

Research Article

LMC and SDL Complexity Measures: A Tool to Explore Time Series

José Roberto C. Piqueira ¹ and Sérgio Henrique Vannucchi Leme de Mattos ²

¹Escola Politécnica da Universidade de São Paulo, Avenida Prof. Luciano Gualberto, travessa 3, n. 158, 05508-900 São Paulo, SP, Brazil

²Universidade Federal de São Carlos, Rod. Washington Luís km 235, SP-310, 13565-905 São Carlos, SP, Brazil

Correspondence should be addressed to José Roberto C. Piqueira; piqueira@lac.usp.br

Received 21 September 2018; Revised 24 November 2018; Accepted 10 December 2018; Published 2 January 2019

Guest Editor: Jose Garcia-Rodriguez

Copyright © 2019 José Roberto C. Piqueira and Sérgio Henrique Vannucchi Leme de Mattos. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This work is a generalization of the López-Ruiz, Mancini, and Calbet (LMC) and Shiner, Davison, and Landsberg (SDL) complexity measures, considering that the state of a system or process is represented by a continuous temporal series of a dynamical variable. As the two complexity measures are based on the calculation of informational entropy, an equivalent information source is defined by using partitions of the dynamical variable range. During the time intervals, the information associated with the measured dynamical variable is the seed to calculate instantaneous LMC and SDL measures. To show how the methodology works generating indicators, two examples, one concerning meteorological data and the other concerning economic data, are presented and discussed.

1. Introduction

The word complexity, in the common sense meaning, represents systems that are difficult to describe, design, or understand. However, since Kolmogorov presented the concept of computational complexity [1], new ideas have been associated with this word, mainly in life sciences [2], relating complexity, and information [3].

As a consequence, complexity started to be associated with systems and with the emergence of unexpected behaviors, due to nonlinearities [4, 5] and, concerning system theory [6], a new meaning was carved, postulating that complexity is half way of the equilibrium and disequilibrium [7].

Developing this idea, in a seminal paper [8], López-Ruiz, Mancini, and Calbet proposed the LMC (López-Ruiz, Mancini, and Calbet) complexity measure for a random distribution by using informational entropy [9] to evaluate equilibrium, and the quadratic deviation from the uniform distribution to evaluate disequilibrium.

However, there has been some criticism about the LMC measure, considering that it is inaccurate for some classes of systems obeying Markovian chains and cannot be considered

to represent an extensive variable. Feldman and Crutchfield [10] proposed a correction for the disequilibrium term, replacing it by the relative entropy with respect to the uniform distribution.

Shiner, Davison, and Landsberg proposed another modification of the LMC measure, replacing the disequilibrium term by the complement of the equilibrium term. This measure is called SDL (Shiner, Davison, and Landsberg) [11] and presents conclusions similar to that obtained by using LMC, for the majority of usual statistical distributions [2].

The main restriction to LMC and SDL complexity measures is due to Crutchfield, Feldman, and Shalizi, as they argue that an equilibrium system can be structurally complex [12], but this problem could be solved by weighting order and disorder, according to the specific problem to be analyzed.

Since the early 2000s, the idea of adapting LMC and SDL to dynamical systems was successfully applied to different types of time evolution problems: bird songs [13], neural plasticity [14], interactions between species in ecological systems [2], physiognomies of landscapes [15], economic series [16], spread depression [17], and quantum information [18].

With these ideas in mind, this article presents a systematization of the methodology used in the referred papers,

based on LMC and SDL measures, to be applied to temporal series, by defining and calculating the dynamic complexity measures.

The procedure, applied to a temporal series representing some organizational or functional aspect of a system, provides insights regarding the evolution of its complexity.

As the LMC and SDL dynamical measures are based on informational entropy [16], the first task, described in the next section, is to define an alphabet source, associating a probability distribution with the possible system states.

Following the definition of the probability distribution, a new section defines how dynamical LMC and SDL measures can be calculated at each time, based on the individual information associated with the system state at this time, generating temporal series for LMC and SDL measures.

To illustrate the calculation procedure, two examples are presented: one related to a meteorological time series and the other to an economic time series. In both cases section, a practical discussion about how to divide the range of the values assumed by the system state is presented.

The examples were chosen to show that the methodology can be applied to different types of phenomena: precipitation (first example) with strong periodic component and economic time series (second example) that seems to be random.

The work is closed with a conclusion section, emphasizing that the same procedure can be applied to any kind of temporal real numbers series, even with different temporal scale, to calculate complexity measures.

2. Defining Source and Probability Distribution for a Temporal Series

Considering Shannon's model [9] for an information source, a time series $x(n)$ is considered to be a function of the nonnegative integers into a real interval, i.e., $x(n) : Z_+ \rightarrow (a, b)$, associating with each time $t_0 + nT$ a real number belonging to (a, b) , with $t_0 > 0$ being the initial instant and $T > 0$ an arbitrary period, depending on the data availability.

The set $x(t_0), x(t_0 + T), \dots, x(t_0 + nT)$ is assumed to be a sequence of independent random variables and the stochastic process $x(n)$ as a whole is stationary [19].

The first step is to divide the interval (a, b) into N subintervals. For the sake of simplicity, N is chosen equal to $2^k, k \in Z_+$.

At this point, it could be asked how to choose N , as there is a compromise between precision (high values of N) and speed of calculation (low values of N). This question will not be addressed theoretically; however, in the example section, practical hints about this choice are presented.

Consequently, the source alphabet is defined by the intervals $A_i, i = 1, \dots, N$, with $\bigcup_{i=1}^N A_i = (a, b)$ and $A_i \cap A_j = \phi, \forall i \neq j$.

Then, a time interval defined by a given n must be chosen, and for the time sequence $t_0, t_0 + T, \dots, t_0 + nT$ the values of the variable $x(n)$ must be read and associated with the intervals A_i , containing their respective value.

Therefore, for the whole set $t_0, t_0 + T, \dots, t_0 + nT$, each interval A_i belonging to the source alphabet is associated with

$x(n)$ a certain number of times n_i , which defines a relative frequency $p_i = n_i/(n + 1)$.

As $p_i \geq 0$ and $\sum_{i=1}^N p_i = 1$, it can be taken as a probability, associated with each interval A_i .

Following the definition, for each subinterval $A_i \subset (a, b)$, its individual contribution to the whole information entropy is given by $S_i = -p_i \log_2 p_i$; and the maximum value of the informational entropy for the whole source, $S_{max} = \log_2 N = k$, can be calculated [9].

3. Dynamical LMC and SDL

As the source alphabet and individual information were defined, the instantaneous values of $x(n)$ are associated with their respective S_i , allowing the calculation of the instantaneous value of the equilibrium (disorder) term:

$$\Delta(n) = \frac{S_i}{k}. \quad (1)$$

Combining (1) with the different definitions of the disequilibrium (order) terms, dynamical LMC and SDL measures are defined.

3.1. LMC Dynamical Measure. As indicated by López-Ruiz, Mancini, and Calbet [8], the dynamic disequilibrium (order) term can be calculated as the quadratic deviation of the source alphabet probability distribution from the uniform distribution and, consequently, the individual contribution of each interval A_i is

$$D(n) = \left(p_i - \frac{1}{N}\right)^2. \quad (2)$$

Extending the definition of LMC measure, dynamical LMC, calculated in $t_0 + nT$, is given by

$$C_{LMC}(n) = \Delta(n).D(n). \quad (3)$$

3.2. SDL Dynamical Measure. As proposed by Shiner, Davison, and Landsberg [11], the dynamic disequilibrium (order) term can be calculated as the complement of the dynamic equilibrium term:

$$D(n) = (1 - \Delta(n)). \quad (4)$$

Extending the definition of SDL measure, dynamical SDL, calculated in $t_0 + nT$, is given by

$$C_{SDL}(n) = \Delta(n).D(n). \quad (5)$$

4. Applying the Method to Meteorological Data

A monthly meteorological temporal series is studied in this section, showing that the described method can be applied, independently of the natural time scale and periodicity of the phenomenon.

The meteorological data series relative to rain precipitation in Dourados-MS-Brazil [20] is analyzed, only in a

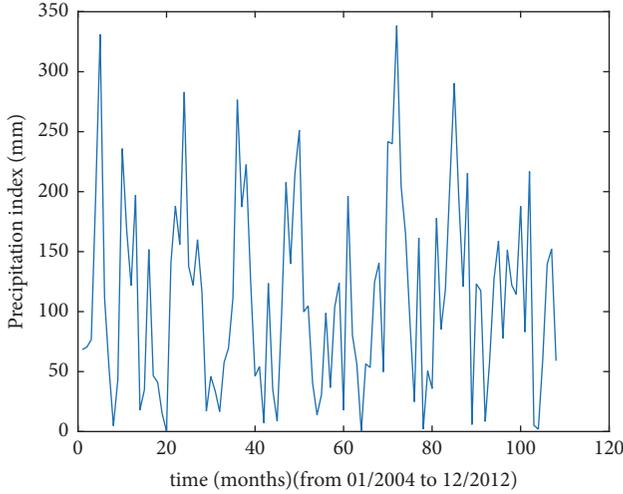


FIGURE 1: Precipitation index (Dourados-MS-Brasil).

methodological point of view, without any meteorological conjecture about the results.

The monthly precipitation index temporal series, from January 2004 to September of 2012, shown in Figure 1, represents the value of $x(n)$ [20], whose complexity is analyzed.

Consequently, the interval (a, b) related to the excursion of x is $(0; 338)$. It is divided into $8(k = 3)$, $16(k = 4)$, $32(k = 5)$, and $64(k = 6)$ subintervals to build the sources and their respective probability distributions.

Based on these probability distributions, $C_{LMC}(n)$ and $C_{SDL}(n)$ are calculated and plotted giving an idea about how the measure choice and the interval division affect the results.

4.1. Equivalence between LMC and SDL. Dividing the range of $x(n)$ into 8 parts, the results of the calculation of $C_{LMC}(n)$ and $C_{SDL}(n)$ measures are shown in Figures 2(a) and 2(b), respectively.

As Figures 2(a) and 2(b) show, in spite of the numerical differences, the time evolutions of $C_{LMC}(n)$ and $C_{SDL}(n)$ are represented by similar curves, in the eight-part division case.

Observing the figures, it is possible to infer that the LMC measure captures the periodic character of the precipitation along the years in a better way. However, the SDL measure assumes its maximum value (.25).

If the range of $x(n)$ is divided into 16 parts, Figures 3(a) and 3(b) show the results for $C_{LMC}(n)$ and $C_{SDL}(n)$.

It can be observed that, in this case (sixteen-division case), $C_{LMC}(n)$ and $C_{SDL}(n)$ differ by a scale factor, with LMC measures presenting better accuracy to express the periodic character of the rain seasons. SDL measure presents high value of peaks but the maximum value (.25) is not reached.

Comparing Figures 2(a) and 3(a), $C_{LMC}(n)$ for different range partitions, the global aspects of the curves are the same and, by increasing the number of divisions, the dynamical range of the measures decreases, and some rapid oscillatory variations similar to noise appear.

Comparing Figures 2(b) and 3(b), $C_{SDL}(n)$ for different range partitions, the whole aspects of the curves are the same

and the noisy aspect due to the increasing number of interval divisions is similar to the presented by $C_{LMC}(n)$.

4.2. Range Interval Partition. As it was observed, the dynamical range of the both measures decreases as the number of divisions increases, as a consequence of the fact that the number of elements of the source increases provoking more uniform distribution of the possible measures.

To better understand this phenomenon, the measures are recalculated by increasing the number of intervals of $x(n)$, and the result for a thirty-two partition is shown in Figure 4(a) for $C_{LMC}(n)$ and in Figure 4(b) for $C_{SDL}(n)$.

By analyzing the results from Figures 2(a), 3(a), and 4(a), it could be observed that, by increasing the number of intervals, the dynamical range of $C_{LMC}(n)$ decreases but, apparently, for this long series, the temporal evolution of $C_{LMC}(n)$ maintains its qualitative behavior mixing noise with accuracy.

By analyzing the results from Figures 2(b), 3(b), and 4(b), it could be observed that, by increasing the number of intervals, the dynamical range of $C_{SDL}(n)$ decreases and its maximum value (.25) is not reached. Apparently, for this long series, the temporal evolution of $C_{SDL}(n)$ maintains its qualitative behavior mixing noise with accuracy.

5. Applying the Method to Economic Data

In this section, the economic series relative to the conversion of currencies studied in [16] is taken as an example, showing the applicability of the methodology for random phenomena.

The temporal series related to the daily dollar to Brazilian real (USD/BR) conversion rate, from January 1999 to September of 2015, shown in Figure 5 [16] is analyzed, only in a methodological point of view, without any economic conjecture about the results.

This conversion rate represents the value of $x(n)$, whose complexity is analyzed.

Consequently, the interval (a, b) related to the excursion of x is $(1.207; 4.178)$. It is divided into $8(k = 3)$, $16(k = 4)$, $32(k = 5)$, and $64(k = 6)$ subintervals to build the sources and the respective probability distributions.

Based on these probability distributions, $C_{LMC}(n)$ and $C_{SDL}(n)$ are calculated and plotted giving an idea about how the measure choice and the interval division affect the results.

5.1. Equivalence between LMC and SDL. Dividing the range of $x(n)$ into 8 parts, the results of the calculation of $C_{LMC}(n)$ and $C_{SDL}(n)$ measures are shown in Figures 6(a) and 6(b), respectively.

As Figures 6(a) and 6(b) show, in spite of the numerical differences, the time evolution of $C_{LMC}(n)$ and $C_{SDL}(n)$ are qualitatively the same and represented by very similar curves, in the eight-part division case.

If the range of $x(n)$ is divided into 16 parts, Figures 7(a) and 7(b) show the results for $C_{LMC}(n)$ and $C_{SDL}(n)$.

It can be observed that, in this case (sixteen-division case), $C_{LMC}(n)$ and $C_{SDL}(n)$ differ only by a scale factor, with the same qualitative time evolution.

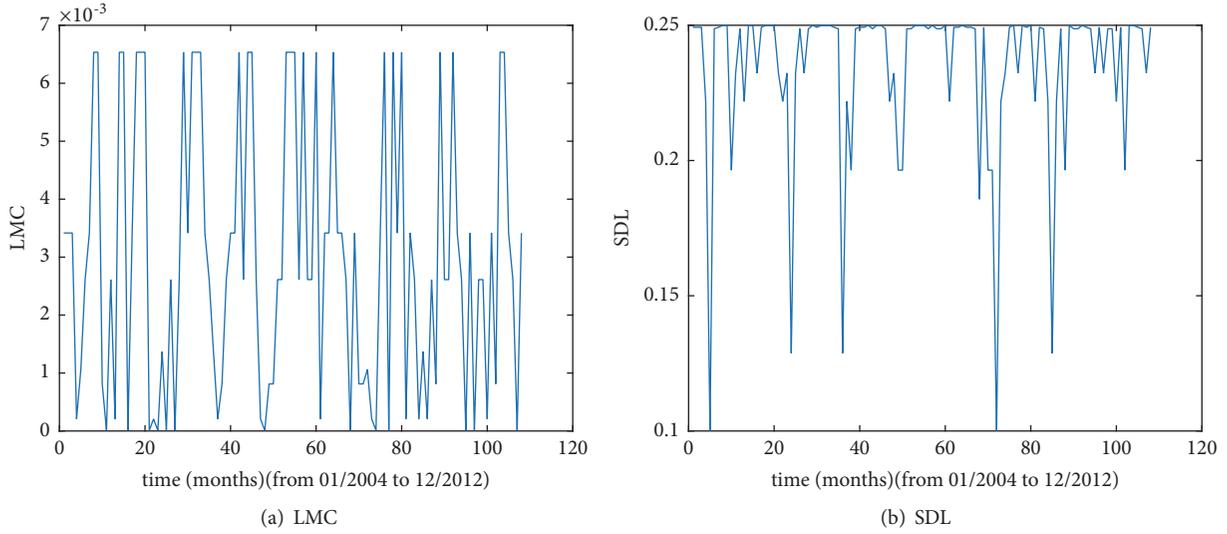


FIGURE 2: Temporal evolution of complexity (8-part division).

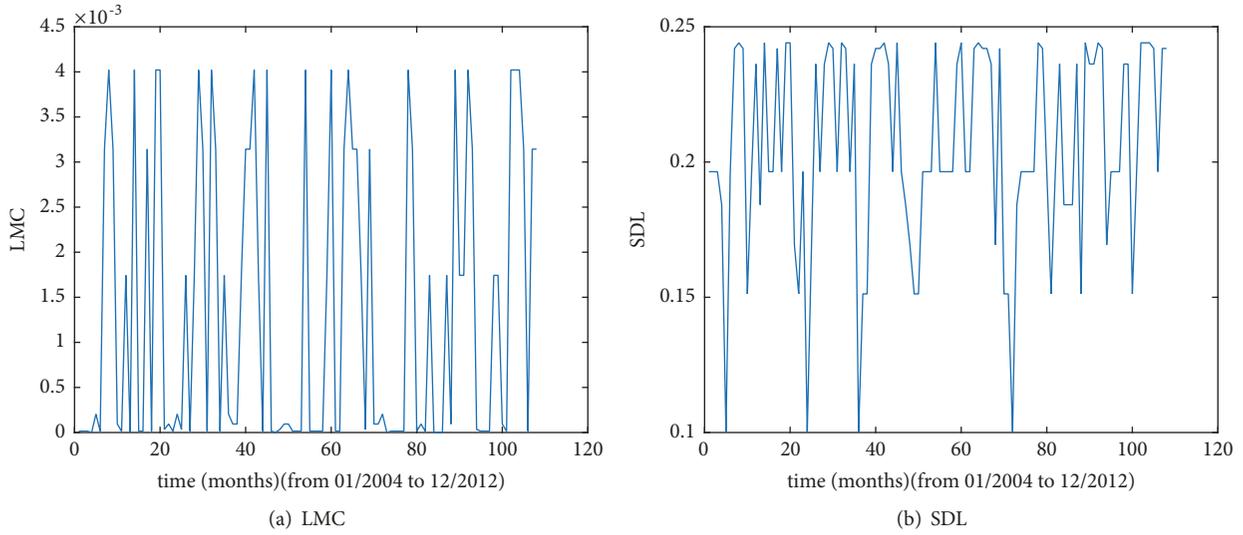


FIGURE 3: Temporal evolution of complexity (16-part division).

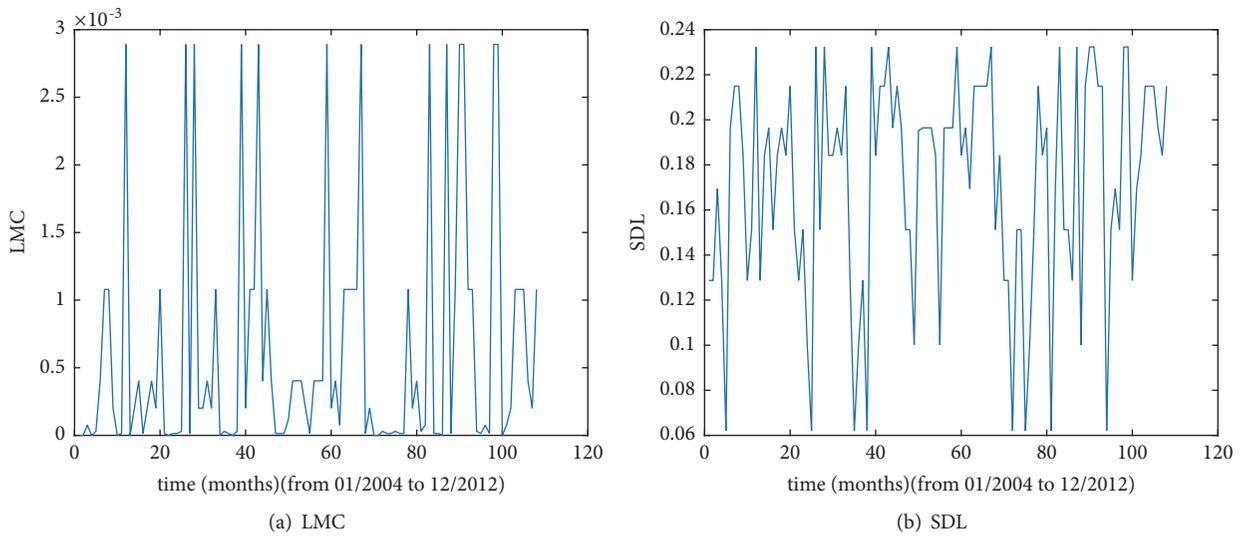


FIGURE 4: Temporal evolution of complexity (32-part division).

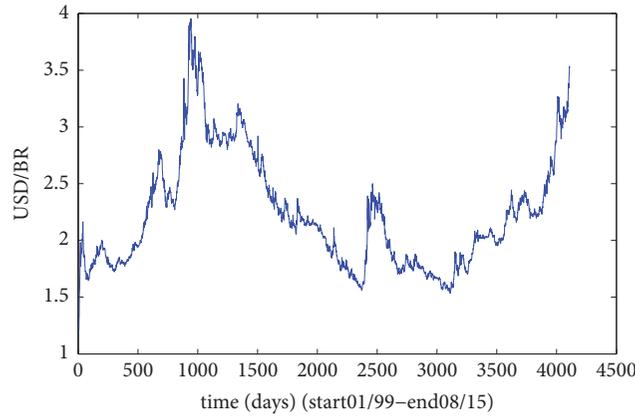
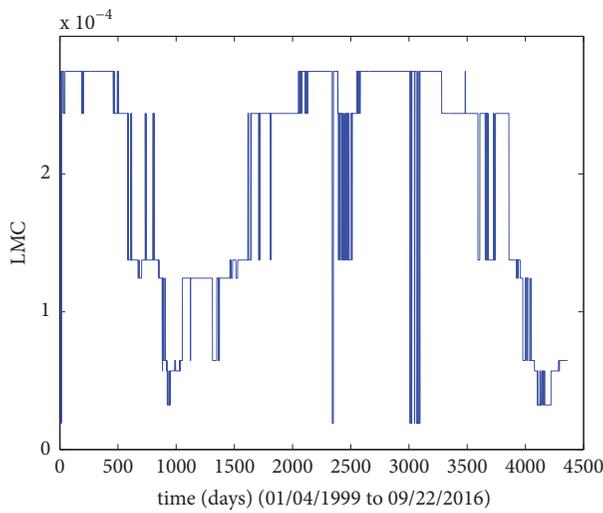
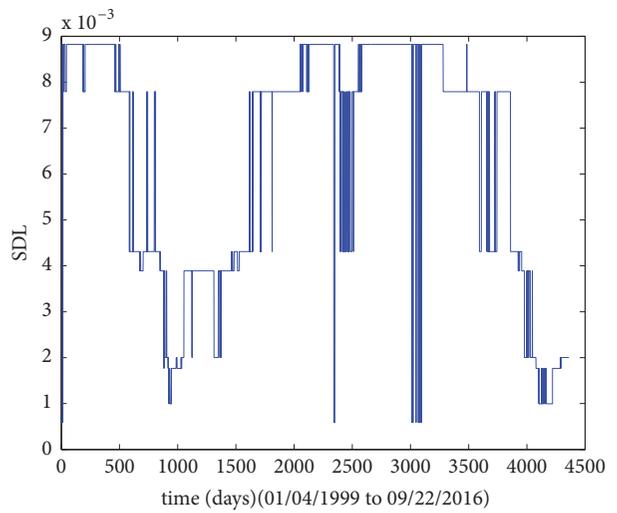


FIGURE 5: USA Dollar-Brazilian real exchanging rate.

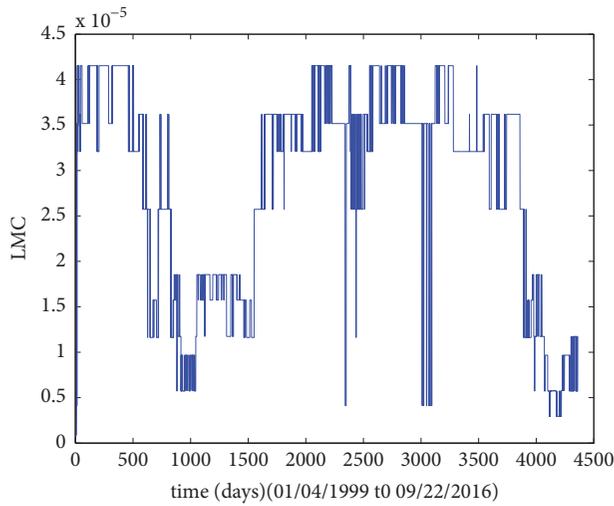


(a) LMC

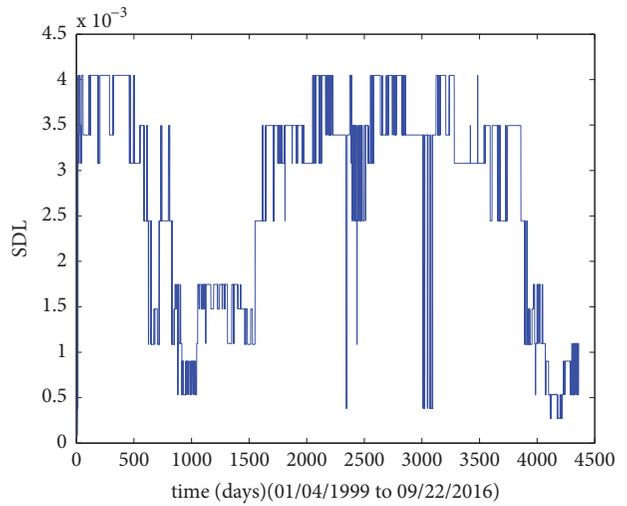


(b) SDL

FIGURE 6: Temporal evolution of complexity (8-part division).



(a) LMC



(b) SDL

FIGURE 7: Temporal evolution of complexity (16-part division).

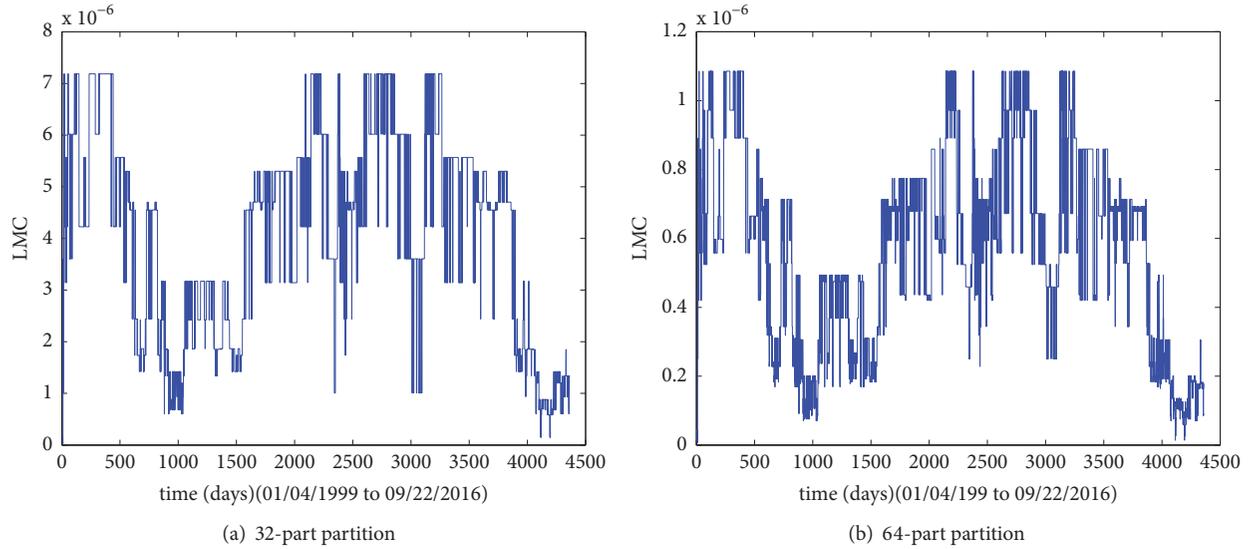


FIGURE 8: Temporal evolution of LMC complexity.

Comparing Figures 6(a) and 7(a), $C_{LMC}(n)$ for different range partitions, the whole qualitative aspects of the curves are the same and by increasing the number of divisions, the dynamical range of the measures changes, implying some rapid oscillatory variations, similar to noise.

Comparing Figures 6(b) and 7(b), $C_{SDL}(n)$ for different range partitions, the whole qualitative aspects of the curves are the same and the noisy aspect due to the increasing number of interval divisions is maintained.

Consequently, from now on, only LMC measure will be analyzed, since SDL presents the same qualitative dynamical behavior and partition sensitivity.

5.2. Range Interval Partition. By increasing the number of intervals of $x(n)$ and recalculating $C_{LMC}(n)$, the result for a thirty-two partition is shown in Figure 8(a) and, for a sixty-four partition, in Figure 8(b).

By analyzing the results from Figures 6(a), 7(a), 8(a), and 8(b), it could be observed that, by increasing the number of intervals, the maximum value of $C_{LMC}(n)$ decreases improving the precision but, apparently, for this long series, the temporal evolution of $C_{LMC}(n)$ maintains its qualitative behavior mixing noise with accuracy.

Attempting to be more precise about how the range interval partition, $C_{LMC}(n)$ is calculated for the several partitions, but considering a shorter time period for the data. The interval between July and December of 2002 is chosen, because, as explained in [16], it is critical concerning the conversion rates in Brazil.

Figures 9(a), 9(b), 9(c), and 9(d) show the LMC dynamical measure calculated, for the initial set of data, with the range interval divided into 8, 16, 32, and 64 parts, respectively.

It can be observed from these results that, for shorter intervals, the general qualitative characteristics of the time evolution appear, independently on the partition. However, as the number of subintervals increases, the instantaneous

numerical values change but the precision increases, allowing more accurate analysis.

6. Conclusions

A methodology for calculating LMC and SDL dynamical complexity was developed, starting with the construction of a source and a probability distribution, for any temporal series. The contribution is just concerning to extend ideas, mainly applied to static situations, to temporal evolution of variables representing some kind of organization phenomenon.

LMC and SDL measures were observed to be equivalent in some temporal analyses but, when there is a strong oscillatory component, the LMC measure seems to be more accurate to express the temporal evolution of the complexity, as the meteorological data analysis shows.

For more randomly distributed data, the two measures (LMC and SDL) present the same accuracy, as the economic data analysis shows.

A point that is always an object of discussion is the range interval partition. The choice of the number of subintervals is a matter of experience.

Long time intervals are not so sensitive to the increase of the number of divisions, in spite of meteorological data being more sensitive than economic. However, for short time intervals, increasing the number of divisions produces a less precise analysis, introducing noise.

The examples presented were just to illustrate the methodological approach, without any compromise with meteorologic or economic conclusions that can be inferred by a specialist, by using the developed tool.

Data Availability

All the data used in this paper are available in the Internet by following the links given in the references.

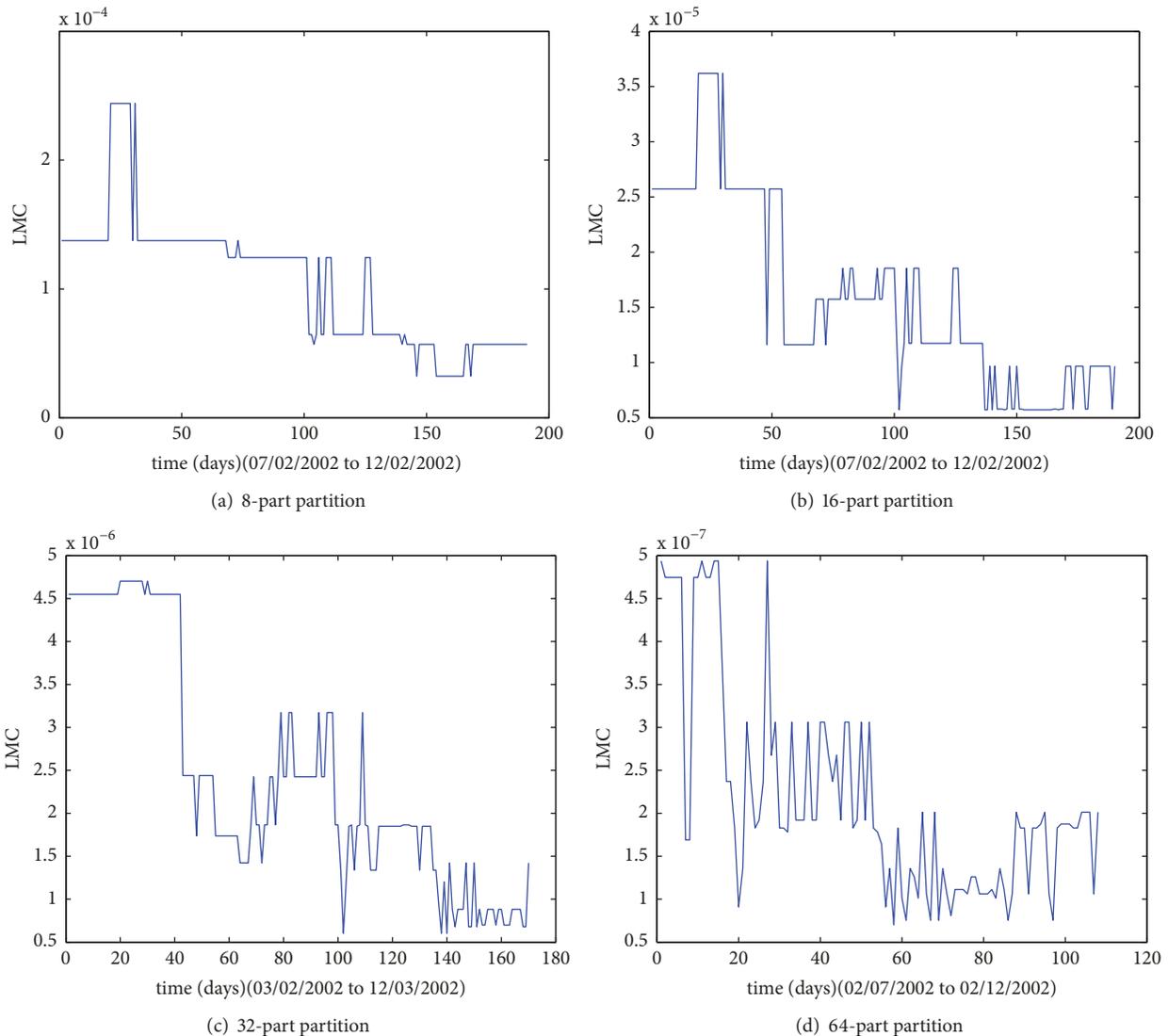


FIGURE 9: LMC temporal evolution for shorter time intervals.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

All the data used in this paper are available in the Internet by following the links given in the references. This work was supported by CNPq, Brazil.

References

- [1] A. N. Kolmogorov, "Three approaches to the definition of the concept "quantity of information", *Akademiya Nauk SSSR. Institut Problem Peredachi Informatsii Akademii Nauk SSSR. Problemy Peredachi Informatsii*, vol. 1, no. vyp. 1, pp. 3–11, 1965.
- [2] M. Anand and L. Orlóci, "Complexity in plant communities: The notion and quantification," *Journal of Theoretical Biology*, vol. 179, no. 2, pp. 179–186, 1996.
- [3] H. Haken, *Information and Self-Organization*, Springer Series in Synergetics, Springer-Verlag, Berlin, Second edition, 2000.
- [4] E. Morin, *On Complexity*, Hampton Press, New York, NY, USA, 2008.
- [5] G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems*, John Wiley & Sons, USA, 1977.
- [6] L. Von Bertalanffy, *General System Theory: Foundations, Development, Applications*, George Braziller Inc., New York, NY, USA, 1968.
- [7] K. Kaneko and I. Tsuda, *Complex Systems: chaos and beyond*, Springer Verlag: Berlin, Germany, 2001.
- [8] R. López-Ruiz, H. L. Mancini, and X. Calbet, "A statistical measure of complexity," *Physics Letters A*, vol. 209, no. 5-6, pp. 321–326, 1995.
- [9] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, Ili Books Edition, Chicago, USA, 1993.
- [10] D. P. Feldman and J. P. Crutchfield, "Measures of statistical complexity: Why?" *Physics Letters A*, vol. 238, no. 4-5, pp. 244–252, 1998.

- [11] J. S. Shiner, M. Davison, and P. T. Landsberg, “Simple measure for complexity,” *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 59, no. 2, pp. 1459–1464, 1999.
- [12] J. P. Crutchfield, D. P. Feldman, and C. R. Shalizi, “Comment I on “Simple measure for complexity”,” *Physical Review E: Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 62, no. 2, pp. 2996–2997, 2000.
- [13] M. L. Da Silva, J. R. C. Piqueira, and J. M. E. Vieliard, “Using Shannon entropy on measuring the individual variability in the Rufous-bellied thrush *Turdus rufiventris* vocal communication,” *Journal of Theoretical Biology*, vol. 207, no. 1, pp. 57–64, 2000.
- [14] M. Pinho, M. Mazza, J. R. C. Piqueira, and A. C. Roque, “Shannons entropy applied to the analysis of tonotopic reorganization in a computational model of classic conditioning,” *Neurocomputing*, pp. 923–928, 2002.
- [15] S. H. V. L. De Mattos, L. E. Vicente, A. P. Filho, and J. R. C. Piqueira, “Contributions of the complexity paradigm to the understanding of Cerrado’s organization and dynamics,” *Anais da Academia Brasileira de Ciências*, vol. 88, no. 4, pp. 2417–2427, 2016.
- [16] L. P. D. Mortoza and J. R. C. Piqueira, “Measuring complexity in Brazilian economic crises,” *PLOS ONE*, vol. 12, no. 3, Article ID e0173280, 2017.
- [17] J. R. C. Piqueira, V. M. F. De Lima, and C. M. Batistela, “Complexity measures and self-similarity on spreading depression waves,” *Physica A: Statistical Mechanics and Its Applications*, vol. 401, pp. 271–277, 2014.
- [18] Y. C. Campbell-Borges and J. R. C. Piqueira, “Complexity Measure: A Quantum Information Approach,” vol. 10, p. 19, 2012.
- [19] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, Mc Graw Hill, USA, 4th edition, 2002.
- [20] “EMBRAPA (Empresa Brasileira de Produção Agropecuária),” <https://www.embrapa.br/agropecuaria-oeste/biblioteca/acervo>, 2018.

