British Journal of Applied Science & Technology
18(1): 1-16, 2016; Article no.BJAST.29596
ISSN: 2231-0843, NLM ID: 101664541

SCIENCEDOMAIN *international*
www.sciencedomain.org

Size-Related Properties of Area1 of Approximate Entropy to Characterize Time-series Organization

José Eduardo Soubhia Natali[1], Paulo Nogueira Starzynski[1], Ingird Machado Cusin Ahmed El-Dash[1], Thiago Paes de Barros de Luccia[1], Vivian Machado Cusin Ahmed El-Dash[1] and José Guilherme Chaui-Berlinck[1*]

[1]Department of Physiology, Biosciences Institute, University of São Paulo, Rua do Matão, 101. CEP: 05508-090, Brazil.

Authors' contributions

This work was carried out in collaboration between all authors. Authors JESN and JGCB designed the study, performed the simulations and statistical analysis. Authors PNS, IMCAED, TPBL and VMCAED carried out the ECG data collection and analysis. Author JGCB wrote the first draft of the manuscript and managed literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJAST/2016/29596
Editor(s):
(1) Vyacheslav O. Vakhnenko, Division of Geodynamics of Explosion, Subbotin Institute of Geophysics, National Academy of Sciences of Ukrainian, Ukraine.
Reviewers:
(1) Thomas L. Toulias, Technological Educational Institute of Athens, Greece.
(2) S. B. Ota, Institute of Physics, Sachivalaya Marg, Bhubaneswar, India.
Complete Peer review History: http://www.sciencedomain.org/review-history/16856

Original Research Article

Received 19th September 2016
Accepted 29th October 2016
Published 10th November 2016

ABSTRACT

Aims: There are several entropy estimators to address the organization of time-series. However, the behavior of a given estimator in relation to the size *N* of the data is not often studied in terms of improving the analysis. Here, we investigate size-related properties of the estimator a1ApEn (area1 of approximate entropy) in order to establish how such properties can improve time-series analysis.
Study Design/Methodology: We established a set of 14 different generating processes, including deterministic maps and limited and unlimited random distributions. Then, we created several vectors of five different sizes (*N* = 100, 200, 400, 500, 1000) for each process, and a set of indicators (*maximum, minimum* and *mean* a1ApEn values) was taken. The correlation between a given indicator and $\log_{10}(N)$ was classified as greater or lower than zero, or non-significant, creating a pattern of correlations for each process. Next, we perform a similar analysis in a resampling procedure from vectors of 2,000 points for the same generating processes. In addition, we analyzed heart rate dynamics and solar wind cycles with this method in order to show the applicability of the technique.

**Corresponding author: E-mail: jgcb.fisio.teor@gmail.com, jgcb@usp.br;*

Results: The main result is that the patterns of the correlations between indicators and $\log_{10}(N)$ are able to segregate the different generating process.
Conclusion: The use of a resampling procedure along with the size-related correlations of the nonlinear estimator a1ApEn is an effective method to discern different generating processes underlying empirical time-series. The method allows for the use of data sets of different sizes in comparisons among results.

Keywords: Informational entropy; time-series analysis; data-vector size; approximate entropy; heart rate; solar wind.

1. INTRODUCTION

The discursive concept of complexity has no unique mathematical counterpart. Usually, the diversity of the elements of a time-series (or data set) is addressed in the efforts to give a number to complexity, and how to measure such a diversity comprises many different approaches. However, even what would be measured in the analysis might vary dramatically (e.g., [1]).

Within the "how to measure" question above, there are numberless informational entropies with different properties and different scopes each one of them. Important examples are the Rényi entropy, S_R, and Tsallis entropy, S_q, as discussed by Masi [2]. These entropies are devoted to extend Shannon's uncertainty measure to a broader range of cases.

From a different perspective, we have information entropies or measures alike (e.g., false neighborhood [3], directed weighted complex network [4]) that seek to establish a direct approach to data to give an estimation of the degree of organization of the series. In this sense, such measures are much more "statistical" estimators than entropies, indeed.

Approximate Entropy [5] falls into the category of the so-called "practical estimators" we just created above. It is interesting to note that Pincus envisaged a measure that would attain a limiting value as N tends to infinity [5,6] (see, also [7]), and employed the notation ApEn(m,r,N) to indicate a value coming from a finite sample size.

Nevertheless, what is meant by extensive can have diverse interpretations. Formally, a function of extensive variables is extensive if it is homogeneous of degree 1 ([8] - chapter 5). However, such a discussion diverges from what the measures of organization are truly intended for. On a rather practical way, organization is somehow related to the occupancy of a phase-space created by the estimation procedure itself [3,4], and "extensive" should allude for the relationship between the value of the estimator and the size N of the data set. In fact, size-related issues are of great importance for estimators in general, either linear or nonlinear ones (consider, for instance, the mean and the variance), since one needs to know the contingent dependence of the estimator on the size of the data under analysis.

As a direct example, Bruijn et al. [9] tested two nonlinear methods to quantify certain patterns of walking stability and found that while one of the methods presented an increase in its value as the length of the time-series increased, the other one presented a decrease. These results lead the authors to conclude that, in order to make comparisons meaningful, "a fixed number of strides should be analyzed" [9]. Thus, it becomes implicit that size of the sample might become an obstacle to compare different data.

Even though its general acceptance, ApEn suffers from two subtle, and interconnected, drawbacks (see [10–14]). The first one is the lack an objective procedure to choose the parameters of analysis (namely, the tolerance for differences, r, and the size of the probing window, m). The other is the lack of consistency: two time-series can be classified in opposite ways depending on the choice of the parameters made by the observer.

Recently, our group envisaged a new method, derived from ApEn, which we proved to be of greater consistency and completely objective [14]. The procedure is based on the construction of the area under the curve of ApEn versus a normalized tolerance vector **r** for the probing window size m = 1. The tool was named area1-ApEn, or, simply, a1ApEn (see [14] for details of the method).

Therefore, it is very important to know the size-related properties of a1ApEn. Moreover, it would be extremely relevant if these properties could be employed for a more complete characterization of empirical time-series without the obstacle of sample size. In the present study, we explore these properties and show that a1ApEn has size-related features useful to recognize different processes.

2. METHODS

2.1 An Overview of the a1ApEn Estimator

Notice that the complete description of the procedures is found elsewhere [14]. Here we simply give an outline of the a1ApEn estimator.

2.1.1 Approximate entropy

Within a time-series of size *N*, ApEn is devised as a counting of sub-vectors equal to each other along the series. The distance between a pair of sub-vectors is given by the Heavside distance, and equality is considered within a certain tolerance.

Eqs 1-3 below detail the numerical procedure. A sub-vector **i** of size *m* is compared to every other sub-vector of size *m* along the original time-series. For each paring, a match is generated if the distance is equal or less than a given tolerance *r*. The counting *C* for a sub-vector **i** is the number of matches (#) weighted in relation to the total number of comparisons made. Thus, for a given sub-vector **i**:

$$C_{\mathbf{i}}^{m}(r) = \frac{\#_{\mathbf{i}}^{m}}{N - m + 1} \tag{1}$$

For the whole set of comparisons, ϕ is a variable computed as:

$$\phi^{m}(r) = \frac{1}{N-m+1} \cdot \sum_{1}^{N-m+1} \ln\left(C_{\mathbf{i}}^{m}\right) \tag{2}$$

Notice that the denominator in Eqs 1 and 2 is the same for all the sub-vectors **i**, and that for each **i** there is, at least, one positive match coming from the comparison of **i** and **i** itself.

Finally, the ApEn estimator is obtained through:

$$\mathrm{ApEn}(m, r, N) = \phi^{m}(r) - \phi^{m+1}(r) \tag{3}$$

2.1.2 The estimator a1ApEn

The first and fundamental step is the construction of a tolerance vector, which, for the sake of simplicity, we shall call **r**. This vector contains multiple values of tolerances.

In ordinary ApEn computations, the usual practice to establish the single value of tolerance for comparisons is to calculate it as a fraction of the standard deviation of the time-series (see [15]). For instance, some figure between 15% and 25% of the standard deviation is the most typical choice for ApEn computations.

Here, we do not use the standard deviation as a milestone to compute *r* values to compose the tolerance **r** vector. Instead, we obtain a detailed set of distances among the members of the time-series (code given in Supporting Information at the end of the manuscript). This set goes from the lowest value of distance that can be found among all the possible pairings in the time-series to the highest value of distance that can be formed among pairs. Then, from these distances, a new tolerance vector $\mathbf{r}^{*}$ is constructed by scaling the original **r** from zero to one:

$$\mathbf{r}^{*} = \frac{1}{\max(r)} \cdot \mathbf{r} \tag{4}$$

The next step is to construct a curve of ApEn values using the original **r** vector for a window size *m* = 1. Such a curve describes the behavior of the function:

$$\mathrm{ApEn}(1, \mathbf{r}, N) = \phi^{1}(\mathbf{r}) - \phi^{2}(\mathbf{r}) \tag{5}$$

The last step is to compute the numeric integral of Eq 5 in relation to the normalized tolerance vector $\mathbf{r}^{*}$:

$$\mathrm{a1ApEn}(N) = \int_{0}^{1} \mathrm{ApEn}(1, \mathbf{r}^{*}, N)\, dr^{*} \tag{6}$$

2.2 Scaling of a1ApEn

The main sketch of the present work is as follows. In a first step, we studied fourteen prototypic processes generated *in machina* and addressed the question of what happens to the a1ApEn measure as the size of the vectors increases.

In a second step, from a single time-series of 2,000 points of each prototypic process, we repeatedly resampled the series, thus forming vectors of diverse sizes. Next, we checked whether the size-related properties detected in the first step are still the same. The answer here is yes.

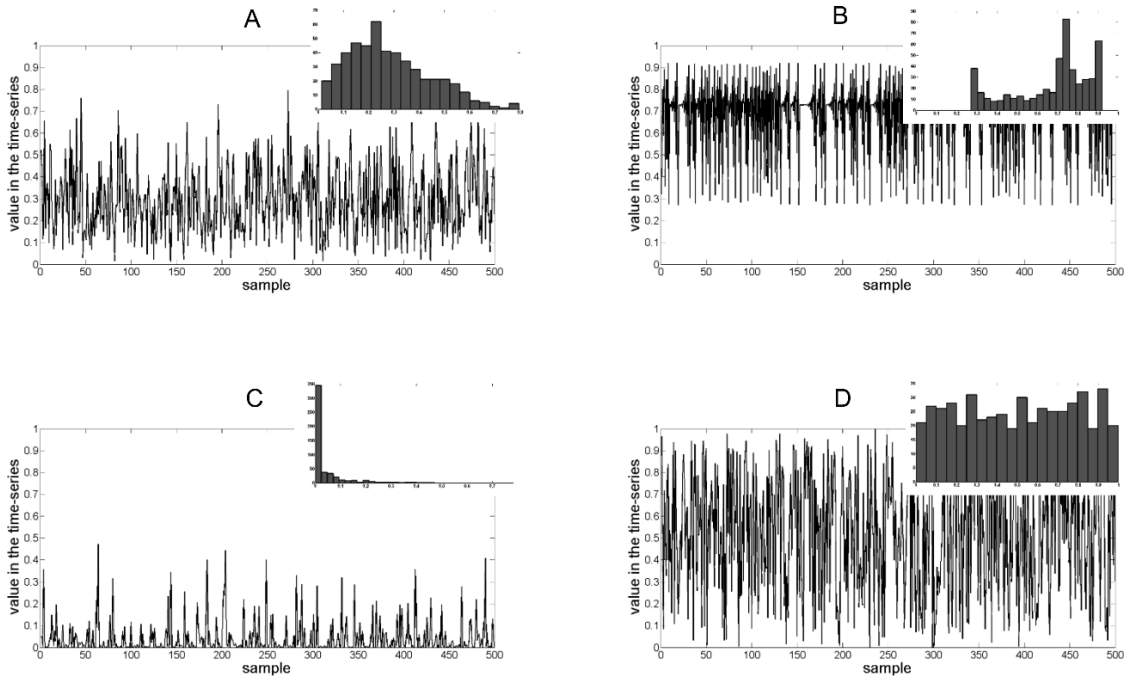

Fig. 1. Some examples of prototypic time-series

These series are contained in the]0,1[interval and their respective histograms are shown in the insets; N = 500. (A) Beta2; (B) L3.68; (C) Beta4; (D) UDRN. See Table 1 for details of the originating processes

Table 1. Prototypic processes and their designations (in alphabetical order)

Designation	Process/Description
AR1	Autoregressive 1st order. The weight of the previous datum in the next ndrn datum was set as 20% (see NDRN entry below).
AR2	Autoregressive 2nd order. The weight of the previous and the second previous data in the next ndrn datum were 20% and 40%, respectively (see NDRN entry below).
Beta1	Beta random distribution ($\alpha = \beta = 0.4$); U-shaped.
Beta2	Beta random distribution ($\alpha = 2$, $\beta = 5$); left-skewed.
Beta3	Beta random distribution ($\alpha = \beta = 0.1$); extremely U-shaped.
Beta4	Beta random distribution ($\alpha = 0.2$, $\beta = 5$); extremely left-skewed.
Henon	Henon map. Parameters a = 1.4 and b = 0.3. Chaos region. Initial conditions were randomly assigned around $x_0 \cong 0.63$ and $y_0 \cong 0.179$
L3.6	Logistic map with control parameter μ = 3.6. Period 2 region, jumping between]0.30 , 0.61[to]0.78 , 0.91[. Initial condition randomly assigned.
L3.68	Logistic map with control parameter μ = 3.68. Pomeau–Manneville scenario (transition to chaos due to intermittency). Initial condition randomly assigned.
L3.9	Logistic map with control parameter μ = 3.9. Chaos region. Initial condition randomly assigned.
L3.99	Logistic map with control parameter μ = 3.99. Chaos region. Initial condition randomly assigned.
Lévy	Lévy distribution.
NDRN	Normally distributed random numbers with zero mean and standard deviation of 1.
UDRN	Uniformly distributed random numbers in the]0,1[interval.

The third and final step was to apply the resampling method to empirical data (resting heart rate and solar wind) to illustrate the usefulness of the size-related properties of a1ApEn.

Table 1 specifies and describes the prototypic processes studied here. Notice that are three grand subsets of processes: (1) deterministic maps (Hénon and Logistic); (2) random numbers limited to a given interval (beta and uniform distributions); (3) random numbers not limited (autoregressive models, Lévy and normal distributions). Fig. 1 illustrates examples of time-series generated by some of these processes. Data were generated *in machina* using Matlab R2013a.

3. RESULTS AND DISCUSSION

3.1 First Step – Independent Vectors

Vectors of varying sizes for each process were generated. The sizes were N = 100; 200; 400; 500; 1,000. For each size, 60 independent vectors were created. Thus, 300 vectors (5x60) were generated per process, and 4,200 vectors were analyzed (14x300).

From the 60 vectors of each size (in each process), we established, as indicators, the *maximum value* (upper bound), the *mean value* and the *minimum value* (lower bound) of the a1ApEn estimator. The correlation between $\log_{10}(N)$ and each indicator was obtained for each process. The use of the logarithm of the size is for linearize the order of magnitude of change in N, thus avoiding false correlations due to the extremity effects in the statistics.

The results are summarized in Table 2. As the table shows, depending on the process and the indicator chosen, there are positive, "zero" or negative correlations, where "zero" is for the non-significant correlations. Fig. 2 illustrates different behaviors of these indicators, for the same processes shown in Fig. 1.

Therefore, depending on the process and on the indicator, a1ApEn can be an increasing, a decreasing or an unchanging measure as a function of the time-series size (as exemplified in Fig. 2).

Table 3 gives a symbolic pattern of correlations for the results presented in Table 2. There, we grouped the processes accordingly to their symbolic pattern. It is interesting to note that the size-related properties of a1ApEn are able to discern these originating processes and group them into those three grand subsets that we know beforehand. Notice, thus, that we are not concerned with the values of the correlations: our focus is in patterns.

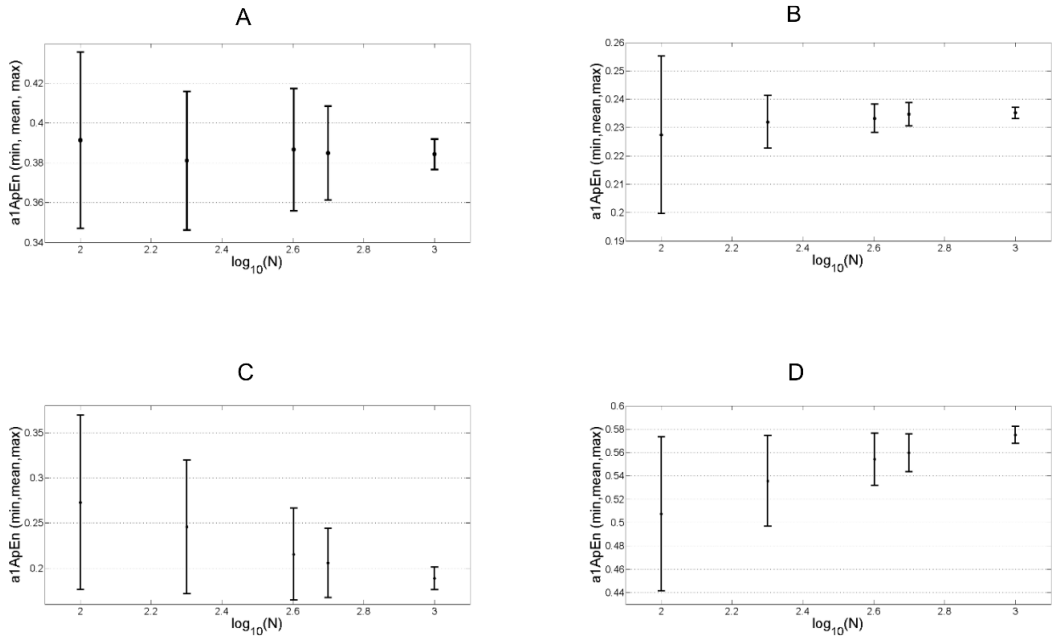

Fig. 2. Upper bound, mean and lower bound of a1ApEn for some prototypic series
Results from the resampling procedure (step two) versus $\log_{10}(N)$. (A) Beta2; (B) L3.68; (C) Beta4; (D) UDRN

Table 2. Correlation between log(*N*) and a1ApEn values for the indicators (*maximum, mean* and *minimum*)

Process		*Maximum*	*Mean*	*Minimum*
AR1	r	-0.70	0.55	0.95
	F-value	2.89	1.3	**29.77**
AR2	r	-0.90	-0.89	0.00
	F-value	***12.44***	***11.56***	0.01
Beta1	r	-0.95	-0.97	0.00
	F-value	***27.3***	***67.44***	0.01
Beta2	r	0.96	0.98	0.95
	F-value	**34.74**	**79.6**	**30.81**
Beta3	r	-0.33	-0.61	0.87
	F-value	0.38	1.8	8.93
Beta4	r	0.96	0.96	0.89
	F-value	**41**	**34.23**	**12.16**
Henon	r	-0.99	-0.99	0.70
	F-value	***143***	***194***	2.87
L3.6	r	0.87	0.99	0.99
	F-value	9.68	**551**	**304**
L3.68	r	0.77	0.91	0.94
	F-value	4.42	**14.07**	**24.09**
L3.9	r	0.00	0.96	0.91
	F-value	0.03	**45.24**	**14**
L3.99	r	0.00	0.96	0.97
	F-value	0.04	**41.5**	**52.6**
Lévy	r	0.40	0.92	0.91
	F-value	0.58	**15.72**	**13.41**
NDRN	r	-0.99	-0.98	-0.77
	F-value	***617***	***86.73***	4.58
UDRN	r	-0.91	-0.96	0.42
	F-value	***16.16***	***37.37***	0.65

r: correlation coefficient. In bold: significant positive correlation; italic-bold: significant negative correlation. F critical = 10.13 for 95% confidence interval ($v_1 = 1$, $v_2 = 3$)

Table 3. Symbolic representation of the correlations of the prototypic processes, organized by pattern similarity

Process	*Maximum*	*Mean*	*Minimum*
Beta2	0	0	0
Henon	0	+	+
L3.6	0	+	+
L3.68	0	+	+
L3.9	0	+	+
L3.99	0	+	+
AR1	-	-	0
AR2	-	-	0
Beta4	-	-	0
Lévy	-	-	0
NDRN	-	-	0
Beta1	+	+	+
Beta3	+	+	+
UDRN	+	+	+

+: significant positive correlation; -: significant negative correlation; 0: non-significant correlation

An exception seems to be the Beta2 random distribution (parameters $\alpha = 2$, $\beta = 5$), that ended up in an isolated pattern. This distribution is somewhat skewed, but not too much. If it "progresses" towards a more prominent skewness (as Beta4), it will resemble a non-limited process, and its size-related pattern would go as the one of the normally distributed random numbers or the Lévy distribution (which is an extreme of an unlimited process). On the other hand, if it "progresses" towards a less skewed distribution, its size-related pattern would, then, go as those of the clearly limited processes, as the uniformly distributed random numbers. Therefore, the size-related properties of our Beta2 distribution help to explain the size-related patterns of other distributions, as we present below.

The set formed by AR1, AR2, Beta4, Lévy and NDRN has negative correlations of the maximum and of the mean a1ApEn values, while minimum values have non-significant correlations. The members of this set comprise distributions that have the major part of their data within some limits but, every now and then, a discrepant value emerges. Then, from the standpoint of a1ApEn, the phase-space seems less occupied when very discrepant values are found in the time-series (as it were a less dense set). As the

size of the series increases the chance of discrepant values to emerge also increases, thus the mean and the maximum tend to decrease with an increasing *N*. On the other hand, minimum a1ApEn values are not affected by this condition (thus, the zero correlation for the minimum values).

The opposite holds true for the set comprising Beta1, Beta3 and UDRN. These processes are limited and, as their sizes increase, the phase-space is gradually more occupied, resulting in a1ApEn values progressively higher.

3.2 Second Step – Resampling Procedure

The results just described indicate that the size-related properties of a1ApEn are, likely, a relevant topic to address the temporal organization of a time-series and its putative underlying process. However, for one to obtain 60 independent vectors with different sizes as we did in the preceding section is highly improbable in the real world (132,000 data points for each process were necessary for those analysis).

Therefore, we adopt another approach, one that could be feasible for empirical data analysis. From a single series of 2,000 points for each process, we conducted a resampling procedure. The resampling procedure consisted of selecting an initial point at random and, then, extract a vector of *N*-1 consecutive points from that initial one. This was done 30 times for each size *N* (100, 200, 400, 500 and 1,000), and so, 1,500 random sub-vectors of different sizes were created from the 2,000-points original one.

Minimum, mean and *maximum* values of a1ApEn for the 30 vectors of each size were obtained and we computed the correlation with log10(*N*), as in the previous section. Table 4 shows the symbolic pattern for this resampling procedure.

Basically, the same groupings are obtained. A noticeable change occurred for the autoregressive model of 1st order, that, now, resembles the Beta2 pattern from the 60 independent samples.

Anyway, it is still clear that one can use the size-related properties of a1ApEn in a resampling procedure to distinguish among different processes. Maps and limited random processes contrast each other by the size-related behavior of maximum values, and unlimited random processes present a clear diverse pattern, essentially due to negative correlations.

Table 4. Symbolic representation of the correlation of the resampled prototypic processes for the indicators (maximum, mean and minimum)

Process	*Maximum*	*Mean*	*Minimum*
Beta2	0	0	+
Henon	0	+	+
L3.6	0	+	+
L3.68	0	+	+
L3.9	0	+	+
L3.99	0	+	+
AR1	0	0	0
AR2	0	-	0
Beta4	-	-	0
Lévy	-	-	0
NDRN	-	-	0
Beta1	+	+	+
Beta3	+	+	+
UDRN	+	+	+

+: significant positive; -: significant negative; 0: non-significant correlation

3.3 Third Step – Empirical Data Analysis

In order to test the real application of the size-related properties of the a1ApEn measure, we investigated two types of data: heart rate dynamics and solar wind.

3.3.1 Heart rate dynamics

Electrocardiographic records (ECG) were obtained from seven resting individuals for 30 minutes using superficial electrodes. Table 5 gives a descriptive summary of the volunteers. From each 30 minutes recording, the first five were discarded at once. The analysis comprises the remaining 25 minutes of data from each subject.

Table 5. Volunteers' descriptive summary

Subject	Gender	Age (years)	Mean HR (bpm)	Size of the <R-R> vector
1	F	38	54	1,322
2	F	22	69	1,749
3	M	32	55	1,289
4	M	32	52	1,288
5	M	23	56	1,372
6	M	36	74	1,990
7	F	22	76	1,967

From the ECG recording of a given individual, a vector of sequential inter-beats intervals (R-R interval) was obtained (the total size of the

vectors changes among the individuals due to different heart rates) – Fig. 3 illustrates one of these series. Then, the same resampling procedure adopted in step two was undertaken. The symbolic patterns obtained are shown in Table 6, and we grouped the individuals accordingly to these patterns.

We can recognize that four individuals have similar patterns, corresponding to the group of the unlimited random processes as those found in Table 4. Two individuals (#2 and #7) are identical twin sisters that were subjected to corrective heart surgery in their early childhood and they have a low-grade mitral valve prolapse. Interestingly, the size-related properties of a1ApEn have grouped them separately from the other volunteers. Finally, individual #5 falls in another subset that do not correspond to any pattern of the processes studied here. This subject has no phenotypic particularity that could explain his results. Nevertheless, it is relevant to note that the a1ApEn size-related pattern of heart rate dynamics generally resembles the unlimited random processes, and seems to detect some particularities of individuals.

3.3.2 Solar wind

Data of the magnetic field and the plasma temperature from solar wind were obtained directly from NASA (omniweb.gsfc.nasa.gov/form/dx1.html). Few points were deleted due to an apparent saturation of the magnetic field sensors (measurements above 999.9 nT). Each time-series was partitioned in three vectors related to the solar spots cycle: the 23rd increasing phase (from May 01 1996 to December 31 2002); the 23rd decreasing phase (from January 01 2003 to January 03 2008); and the 24th increasing phase (from January 04 2008 to February 28 2014). Fig. 4 shows the data and Table 7 presents the results of the analysis as symbolic patterns. There are few cycles to analyze, so the results are more to appreciate an application of the tool and not for a full evaluation of the phenomenon itself.

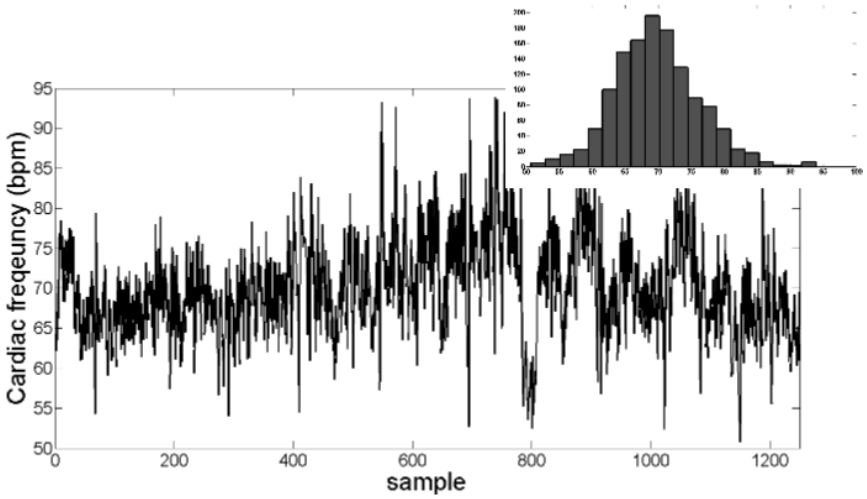

Fig. 3. Example of a heart rate series
Cardiac frequency in bpm, data from subject 4. Histogram of data distribution in the inset

Table 6. Symbolic representation of the correlation between vector size and a1ApEn indicators for the <R-R> data of the seven experimental individuals

Subject	Maximum	Mean	Minimum	Process alike
1	-	-	0	
3	-	-	0	Beta4/Lévy
4	-	-	0	
6	-	-	0	
5	-	-	+	none
2	0	0	0	AR1 / Beta2
7	-	0	0	

Symbols are the same as in Tables 3 and 4

Table 7. Symbolic representation of the correlation between vector size and a1ApEn indicators for the solar wind data. MF: magnetic field; PT: plasma temperature

Variable	Cycle	Maximum	Mean	Minimum	Process alike
MF	23rd increasing	-	-	0	Beta4/Lévy
	23rd decreasing	0	-	-	none
	24th increasing	-	-	0	Beta4/Lévy
	23rd increasing	-	-	+	none
PT	23rd decreasing	-	-	0	Beta4/Lévy
	24th increasing	-	-	0	Beta4/Lévy

Symbols are the same as in Tables 3 and 4

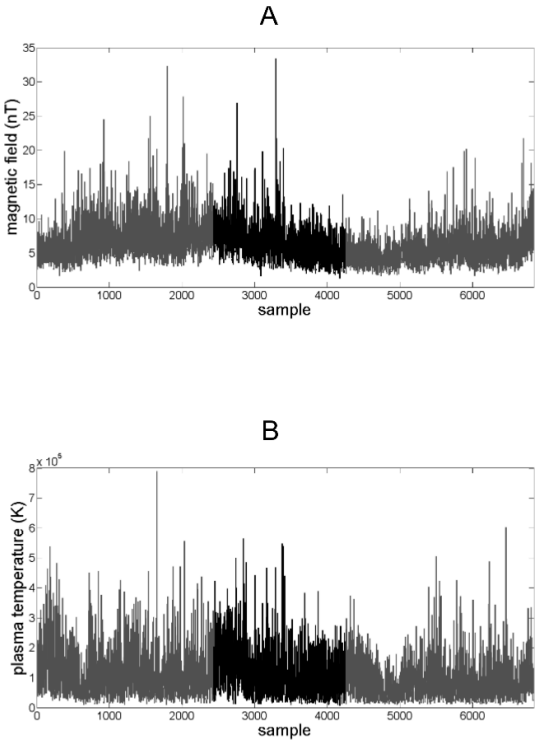

Fig. 4. Solar wind data. The whole series was partitioned in three sequential sub-series (grey-black-grey) accordingly to the cycle and phase

(A) Magnetic field [nT]; (B) Plasma temperature [K]. Data from NASA (http://omniweb.gsfc.nasa.gov/form/dx1.html).

To ascribe a degree of organization for a given time-series is an ever-challenging task. In some specific cases, it is quite simple, indeed. However, in a vast majority, there seems to be a great number of ways to give a quantitative estimation to the behavior of the series, but these estimators are not necessarily concordant one each other. In addition, a single value of a measure may be misleading for a more deep identification of primary processes that forge the data under analysis. In this sense, the present study shows the importance of the size-related properties of the estimator a1ApEn to a most comprehensive characterization of the organization of time-series.

Information entropy estimators have no a priori rule for their scaling properties. Shannon's uncertainty measure is extensive in relation to the size of the "alphabet" originating the message, but not in relation to the size of the data [16]. Approximate Entropies are expected to grow logarithmically with size, at least from Rukhin analysis [7]. From a different perspective, Costa et al. constructed vectors of decreasing sizes and increasing grainy from an original one, and employed Sample Entropy [17], from the ApEn family, as a measure to address heart rate variability and other biological processes [18,19]. However, due to the concomitant change in size and grainy, it is not possible to discern scaling properties of the measure in these reports.

Here we studied how three values (*minimum, mean* and *maximum*) of a1ApEn are related to the size of the series under analysis. We call these values as "indicators". We employed time-series created *in machina* from some specific generating processes and obtained the correlations of the indicators in relation to the logarithm of the size.

The correlation can be positive, negative or zero (i.e., non-significant) for each indicator. This creates a pattern and we show that different generating processes present different patterns of correlations. Moreover, for similar generating processes, the emerging pattern is the same.

Therefore, independently of a single plain value of the estimator, and independently of the values of the correlations, the patterns identify the processes. In other words, we use patterns of correlations to identify sets of processes and plain values to locate the level of organization of the data within a given set. To our knowledge, this is the first time an approach like this is

employed with an information measure. This is the main point of the present study.

Despite the fact that a1ApEn has consistency [14], a single a1ApEn value might not be enough to explore the underlying process that generates a time-series. This is also true for other information measures, but the issue is seldom put forward (see, for example, [9,19]).

In the realm of empirical time-series, one has often to face the problem of having vectors of very different sizes. So, what one should do? To discard part of the data to have all vectors of the same size in order to make comparisons reliable seems a not quite cunning procedure (cf. [9] in Introduction), since data are not easily acquired most of the time. A useful procedure, sometimes employed, is to fix a "small *N*" and to obtain multiple values of the estimator along the series by a moving window ([20–22]). The present study offers an alternative and complementary approach that adds new information to the analysis, without discarding data.

Let us show some practical examples. Table 8 contains mean values of a1ApEn computed for *N* = 200. Clearly, the logistic map with control parameter μ = 3.6 is recognized as "less complex" than the one with μ = 3.99. This is a fair conclusion, but it says nothing more than this. However, when we take the pattern of correlations into account, a much more interesting picture emerges: the same process might originate these two time-series (obviously, here, we know that this is the case, indeed).

Consider the plain values of a1ApEn of the heart rate data (Table 8). It is tempting to classify them within the range of the deterministic maps set. This would incite heated debates regarding the origin of such a deterministic process underlying heart rate control ([23–27]). On the other hand, when we exam the pattern of the size-related behavior of a1ApEn for these time-series, we see that these empirical data resemble unlimited random processes. A similar conclusion is put forward by Lake [28], who concludes that the Gaussianity of heart rate complexity is associated with adequate physiological responses. The reasons for that are beyond the scope of the present study.

An analogous argument occurs regarding the interpretation of solar wind data, with some advocating a deterministic origin of the process that generates variability [29–31] while others consider it as a stochastic process [32–35]. From the values in Table 8, one would consider the process as belonging to the deterministic set. On the other hand, as Table 7 shows, the patterns detected belong to unlimited random processes and the size-related behavior of a1ApEn strongly suggests that the process is from a stochastic nature. Once again, it is not our purpose to take one side in the debate. Instead, our focus is to demonstrate the usefulness of the present approach and its potential to unveil additional information even in commonly studied processes.

Table 8. Mean a1ApEn values for the resampling procedure with *N* = 200

	Data source	a1ApEn *N* = 200
in machina generated processes	Beta2	0.381
	Henon	0.325
	L3.6	0.147
	L3.68	0.232
	L3.9	0.313
	L3.99	0.382
	AR1	0.340
	AR2	0.335
	Beta4	0.246
	Lévy	0.077
	NDRN	0.344
	Beta1	0.669
	Beta3	0.729
	UDRN	0.536
<R-R> interval	1	0.298
	2	0.201
	3	0.233
	4	0.288
	5	0.276
	6	0.171
	7	0.197
Solar wind	MF(23^{rd} inc)	0.254
	MF(23^{rd} dec)	0.261
	MF(24^{th} inc)	0.283
	PT(23^{rd} inc)	0.260
	PT(23^{rd} dec)	0.283
	PT(24^{th} inc)	0.253

4. CONCLUSION

The combination of plain values of a1ApEn and the pattern of the correlations of its size-related properties are highly informative in terms of the identification of the underlying process that generates a given time-series.

ETHICAL APPROVAL

ECG data collection was approved by the Comissão de Ética no Uso de Animais - Instituto de Biociências (CEUA-IB/ CAAE:

34609214.6.0000.5464) and each volunteer gave informed consent.

ACKNOWLEDGEMENTS

This study was supported by a research grant from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPEPSP #2014/08842-3).

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Piqueira JRC, Serboncini FA, Monteiro LHA. Biological models: Measuring variability with classical and quantum information. J. Theor. Biol. 2006;242:309–313.
DOI: 10.1016/j.jtbi.2006.02.019
2. Masi M. A step beyond Tsallis and Rényi entropies. Phys. Lett. A. 2005;338:217–224.
DOI: 10.1016/j.physleta.2005.01.094
3. Kantz H, Schreiber T. Nonlinear time series analysis. 2nd ed., Cambridge University Press; 2004.
4. Gao Z-K, Jin N-D. A directed weighted complex network for characterizing chaotic dynamics from time series. Nonlinear Anal. Real World Appl. 2012;13:947–952.
DOI: 10.1016/j.nonrwa.2011.08.029
5. Pincus SM. Approximate entropy as a measure of system complexity. Proc. Natl. Acad. Sci. 1991;88:2297–2301.
DOI: 10.1073/pnas.88.6.2297
6. Pincus SM, Huang W. Approximate entropy: Statistical properties and applications. Commun. Stat. - Theory Methods. 1992;21:3061–3077.
DOI: 10.1080/03610929208830963
7. Rukhin AL. Approximate entropy for testing randomness. J. Appl. Probab. 2000;37:88–100.
DOI: 10.1239/jap/1014842270
8. Grandy WTJ. Entropy and the time evolution of macroscopic systems. Oxford University Press, Oxford; 2012.
9. Bruijn SM, van Dieën JH, Meijer OG, Beek PJ. Statistical precision and sensitivity of measures of dynamic gait stability. J. Neurosci. Methods. 2009;178:327–333.
DOI: 10.1016/j.jneumeth.2008.12.015
10. Lu S, Chen X, Kanters JK, Solomon IC, Chon KH. Automatic selection of the threshold value R for approximate entropy. IEEE Trans. Biomed. Eng. 2008;55:1966–72.
DOI: 10.1109/TBME.2008.919870
11. Chon K, Scully CG, Lu S. Approximate entropy for all signals. IEEE Eng. Med. Biol. Mag. 2009;28:18–23.
DOI: 10.1109/MEMB.2009.934629
12. Castiglioni P, Di Rienzo M. How the threshold "r" influences approximate entropy analysis of heart-rate variability. Comput. Cardiol. 2008;561–564.
DOI: 10.1109/CIC.2008.4749103
13. Santos BT, Martins RA, Natali JES, Rodrigues VH, Marques FS, Chauí-Berlinck JG. Consistency in approximate entropy given by a volumetric estimate. Chaos, Solitons & Fractals. 2009;42:322–334.
DOI: 10.1016/j.chaos.2008.12.002
14. Natali JES, Chaui-Berlinck JG. Area 1 of approximate entropy as a fast and robust tool to address temporal organization. Br. J. Appl. Sci. Technol. 2016;13:1–11.
DOI: 10.9734/BJAST/2016/22726
15. Pincus SM. Approximate entropy as a measure of system complexity. Proc. Natl. Acad. Sci. USA. 1991;88:2297–2301.
DOI: 10.1073/pnas.88.6.2297
16. Shannon CE. A mathematical theory of communication. ACM SIGMOBILE Mob. Comput. Commun. Rev. 1948;27:379–423.
17. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am. J. Physiol. Heart Circ. Physiol. 2000;278: H2039–H2049.
18. Costa M, Goldberger AL, Peng C-K. Multiscale entropy analysis of complex physiologic time series. Phys. Rev. Lett. 2002;89:68-102.
DOI: 10.1103/PhysRevLett.89.068102
19. Costa M, Goldberger AL, Peng C-K. Multiscale entropy analysis of biological signals. Phys. Rev. E. 2005;71:21906.
DOI: 10.1103/PhysRevE.71.021906
20. Zhou P, Barkhaus PE, Zhang X, Rymer WZ. Characterizing the complexity of spontaneous motor unit patterns of amyotrophic lateral sclerosis using approximate entropy. J. Neural Eng. 2011;8:66010.
DOI: 10.1088/1741-2560/8/6/066010

21. Hu X, Miller C, Vespa P, Bergsneider M. Adaptive computation of approximate entropy and its application in integrative analysis of irregularity of heart rate variability and intracranial pressure signals. Med. Eng. Phys. 2008;30:631–639.
DOI: 10.1016/j.medengphy.2007.07.002
22. Chen L, Luo W, Deng Y, Wang Z, Zeng S. Characterizing the complexity of spontaneous electrical signals in cultured neuronal networks using approximate entropy. IEEE Trans. Inf. Technol. Biomed. 2009;13:405–10.
DOI: 10.1109/TITB.2008.2012164
23. Goldberger AL. Non-linear dynamics for clinicians: Chaos theory, fractals, and complexity at the bedside. Lancet. 1996;347:1312–1314.
DOI: 10.1016/S0140-6736(96)90948-4
24. Pikkujamsa SM, Makikallio TH, Sourander LB, Raiha IJ, Puukka P, Skytta J, Peng C-K, Goldberger AL, Huikuri HV. Cardiac interbeat interval dynamics from childhood to senescence : Comparison of conventional and new measures based on fractals and chaos theory. Circulation. 1999;100:393–399.
DOI: 10.1161/01.CIR.100.4.393
25. Glass L, Goldberger AL, Courtemanche M, Shrier A. Nonlinear dynamics, chaos and complex cardiac arrhythmias. Proc. R. Soc. A Math. Phys. Eng. Sci. 1987;413:9–26.
DOI: 10.1098/rspa.1987.0097
26. Poon C-S, Merrill CK. Decrease of cardiac chaos in congestive heart failure. Nature. 1997;389:492–495.
DOI: 10.1038/39043
27. Garfinkel A, Spano M, Ditto W, Weiss J. Controlling cardiac chaos. Science. 1992;257(80);1230–1235.
DOI: 10.1126/science.1519060
28. Lake DE. Renyi entropy measures of heart rate Gaussianity. IEEE Trans. Biomed. Eng. 2006;53:21–27.
DOI: 10.1109/TBME.2005.859782
29. Macek WM. Testing for an attractor in the solar wind flow. Phys. D Nonlinear Phenom. 1998;122:254–264.
DOI: 10.1016/S0167-2789(98)00098-0
30. Redaelli S, Macek WM. Lyapunov exponent and entropy of the solar wind flow. Planet. Space Sci. 2001;49:1211–1218.
DOI: 10.1016/S0032-0633(01)00062-9
31. Komm RW. Hurst analysis of Mt. Wilson rotation measurements. Sol. Phys. 1995;156:17–28.
DOI: 10.1007/BF00669572
32. Oliver R, Ballester JL. Is there memory in solar activity? Phys. Rev. E. 1998;58: 5650–5654.
DOI: 10.1103/PhysRevE.58.5650
33. Mininni PD, Gómez DO, Mindlin GB. Stochastic relaxation oscillator model for the solar cycle. Phys. Rev. Lett. 2000;85: 5476–5479.
DOI: 10.1103/PhysRevLett.85.5476
34. Price CP, Prichard D, Hogenson EA. Do the sunspot numbers form a “chaotic” set? J. Geophys. Res. 1992;97:19113.
DOI: 10.1029/92JA01459
35. Rypdal M, Rypdal K. Testing hypotheses about sun-climate complexity linking. Phys. Rev. Lett. 2010;104:128501.
DOI: 10.1103/PhysRevLett.104.128501

APPENDIX

SUPPORTING INFORMATION - CODE FOR COMPUTING TOLERANCES

```
function [ro,rn,GG] = tolerances_a1(x,LL)
%Computes tolerances (absolute and relative) covering
%the whole interval from 0% to 100% of equality among the values in the
%time series.
%
%Call:
%[r_orig , r_norm] = tolerances_a1(x , grain_level);
%
%Input
%x = time series (vector with length = N)
%
%grain_level = a value between 0 and 1 (exclusive) that allows the user to
%set a level to generate a "coarse grainy" warning (without halting the
%computations) - default = 0.9
%
%%Output:
%r_orig = tolerance vector
%
%r_norm = normalized tolerance vector
%
%grainy of the series - the finest the grainy, the more diverse are
%the values in the data (more number of states).
%The maximum value GRAIN may attain is (N-1)/N,
%corresponding to a binary series (and a coarse grainy). The minumum value is
%1/N, corresponding to a fine grainy (and, in such a case, all the elements
%of the series are different from each other).

%Checking the grainy level input
if nargin == 1
   LL = 0.9;
else
   if LL >=1 | LL <= 0
      LL = 0.9;
   end
end

Tam = length(x);

%Deltas in the time series
z = sort(x);
zz=z;
z(Tam)=[];
zz(1)=[];
%delta vector
D = zz-z;

%Maximum distance between values in the time series
Dmax = sum(D);

%Preparing for next steps
D = sort(D);

%Finding zeros (if there are equal values in the time series)
```

```
a = find(D == 0);

%deltas' base
Da = D;

%extracting the zeros
Da(a) = [];

%Size of the non-zero elements
CDelta = length(Da);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%
%Grainy
%Note: the time-series was sorted so elements of equal values are located
%next to each other, generating deltas = 0.
%The size of the non-zeros deltas (CDelta above) is, therefore, the number
%of non-equal elements in the series minus one (because we need a pair of
%values to generate one delta). The greater the size of CDelta, the more
%grainy the series is.
%Thus, the grainy is defined as a ratio between the size of the equal
%elements and the size of the series. The closer such a ratio is to 1, the
%less grainy the series is (i.e., coarse grain)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%

gg = (Tam - CDelta)/Tam;

%if gg == (Tam - 1)/Tam
   %This is the highest value of gg in a series of size Tam
%     error('Time series identified as binary. Use adequate tools for analysis')

 %    elseif gg >= LL*((Tam - 1)/Tam)

 %    Grainy = gg
 %    Number_of_states = CDelta

 %    warning('Coarse grain detected. Be careful with analysis/interpretation')
%end

GG = [gg CDelta];

%Partitioning for the tolerance vector

Dmin = min(Da);
Amplitude = Dmax - Dmin;

%Preparing, once again, the delta vector to construct the tolerance vector
a = find(D == 0);
Da = D;
Da(a) = [];
tamanho = length(Da);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%
%   Absolute Tolerance Vector
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%

%The contruction of the tolerance vector will depend on the size of the
%delta vector

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%
if tamanho <= 300
   %The first value of the tolerance vector is zero
   ro(1) = 0;
   for nn = 2:tamanho
      ro(nn,1) = sum(Da(1:nn-1));
   end
   % Complete Original Tolerance Vector
   ro(tamanho+1,1) = Dmax;

else
   %For time series with more than 300 delta values, the first 50 deltas
   %are employed directly. The rest of the tolerance vector is constructed
   %using fractions (e.g., 0.1 etc.) of the total amplitude.
   for nn = 1:50
      r1(nn) = sum(Da(1:nn));
   end

   ref2 = sum(Da(1:51));
   ref1 = ref2/Amplitude;

   if ref1 <= 0.02
      pontos = 10*(1 + ceil((5-(100*ref1))));
      p5 = 0.05*Amplitude;
      p35 = 0.35*Amplitude;
      r3 = (ref2 : ((p5-ref2)/pontos) : p5);
      r4 = (p5 : ((p35-p5)/200) : p35);
      r5 = (p35 : ((Dmax-p35)/100) : Dmax);
      r4(1) = [];
      r5(1) = [];
      r2 = [r3 r4 r5];
   elseif ref1 > 0.02 & ref1 < 0.1
      pontos = 10*(1 + ceil((11-(100*ref1))));
      p11 = 0.11*Amplitude;
      p35 = .35*Amplitude;
      r3 = (ref2 : ((p11-ref2)/pontos) : p11);
      r4 = (p11 : ((p35-p11)/150) : p35);
      r5 = (p35 : ((Dmax-p35)/100) : Dmax);
      r4(1) = [];
      r5(1) = [];
      r2 = [r3 r4 r5];
   elseif ref1 >= 0.1 & ref1 <= 0.2
      p35 = .35*Amplitude;
      r3 = (ref2 : ((p35-ref2)/150) : p50);
      r4 = (p35 : ((Dmax-p35)/100) : Dmax);
      r4(1) = [];
      r2 = [r3 r4];
```

```
  elseif ref1 > 0.2 & ref1 <= 0.35
    p50 = .5*Amplitude;
    r3 = (ref2 : ((p50-ref2)/150) : p50);
    r4 = (p50 : ((Dmax-p50)/50) : Dmax);
    r4(1) = [];
    r2 = [r3 r4];
  elseif ref1 > 0.35 & ref1 <= 0.5
    r2 = (ref2 : ((Dmax-ref2)/100) : Dmax);
  elseif ref1 > 0.5
    r2 = (ref2 : ((Dmax-ref2)/50) : Dmax);
  end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%
  % Complete Original Tolerance Vector - note the first element is zero

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%
  ro = [0 r1 r2]';

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%
% Relative Tolerances
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%
rn = ro./Dmax;
```


Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/16856