value in a previous position in the series, while "0" means the opposite. We can use TS of any size, where the size is the amount of "0" and "1" considered. If we choose TS of size 4 we will have 16 different sequences at our disposal, from "1111" to "0000". What we do, then, is to compute the relative frequency of occurrence of each of them in a particular time series and to compare the frequencies in both groups.

We collected RR time series from two groups of men at rest: one group with a cardiac chagasic disease (24 individuals) and a second one of healthy people (21 individuals). We analyzed TS of sizes from 2 to 5 and found significant TS that discriminate the two groups at all sizes. The statistical test used was a t-test and some of the TS useful, with a $p < 0.01$ were: (a) for $n=2$, TS "00" ($p=0.0009$); (b) $n=3$, TS "101" ($p=0.0055$) and TS "000" ($p=0.0061$); (c) for $n=4$, TS "1011" ($p=0.0059$), TS "1000" ($p=0.0071$) and TS "0001" ($p=0.0073$), etc. The set of significant TS can be useful, than, for diagnostic purposes.

1. Introduction

If a system in nature can give us a sequence of data

obtained at different times, we can try do analyze the behavior, the state and the dynamics of it analyzing the data. Time series analysis is, then, a key tool to study the dynamics and the characteristics of the systems.

We can use it, for instance, to characterize the system or to forecast some variable. This is particularly important in cardiology, where we have a noninvasive way to study an individual: the ECG recordings. Through the ECG recordings we can get, for instance, RR time series. We can analyze these series in the time domain, in the frequency domain and for a lot of dynamical variables that arise in the study of nonlinear dynamical systems. Our group has published many papers studying some nonlinear dynamics variables [1-6] in RR time series. In this paper, we are analyzing the RR time series for two groups of individuals, one (control group) of healthy individuals and another group where each individual has a chagasic cardiac disease.

The tool we used to perform this analysis was recently proposed in the literature and used to study some financial time series [7]. The objects exhibited in that paper are called "Trend Strips" (TS). In the following we will introduce this tool.

Let $X = \{x_1, x_2, \dots, x_n\}$ be a time series with a discrete index t (usually the time) that spans an interval from 1 to n . Starting from this series X we can generate another time series (an auxiliary series, that we will call Y), with $n-1$ elements. In this auxiliary series each element is either 0 or 1, which are, in this model, just nominal variables. In the general case, we will have "s" states – all nominal variables – for each position, meaning that we will have "s" states attributed to every data in the system. The case with just two states is a binary one.

The meaning of "1" is that the value of x at some position has increased (or remained the same) over the value of x at a previous position, while "0" means that the value has decreased:

$$y_{t+1} = \begin{cases} 1 & \text{if } x_{t+1} - x_t \geq 0 \\ 0 & \text{if } x_{t+1} - x_t < 0 \end{cases}$$

The Y series has, then, $n-1$ elements because we have constructed it starting from the second position in the X series, where the value of X is x_2 . Defining this way, the Y series is a sequence of “0” e “1”, as can be viewed in Table 1, which is just a small piece of the entire X and Y series.

Table 1. Binary Trend Strips (TS) with lengths from 1 to 3. $x(i)$ stands for the original time series, while $y(i)$ stands for the auxiliary time series. The code for $TSa[b]c$ is: a = length; b = number of states (2 for binary states) and c = position in the original time series.

i	1	2	3	4	5
$x(i)$	37	45	59	44	68
$y(i)$	-	1	1	0	1
$TS1[2](i)$	-	1	1	0	1
$TS2[2](i)$	-	-	11	10	01
$TS3[2](i)$	-	-	-	110	101

From the Y series, we extract the last values and form a strip with them. Here we have to fix a new variable: the length of the strips. Let us say we have chosen this length to be five. This means we have a small strip with five cells and each cell is filled with a “0” or with a “1”. These strips form a kind of symbolic dynamics and were named “Trend Strips” (TS) in [7].

If we have, for instance, a binary strip with six cells we can have $2^6 = 64$ different trend strips. So, from the first six values of Y series we extract the first TS; from the second till the seventh value we extract the second TS and so on. In the example of Table 1 this first TS of length 3 is “110”. We, then, look at the next value in the Y series and hold it. The next step is to construct a new TS: we erase the oldest (the first) value (a “1” in this example) of the first TS, displace the remaining values one cell to the left and add a new one (another “1”, in this case) to the now empty cell at the right of the TS. At every new TS we look at the next value in the Y series and hold it. The other TS are constructed in a similar way. We can see all the TS of this type we can form for the example in Table 1.

The present paper looks for the existence of some patterns in the Y series formed from RR time series obtained from two groups of individuals: one healthy and the other with chagasic disease.

2. Methods and tools

This work was approved by the Research Ethics

Committee of Hospital of the Faculty of Medicine of Ribeirao Preto, University of Sao Paulo, Brazil.

The RR time series were collected from two groups of men: (i) one group with a cardiac chagasic disease (24 individuals) and (ii) a second one of healthy people (21 individuals). All the series, for each individual, were collected in two positions: supine and after passive 70 degree head-up tilting, with volunteers seated in a saddle (dismissing the initial transient changes). In order to prevent spurious influences (changes in their standards due to biological rhythms) all the series were collected at the same period of the day. The variable recorded were the RR intervals (ms), each one lasting for a little more than fifteen minutes. Specific software was used to detect R waves of ECG signals and the respective periods [8-9]; the RR interval were then obtained.

At first, a conventional statistical analysis was done regarding the variability of the RR intervals. Then we computed the frequency (in Matlab code) of the TS for lengths ranging from 2 to 5. The following step was to compare the frequencies of each TS between the two groups. This was done using a t-test.

3. Results

The frequencies of the TS are not the same, even for a random time series. For instance, if we form a time series with 1,000,000 data points from a uniform random number generator (as RAND, found in Matlab), we will have the spectrum displayed in Figure 1 for the case $n = 4$ (there are 16 TS), where base stands for the number of states (state one = 1; state two = 0) and n stands for the number of states in each trend strip.

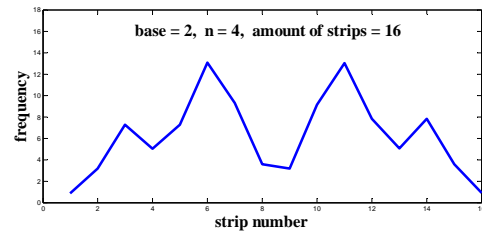


Figure 1. Frequency of occurrence of each one of the 16 trend strips (a binary TS with six positions) for a random time series with 1,000,000 data.

The frequencies of TS for one member of the healthy group can be observed in the dotted line of Figure 2, while the frequencies for one member of the group with chagasic disease (both in supine position) can be observed in the continuous line of the same figure.

As we can see in this figure, the frequencies of each of the 16 TS are different for the two individuals and both frequencies are different from the frequencies for the random case displayed in Figure 1. What we do in this paper is to compare the frequencies between the two

groups through a statistical analysis, in order to see if there are statistically significant differences between them. If it happens, we can use the TS as a noninvasive test to discriminate the two populations. As we shall see, this goal was accomplished.

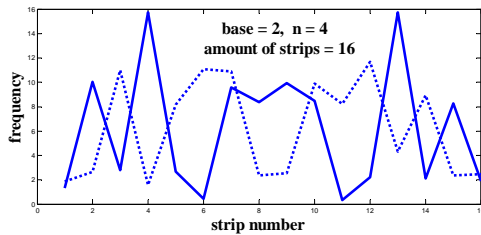


Figure 2. Frequency of occurrence of each one of the 16 TS for: (i) continuous line: one of the individuals with the disease and (2) dotted line: one of the healthy individuals (both cases in supine position).

The results of the t-tests are displayed in Table 2 and Table 3.

Table 2. Comparing the frequencies of each TS (n=1 to n=5) for the two populations in seated position using a t-test (1=significant). Sig is the significance level.

TS\Sig	0.05	0.01	0.001	0.0001
1	1	0	0	0
0	1	0	0	0
11	0	0	0	0
10	1	1	1	0
01	1	1	1	0
00	1	1	1	0
111	1	0	0	0
110	0	0	0	0
101	1	1	1	0
100	0	0	0	0
011	0	0	0	0
010	1	1	0	0
001	0	0	0	0
000	1	1	1	0
1111	0	0	0	0
1110	1	0	0	0
1101	1	1	0	0
1100	1	0	0	0
1011	1	1	0	0
1010	1	0	0	0
1001	1	1	0	0
1000	1	1	1	0
0111	1	0	0	0
0110	1	0	0	0
0101	1	0	0	0
0100	1	0	0	0
0011	1	0	0	0
0010	1	1	0	0

0001	1	1	1	0
0000	1	1	1	0
11000	1	1	1	1
00011	1	1	1	1

Table 3. Comparing the frequencies of each TS (n=1 to n=4) for the two populations in supine position using a t-test (1=significant). Sig is the significance level.

TS\Sig	0.05	0.01	0.001	0.0001
1	1	0	0	0
0	1	0	0	0
11	0	0	0	0
10	1	0	0	0
01	1	0	0	0
00	1	1	1	0
111	0	0	0	0
110	0	0	0	0
101	1	1	0	0
100	0	0	0	0
011	0	0	0	0
010	0	0	0	0
001	0	0	0	0
000	1	1	0	0
1111	0	0	0	0
1110	0	0	0	0
1101	1	0	0	0
1100	0	0	0	0
1011	1	1	0	0
1010	0	0	0	0
1001	0	0	0	0
1000	1	1	0	0
0111	0	0	0	0
0110	0	0	0	0
0101	1	0	0	0
0100	0	0	0	0
0011	1	0	0	0
0010	0	0	0	0
0001	1	1	0	0
0000	1	0	0	0

We have 62 TS for n in the range (1-4). At the significance level of 5% we got 43 TS that show differences between the two groups in seated position (21 in supine position), at 1% we got 22 significant TS (9 in supine position) and we got 16 significant TS (1 in supine position) at 0.1% significance level. As we increase the significance level, a fewer number of TS show this difference, but we still got some (in seated position).

The best results were obtained for significance level of 0.01% with two TS: the “11000” and “00011” in seated position. In the supine position, the best results were obtained for the TS “00”, with a significance level of 0.1%.

4. Conclusions

We got a lot of significant TS. This means that some TS can be used (with a significance as low as 0.01%) to distinguish the two populations. The TS “11000” and “00011” are the best examples. In fact, we have a whole set of significant TS to do the job.

A next step will be to use the TS to distinguish not only populations but individuals. If successful, the TS may be used in future as a noninvasive aid in clinical diagnosis.

References

- [1] Silva FMHSP et al. Identification of anaerobic threshold during dynamic exercise in healthy man using Kolmogorov-Sinai entropy. *Computers in Cardiology* 2005; 32:731-734.
- [2] Silva Filho AC et al. Using the lag of autocorrelation function in order to identify the anaerobic threshold during dynamic physical exercise. *Computers in Cardiology* 2006; 33: 625-628.
- [3] Silva Filho AC et al. An artificial neural network as a tool to identify the anaerobic threshold during dynamic physical exercise. *Computers in Cardiology* 2007; 34:597-600.
- [4] Silva FMHSP et al. Modifications in the heart dynamics of patients with cardiac disease. *Computers in Cardiology* 2009; 36:421-424.
- [5] Silva Filho AC et al. A new parameter in the nonlinear dynamics of the heart: the higher reconstruction step. *Computers in Cardiology* 2010; 37:959-962.
- [6] Silva FMHSP et al. Applying Lyapunov Exponents in heart rate time series to identify the anaerobic threshold in healthy men. *Computing in Cardiology* 2012; 39:629-632.
- [7] Silva Filho AC, Souza RM, Lima FG. Trend Strips: A new tool to analyze financial markets. *Journal of Academy of Business and Economics* 2012, 12:49-54.
- [8] Silva E et al. Design of a computerized system to evaluate the cardiac function during dynamic exercise. *Physics in Medicine & Biology* 1994; 33: 409.
- [9] Gallo Jr L et al. Control of heart rate during exercise in health and disease. *Brazilian Journal of Medicine and Biological Research* 1995; 28: 1179-84.

Address for correspondence:

Antônio Carlos da Silva Filho
Centro Universitário de Franca – Uni-FACEF
Av. Major Nicácio 2433
Franca – SP – Brasil
CEP 14401-135
E-mail: acdasf@bol.com.br