Contributions to the Study of Time Series and Images with the Entropy-Complexity Plane

Eduarda T. C. Chagas[1], **Heitor S. Ramos**[1], **Alejandro C. Frery**[2],

[1] Departamento de Ciência da Computação, Universidade Federal de Minas Gerais
Brazil

{eduarda.chagas, ramosh}@dcc.ufmg.br

[2]School of Mathematics and Statistics, Victoria University of Wellington
New Zealand

alejandro.frery@vuw.ac.nz

Abstract. *In the context of non-parametric analysis of time series, the use of Ordinal Patterns combined with descriptors of Information Theory proved being powerful in characterizing processes underlying the data dynamics. Two are prominent among those descriptors: Shannon's entropy and Statistical Complexity; together, they define the Entropy-Complexity Plane (HC). Although powerful, this approach suffers from two major shortcomings: (i) there are no statistical tests, and (ii) there is some loss of valuable information when discarding the signal amplitude. This work brings solutions to those problems with (I) empirical tests in the HC plane, and (II) a modification in the transition graph of ordinal patterns, the Weighted Amplitude Transition Graph, which weights its edges using amplitude information. We show applications to white noise analysis, and to discrimination and classification of textures in remotely-sensed images. We also provide the code and data that promote reproducibility and replicability of these results.*

1. Introduction

In recent years we have seen significant growth in the number of intelligent applications involving analysis, data mining, and classification, consequently causing an increase in the diversity and volume of information used. With this, the level of complexity of the investigations, the interdisciplinarity and the number of features necessary to carry out such activities have also increased. Thus, the study of simple approaches, inexpensive computationally and independent of the data type for the extraction and characterization of patterns has become fundamental.

Some requirements are necessary to make good and efficient inferences in data analysis studies, such as (i) make few or no assumptions about the underlying process; and (ii) be simple, fast, and transparent; (iii) be resilient before outliers. When analyzing traditional statistical techniques, we found that they cannot obtain good results without assuming the data's characteristic properties, such as the shape of the probability distribution of the samples. In this context, the analysis of Ordinal Patterns coupled with the use of Information Theory descriptors, in addition to meeting the requirements above, has been able to detect causal information related to the unobserved variables that control the system, in addition to identifying chaotic components, assist in the visualization and characterization of different dynamic regimes, among other applications.

[Bandt and Pompe 2002] proposed the Ordinal Patterns (OPs) as a way to analyse time series. An OP is a mapping of a subset of D values into the sequence of indexes that sorts the observations in, e.g., increasing order. The time series $\boldsymbol{z} = (z_1, z_2, \dots, z_{T+D-1})$ is then transformed into de sequence of patterns $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_D)$. There are $D!$ possible patterns, provided all observations are different. This operation is called "symbolization" and, despite its simplicity, it yields methods which are robust to noise and produce good results in a variety of situations. It is noteworthy that Bandt & Pompe's original paper has received to date more that 2100 citations[1].

The sequence of patterns $\boldsymbol{\pi}$ can then be summarized in two ways: by a *marginal* or by a *transitional* approach. The former forms the histogram of $\boldsymbol{\pi}$, while the latter analyses the transitions (π_i, π_{i+1}), mostly as Complex Networks. Once one of these intermediate representations has been formed, i.e., a histogram or a complex network, information theory descriptors (usually Shannon Entropy and Statistical Complexity) allow us to extract a few quantifiers that, according to the literature, reveal important characteristics about the phenomenon that produced the data.

Although the proposal has twenty years, results on the statistical properties of OPs and derived features are scarce. Most references refer to successful applications, but there is no theory for building statistical tests. With this, their investigation, for example, in conjunction with Artificial Intelligence, is still in its early stages, standing out only in the use of machine learning algorithms to enhance the functionalities in the data characterization/classification processes. Another problem we noticed is that the application of OP to multidimensional signals, e.g., images, is limited.

Under this context, this dissertation advances the state-of-the-art in this field with the following studies and solutions:

HC-PCA: We provide the first confidence regions in the Complexity-Entropy Plane using true random sequences from physical devices.

Testing White Noise in the confidence regions: We present and evaluate a new method of building empirical confidence regions in the Complexity-Entropy plane for the analysis of white noise. We also provide an algorithm that, using geometrical arguments, computes the empirical p-value of a sequence under the white noise null hypothesis.

Weighted Amplitude Transition Graphs: We propose the first approach of transition graphs of weighted ordinal patterns using amplitude information of the analyzed sequences.

Analysis and Classification of SAR Textures using Information Theory: We propose a new representation of textures, which allows a low-dimension characterization useful for, among other applications, their classification. We apply this methodology to the difficult problem of characterizing textures corrupted by speckle noise.

[1] According to Web of Science, May 20, 2022

2. Background

2.1. Bandt-Pompe Symbolization

Consider $\mathcal{X} = \{x_t\}_{t=1}^{T}$, a real valued time series of length T. Let $\mathfrak{A}_D$ (with $D \geq 2$ and $D \in \mathbb{N}$) be the symmetric group of order $D!$ formed by all possible permutations of order D, and the set of unique symbols $\boldsymbol{\pi}^{(D)} = \{\pi_1, \pi_2, \dots, \pi_D\}$. The time delay embedding representation of $\mathcal{X}$ with embedding dimension $D \geq 2$ and time delay $\tau \geq 1$ ($\tau \in \mathbb{N}$, also called "embedding time," "time delay", or "delay") is:

$$\mathbf{X}_t^{(D,\tau)} = (x_t, x_{t+\tau}, \dots, x_{t+(D-1)\tau}), \tag{1}$$

for $t = 1, 2, \dots, N$ with $N = T - (D-1)\tau$. Then, the vector $\mathbf{X}_t^{(D,\tau)}$ can be mapped onto a symbol vector $\boldsymbol{\pi}_t^D \in \mathfrak{A}_D$. This mapping preserves the order relationships between the elements $x_t \in \mathbf{X}_t^{(D,\tau)}$, and all $t \in \{1, \dots, T-(D-1)\tau\}$ that share this pattern (also called "motif") are mapped onto the same $\boldsymbol{\pi}_t^D$. We define the mapping $\mathbf{X}_t^{(D,\tau)} \mapsto \pi_t^D$ by ordering the observations $x_t \in \mathbf{X}_t^{(D,\tau)}$ in increasing order.

The classic (marginal) approach to calculating the probability distribution of ordinal patterns is through the frequency histogram. The Bandt-Pompe probability distribution is the relative frequency of symbols in the series against the $D!$ possible patterns $\{\widetilde{\pi}_t^D\}_{t=1}^{D!}$:

$$p(\widetilde{\pi}_t^D) = \frac{\#\left\{\mathbf{X}_t^{(D,\tau)} \text{ is of type } \widetilde{\pi}_t^D\right\}}{T-(D-1)\tau}, \tag{2}$$

where $t \in \{1, \dots, T-(D-1)\tau\}$. These probabilities meet the conditions $p(\widetilde{\pi}_t^D) \geq 0$ and $\sum_{i=1}^{D!} p(\widetilde{\pi}_t^D) = 1$, are invariant before monotonic transformations of the time series values, and, being based on order statistics, are robust to contamination.

2.2. Information-Theoretic Descriptors

After computing all the symbols and their probabilities, the next step into the characterization of the time series is computing descriptors. Shannon Entropy measures the disorder or unpredictability of a system, and its normalized version is:

$$H(\mathbb{P}) = -\frac{1}{\log D!} \sum_{\ell=1}^{D!} p_\ell \log p_\ell, \tag{3}$$

where p_ℓ is the probability obtained from the symbolization. Although very expressive, the Shannon Entropy is not able to describe all possible underlying dynamics. To this aim, [López-Ruiz et al. 1995] proposed the use of disequilibrium Q, a measure of how far $\mathbb{P}$ is from equilibrium or non-informative distribution $\mathbb{U}$, e.g., the uniform law. We calculate this descriptor as:

$$Q'(\mathbb{P}, \mathbb{U}) = \sum_{\ell=1}^{D!} \left(p_\ell \log \frac{p_\ell}{u_\ell} + u_\ell \log \frac{u_\ell}{p_\ell}\right), \tag{4}$$

and then we normalize it $Q = Q'/\max\{Q'\}$. With this, the Statistical Complexity measures the dependence structures among the elements and is given by $C = H \cdot Q$. We can then map a time series onto the point (h, c), and the set of all possible points is the Entropy-Complexity plane ($H \times C$).

3. Contributions and Results

3.1. Weighted Amplitude Transition Graphs

Texture is an elusive trait. When dealing with remotely sensed images, the texture of different patches carries relevant information that is hard to quantify and transform into useful and parsimonious features. This may be since textures, in this context, is a synesthesia phenomenon that triggers tactile responses from visual inputs. This work presents a new way of extracting features from textures, both natural and resulting from anthropic processes, in SAR (Synthetic Aperture Radar) imagery.

SAR systems are a vital source of data because they provide high-resolution images in almost all weather and day-night conditions. They provide basilar information, complementary to that offered by sensors that operate in other regions of the electromagnetic spectrum, for a variety of Earth Observation applications. Although they present rich information, such data have challenging characteristics. Most notably, they do not follow the usual Gaussian additive model, and the signal-to-noise ratio is usually low.

In our approach, we opt to analyze the 1-D signals, linearizing the image samples using the Hilbert-Peano curve [Lee and Hsueh 1994]. With this approach, we reduce the dimensionality of the data while preserving the spatial correlation structure. Observations are then transformed into ordinal patterns with the Bandt-Pompe symbolization. We use Information Theory descriptors to analyze the distributions these patterns induce, both directly and by building transition graphs among subsequent patterns. Those descriptors are the Shannon entropy and the statistical complexity, which are easy to obtain and are interpretable. They reveal important features of the underlying process.

Our proposal, hereinafter referred to as Weighted Amplitude Transition Graph (WATG), differs from the traditional ordinal pattern transition graph by incorporating the absolute difference between successive patterns. First, each time series $\mathcal{X}$ is scaled to $[0,1]$, since we are interested in a metric able to compare datasets:

$$\frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \longmapsto x_i, \tag{5}$$

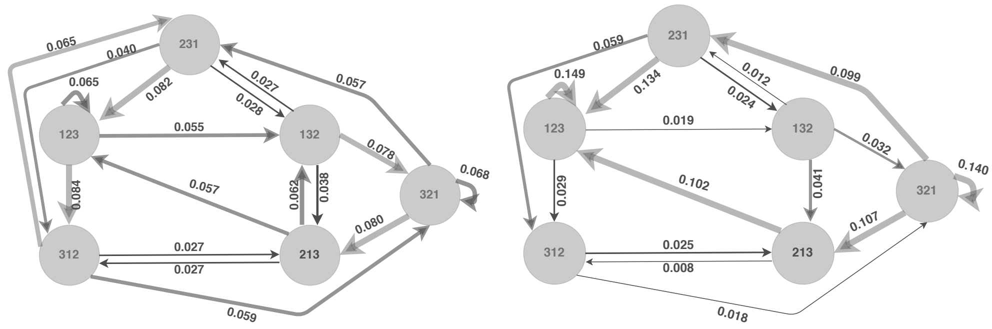
where $x_{\min}$ and $x_{\max}$ are, respectively, the minimum and maximum values of the series. This transformation is relatively stable before contamination, e.g., if instead of $x_{\max}$ we observe $kx_{\max}$ with $k \geq 1$, the relative values are not altered. Nevertheless, other more resistant transformations as, for instance, z scores, might be considered.

Each $\mathbf{X}_t^{(D,\tau)}$ vector is associated with a weight β_t that measures the largest difference between its elements:

$$\beta_t = \max\{|x_i - x_j|\}, \tag{6}$$

where $x_i, x_j \in \mathbf{X}_t^{(D,\tau)}$. We propose that the weight assigned to each edge is proportional to the amplitude difference observed in the transition:

$$w_{v_{\widetilde{\pi}_i^D}, v_{\widetilde{\pi}_j^D}} = \sum_{i:\{\mathbf{X}_t^{(D,\tau)} \mapsto \widetilde{\pi}_i^D\}} \sum_{j:\{\mathbf{X}_t^{(D,\tau)} \mapsto \widetilde{\pi}_j^D\}} |\beta_i - \beta_j|. \tag{7}$$



(a) Transition Graph (b) WATG

Figure 1. Difference of edges weights between the transition graph and the weighted graph of ordinal patterns transitions; urban area, with dimension 3 and delay 1.

Thus, the probability distribution taken from the weighted amplitude transition graph is:

$$\begin{cases} \lambda_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} = 1, \text{ if } (v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}) \in E, \\ \lambda_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} = 0, \text{ otherwise.} \end{cases}, \text{ and} \tag{8}$$

$$p(\tilde{\pi}_i^D, \tilde{\pi}_j^D) = \frac{\lambda_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} \cdot w_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}}}{\sum_{v_{\tilde{\pi}_a^D}, v_{\tilde{\pi}_b^D}} w_{v_{\tilde{\pi}_a^D}, v_{\tilde{\pi}_b^D}}}. \tag{9}$$

Note that $p(\tilde{\pi}_i^D, \tilde{\pi}_j^D) \geq 0$ and $\sum_{\tilde{\pi}_i^D, \tilde{\pi}_j^D} p(\tilde{\pi}_i^D, \tilde{\pi}_j^D) = 1$, so p is a probability function.

To validate our technique in a remote sensing application, we manually selected 200 samples from JPL's Uninhabited Aerial Vehicle SAR (UAVSAR) images patches of size 128×128 to compose the dataset used in the experiments: 40 samples from Guatemalan forests; 40 samples from Guatemalan pasture regions; 80 samples from oceanic regions of Cape Canaveral, divided into two types with different contrast; and 40 samples of urban regions of the city of Munich.

The variation in the magnitude of the targets' backscatter and, consequently, in the intensity of the image pixels, depends on the intrinsic properties of the region under analysis. Urban targets usually exhibit the strongest variation, followed by forest, pasture, forests, and finally, water bodies. By adding such information related to the amplitude, the proposed method is able to increase, compared to traditional methods, the granularity of information captured by ordinal patterns.

As already described, our proposal weights the edges in terms of the difference of amplitudes. The most significant impact is observed in the transition graphs obtained from urban areas, as they present a greater amplitude range between their elements. Fig. 1 shows how this information alters the weights of the transition graph. Notice, in particular, that $(v_{\tilde{\pi}^3_{123}}, v_{\tilde{\pi}^3_{123}})$ almost doubled, while $(v_{\tilde{\pi}^3_{312}}, v_{\tilde{\pi}^3_{231}})$ and $(v_{\tilde{\pi}^3_{213}}, v_{\tilde{\pi}^3_{132}})$ became negligible. We highlight the impact of the weighting on the probability distribution in the two extreme cases observed:

- If the 1-D signal presents a low amplitude variation and intensity peaks between, then the transitions of ordinal patterns that represent the latter have larger weights.

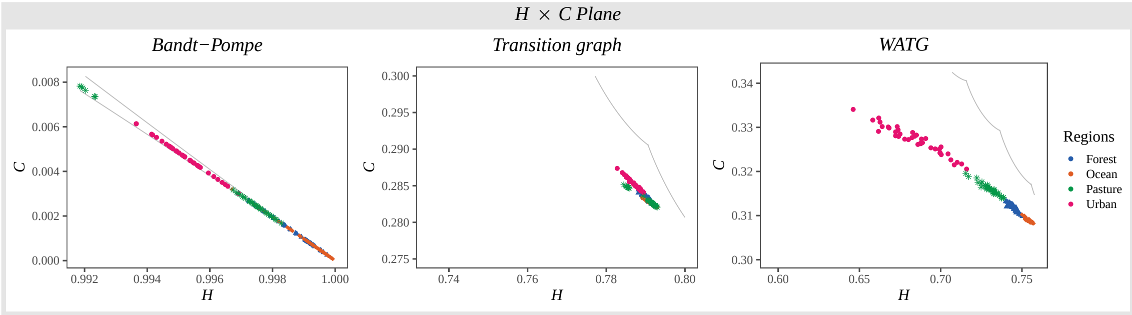


Figure 2. Location of Guatemala (forest), Cape Canaveral (ocean), and Munich (urban) in the $H \times C$ plane for dimension 3 and delay 1. The continuous curves correspond to the maximum and minimum values of C as a function of H.

This contributes so that the probability distribution becomes less uniform among the symbols since it will be more concentrated in these edges. This will also cause a drop in entropy when compared to the traditional method.

- In 1-D signal that shows a uniform amplitude variation, the weights are well distributed between their edges, giving rise to a more random probability distribution, thus obtaining larger entropy.

The Bandt-Pompe symbolization was the first method based on ordinal patterns proposed in the literature. As shown in Fig. 2 left, it provides a limited separation of the textures. Transition graphs (Fig. 2 center) improve the spread of the features, but with some amount of confusion. Our proposal, shown in Fig. 2 right, produces well-separated features. In this way, we were able to obtain, for this experiment, a perfect characterization and, consequently, the high descriptive power of the regions.

Experiments with the k-nearest neighbor algorithm applied to the pairs Entropy-Statistical Complexity descriptors showed that our proposal performs better than Grey-Level Co-occurrence Matrices, Bandt-Pompe, Transition Graphs, SURF (Speeded-Up Robust Features), STFT + SURF (Short-Time Fourier Transform), and other techniques which also employ amplitude information in the analysis of ordinal patterns, and provides the same quality of results obtained with Gabor filters and HOG (Histograms of Oriented Gradients). However, while Gabor filters employ 80 features and HOG uses 54 features, our proposal requires only two. This dimensionality reduction is a huge advantage over the other techniques, with added values: Firstly, by reducing the dimension of the features to 2-D, we can visualize the results. Secondly, for machine learning algorithms, the smaller the number of dimensions is, the faster the training process is, and the less storage space is required. Thirdly, overfitting, a recurring problem in data of high dimensionality, is avoided.

3.2. A Test for White Noise in the Entropy-Complexity Plane

Although the limits of $H \times C$ are well defined, a complete characterization of its intrinsic topology is an open problem, due to the restrictions imposed by its curvilinear space. The lack of knowledge of the joint distribution of the points obtained by this plane, due to the existing correlation between its variables, prevents the studies on test statistics for typical time series in this characterization space. However, with the knowledge of the expected variability of such points, according to the underlying dynamics, we can test hypotheses for a wide variety of models. Results in this direction can be found in the literature. [Larrondo et al. 2006] showed that the Complexity-Entropy plane ($H \times C$) is a good indicator of the results of Diehard tests for pseudo-random number generators. [De Micco et al. 2008] evaluated ways to improve pseudo-random sequences for their representation in this plane.

In this context, a open problem present in the characterization of sequences using the $H \times C$ plane is the absence of a representative metric distance, which makes it difficult to build confidence regions. Thus, in the proposed approach, we opted for the construction of empirical confidence regions obtained through an orthogonal projection of data in the space of principal components. Therefore, the larger is the data set used to build the region, the more representative it will be. In the confidence regions proposed by this work, the white noise hypothesis finds a latent space representation of the data without the restrictions of the plane boundaries. After calculating these regions, we obtain the p-value that a sequence is comprised of independent identically distributed observations.

Our test is based on two sources of true random numbers, both from the observation and measurement of physical phenomena. The first uses vacuum states [Gabriel et al. 2010], and the second employs atmospheric noise captured by a cheap radio receiver with no filter for unwanted static sounds caused by atmospheric noise [Haahr 2018]. We used $54 \times 10^6\,4\mathrm{B}$ words from each physical generator, which approximately amounts to $200\,\mathrm{MB}$ of data. Such wealth of data allowed us to produce $104\,596$ and 2093 independent time series of length 1000 and $50\,000$, respectively. Each time series is mapped onto a point in the Entropy-Complexity plane.

The first step of the proposed technique consists of applying the principal components transformation to the points in the Entropy-Complexity plane $\underset{\sim}{hc}$. With this, we obtain uncorrelated points

$$\underset{\sim}{uv} = \big((u_1, v_1), (u_2, v_2), \dots, (u_N, v_N)\big),$$

in which u_n and v_n are the first and second principal components of h_n and v_n, respectively. This projection allowed us to obtain a "central" point of the data set, around which we will built a rectangular box containing $100\,(1-\alpha)\%$ of the observations. Such box is a variation of the bagplot [Rousseeuw et al. 1999]. Notice that finding the smallest box that encloses k out of N points is difficult; cf. the work by Chan et al. (2020) [Chan and Har-Peled 2020]. For simplicity, and without loss of generality, assume N is odd. We proceeded as follows:

1. Find the indexes that sort the values of the first principal component $\boldsymbol{u} = (u_1, u_2, \dots, u_N)$ in ascending order: $\boldsymbol{r} = (r_1, r_2, \dots, r_N)$, i.e., u_{r_1} is the minimum value, and u_{r_N} is the maximum value.
2. Find the point (u, v) whose first principal component is the median: $(u_{r_{(N+1)/2}}, \cdot)$. Apply the inverse principal components transformation, and obtain $\boldsymbol{P'} = (h', v')$. Call the corresponding time series "emblematic time series."
3. Find the point (u, v) whose first principal component is the quantile $\alpha/2$: $(u_{r_{[N\alpha/2]}}, \cdot)$.
4. Find the point (u, v) whose first principal component is the quantile $1 - \alpha/2$: $(u_{r_{[N(1-\alpha/2)]}}, \cdot)$.

5. The values $u_{r_{[N\alpha/2]}}$ and $u_{r_{[N(1-\alpha/2)]}}$ are the rightmost and leftmost bounds of the box, respectively.
6. The bottom bound of the box is the smallest second principal component value whose first principal component is at least $u_{r_{[N\alpha/2]}}$; denote this values $v_{\min}$.
7. The top bound of the box is the largest second principal value whose first principal component is at most $u_{r_{[N(1-\alpha/2)]}}$; denote this value $v_{\max}$.
8. The corners of the box are $(u_{r_{[N\alpha/2]}}, v_{\min})$, $(u_{r_{[N\alpha/2]}}, v_{\max})$, $(u_{r_{[N(1-\alpha/2)]}}, v_{\min})$ and $(u_{r_{[N(1-\alpha/2)]}}, v_{\max})$.
9. Apply the inverse principal components transformation to these corners obtaining $\boldsymbol{P}_1 = (h_{v_1}, c_{v_1})$, $\boldsymbol{P}_2 = (h_{v_2}, h_{v_2})$, $\boldsymbol{P}_3 = (h_{v_3}, c_{v_3})$ and $\boldsymbol{P}_4 = (h_{v_4}, c_{v_4})$.

The confidence regions obtained provide a powerful tool to make binary assessments about the adequacy of a given time series $\boldsymbol{X}$ to the null hypothesis $\mathcal{H}_0$ that it is white noise. More generally, as we are interested in obtaining the p-value of $\boldsymbol{x}$ under $\mathcal{H}_0$, we devised a procedure to obtain an approximate p-value based on the evidence collected to build the confidence regions.

The procedure operates on the principal components space and consists of measuring the closeness between the "emblematic point" and the observed point. Given the time series $\boldsymbol{x}$ of size T, we want its p-value when contrasted with true white noise random sequences (TWNRS) of the same size at embedding dimension D. We use N TWNRS of size T, compute their points in the $H \times C$ plane, and project them to the corresponding principal components space. We then do the same with $\boldsymbol{x}$, and obtain a new point (u_x, v_x). The closer $\boldsymbol{x}$ is to the emblematic time series, the larger its p-value. Assume that the emblematic time series is represented by (u, v) in the principal components space. We measure this closeness by building a box around (u_x, v_x) that contains (u, v); assume that $u_x > u$, then:

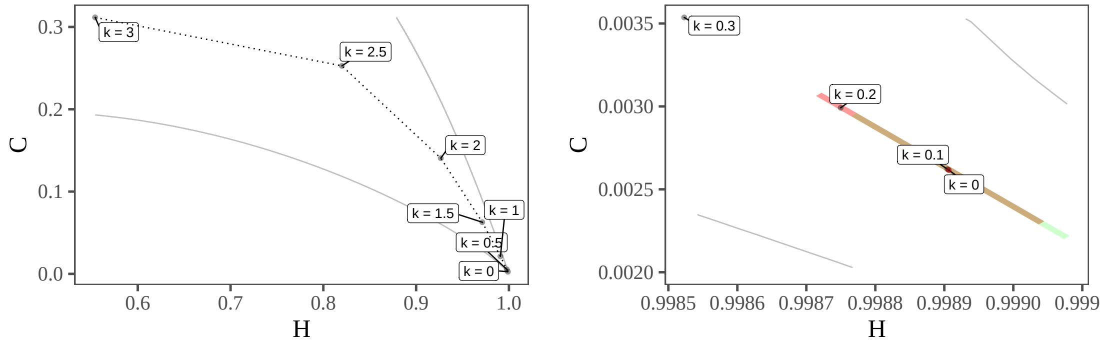
1. the right side of the box is the smallest u_j which is larger that u_x; assume it corresponds to the quantile η_u of $\underset{\sim}{u} = (u_1, u_2, \dots, u_N)$. By definition, $\eta_u \geq 1/2$.
2. the left side of the box is the $1 - \eta_u$ quantile of $\underset{\sim}{u}$.
3. the top side of the box is the smallest v_j which is larger that v_x; assume it corresponds to the quantile η_v of $\underset{\sim}{v} = (v_1, v_2, \dots, v_N)$. By definition, $\eta_v \geq 1/2$.
4. the bottom side of the box is the $1 - \eta_v$ quantile of $\underset{\sim}{v}$.

The definition of the box for the case $u_x < u$ follows naturally. With this approach, we obtain the smallest box that (i) contains the new point, and (ii) it is defined by observed points from TRWNS. Such boxes are less prone to distortions in this space since the distribution of the points becomes less asymmetric than in the Entropy-Complexity plane.

Prior to the study, we conducted an ablation assessment to identify the influence of the parameters T, D, and τ in the construction of empirical confidence regions. We verified that the results involving the time delay parameter variation did not show significant differences in repeated experiments; therefore, in the sequel, we did no consider τ as a determining factor. On the other hand, we found two relevant variables: the length of the sequence and the embedding dimension. We, thus, employed the following factors:

- Sequence length $T \in \mathcal{T} = \{1 \times 10^3, 5 \times 10^4\}$,
- Embedding dimension $D \in \mathcal{D} = \{3, 4, 5, 6\}$.

and kept $\tau = 1$, which is the most frequently used option. The values of D are within the range recommended in the literature [Bandt and Pompe 2002].



(a) Points in the $H \times C$ plane of the emblematic white noise ($k = 0$) and its transformations to become f^{-k} correlated noise with $k = 1/2, 1, 3/2, 2, 5/2, 3$.

(b) Points of the emblematic white noise ($k = 0$), and its f^{-k} correlated noise versions, with $k = 1/10, 1/5, 3/10$ along with the confidence regions for white noise.

Figure 3. Analysis of the test power with correlated f^{-k} noise.

We assessed the power of the test by contrasting time series with different correlation structure (under the f^{-k} model) in the $H \times C$ plane. Our study's basis is the emblematic time series for each length T and dimension embedding D. Recall that the emblematic time was chosen as the most representative of the data set. We use these series, transform them into f^{-k} correlated noise, and verify the new point's location in the $H \times C$ plane.

As we can observe in the plane, as the correlation between the observations increases, that is, $k > 0$, the randomness decreases, and the entropy presented decreases, informing the loss of its stochastic characteristic. Fig. 3(a) shows the overall effect of transforming the emblematic time series into f^{-k} correlated noise, with $k = 1/2, 1, 3/2, 2, 5/2, 3$. At this scale, the emblematic time series $k = 0$ and the one with $k = 1/2$ appear overlapped. As the correlation increases with k, the randomness decreases, causing a drop in the entropy; the series become progressively more predictable. Fig. 3(b) is a zoom close to the $(1, 0)$ point, along with the confidence regions for the white noise. We see that $k = 0$ and $k = 0.1$ are inside the $95\,\%$ confidence region, and $k = 0.2$ is inside the $99\,\%$ box. Notice that the time series with $k = 3/10$ is outside the confidence regions and does not pass the randomness test. The same holds for all $k > 3/10$.

4. Conclusions

The main objective of this work was the investigation of open problems in the Bandt-Pompe methodology of symbolization and their application to the characterization of time series and images. Interested in expanding the range of possible applications, we focused on investigating properties of transition graphs and their limitations. Another objective was the study of the joint distribution of descriptors in the Complexity-Entropy plane, as well as possible linear transformations in this space. This technique proved to be fast, consistent and robust the addition of correlation structures. Thus, we have advanced in the state of the art by proposing innovative and relevant solutions to deal with scenarios unforeseen in the seminal article by Bandt and Pompe.

This work presents several possibilities for future research. For example, the use of WATG can be explored in different application scenarios. Considering that its main characteristic consists of discriminating sequences with variations in amplitude along with the arrangement of its elements, its applicability is not restricted to remote sensing images. In the context of SAR images, modifications can be made to increase the generalizability of the technique. On the other hand, under the context of confidence regions, our work opens up a huge range of related research. The study of regression models on correlated descriptors and the development of specific kernels for the Complexity-Entropy plane are fruitful possibilities for investigation. We also emphasize the need for efforts to build representative metrics. With the advancement of deep metric learning techniques [Barros et al. 2020], we can explore the learning of projections in a linear transformation specific to the plane, which would allow progress to build specific machine learning algorithms for the Complexity-Entropy space.

References

Bandt, C. and Pompe, B. (2002). Permutation entropy: A natural complexity measure for time series. *Physical Review Letters*, 88:174102–1–174102–4.

Barros, P. H., Queiroz, F., Figueredo, F., dos Santos, J. A., and Ramos, H. S. (2020). A new similarity space tailored for supervised deep metric learning.

Chan, T. M. and Har-Peled, S. (2020). Smallest k-enclosing rectangle revisited. *Discrete & Computational Geometry*.

De Micco, L., González, C. M., Larrondo, H. A., Martin, M. T., Plastino, A., and Rosso, O. A. (2008). Randomizing nonlinear maps via symbolic dynamics. *Physica A: Statistical Mechanics and its Applications*, 387(14):3373–3383.

Gabriel, C., Wittmann, C., Sych, D., Dong, R., Mauerer, W., Andersen, U. L., Marquardt, C., and Leuchs, G. (2010). A generator for unique quantum random numbers based on vacuum states. *Nature Photonics*, 4(10):711–715.

Haahr, M. (1998–2018). RANDOM.ORG: true random number service. `https://www.random.org`. Accessed: 2018-06-01.

Larrondo, H. A., Martín, M. T., González, C. M., Plastino, A., and Rosso, O. A. (2006). Random number generators and causality. *Physics Letters A*, 352(4–5):421–425.

Lee, J.-H. and Hsueh, Y.-C. (1994). Texture classification method using multiple space filling curves. *Pattern Recognit. Lett.*, 15(12):1241–1244.

López-Ruiz, R., Mancini, H., and Calbet, X. (1995). A statistical measure of complexity. *Physics Letters A*, 209(5-6):321–326.

Rousseeuw, P. J., Ruts, I., and Tukey, J. W. (1999). The bagplot: A bivariate boxplot. *The American Statistician*, 53(4):382.