

**CONTRIBUTIONS TO THE STUDY OF TIME  
SERIES AND IMAGES WITH THE  
ENTROPY-COMPLEXITY PLANE**

EDUARDA TATIANE CAETANO CHAGAS

**CONTRIBUTIONS TO THE STUDY OF TIME  
SERIES AND IMAGES WITH THE  
ENTROPY-COMPLEXITY PLANE**

Dissertação apresentada ao Programa de Pós-Graduação em Computer Science do Instituto de Ciências Exatas da Federal University of Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Computer Science.

ORIENTADOR: HEITOR SOARES RAMOS FILHO  
COORIENTADOR: ALEJANDRO CESAR FRERY ORGAMBIDE

Belo Horizonte

Março de 2021

EDUARDA TATIANE CAETANO CHAGAS

CONTRIBUTIONS TO THE STUDY OF TIME  
SERIES AND IMAGES WITH THE  
ENTROPY-COMPLEXITY PLANE

Dissertation presented to the Graduate Program in Computer Science of the Federal University of Minas Gerais in partial fulfillment of the requirements for the degree of Master in Computer Science.

ADVISOR: HEITOR SOARES RAMOS FILHO  
CO-ADVISOR: ALEJANDRO CESAR FRERY ORGAMBIDE

Belo Horizonte

March 2021

Chagas, Eduarda Tatiane Caetano.

C433c Contributions to the study of time series and images with the entropy-complexity plane [manuscrito] / Eduarda Tatiane Caetano Chagas – 2021.  
xxv, 76 f. il.

Orientador: Heitor Soares Ramos Filho

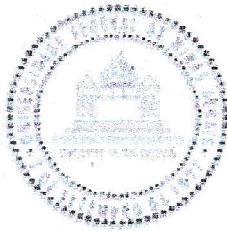
Coorientador: Alejandro Cesar Frery Orgambide.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Ciência da Computação.

Referências: f.67-76

1. Computação – Teses. 2. Teoria da informação – Teses. 3. Entropia (Teoria da informação) – Teses. 4. Estatística não paramétrica – Teses. 5. Análise de séries temporais – Teses. I. Ramos Filho, Heitor Soares. II. Frery Orgambide, Alejandro Cesar. III. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Ciência da Computação. IV. Título.

CDU 519.6\*71(043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

## FOLHA DE APROVAÇÃO

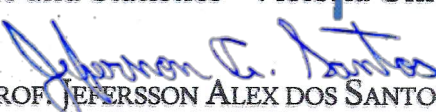
Contributions to the Study of Time Series and Images with the Entropy-  
Complexity Plane

**EDUARDA TATIANE CAETANO CHAGAS**

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

  
PROF. HEITOR SOARES RAMOS FILHO - Orientador  
Departamento de Ciência da Computação - UFMG

  
PROF. ALEJANDRO CÉSAR FRERY ORGAMBIDE - Coorientador  
School of Mathematics and Statistics - Victoria University of Wellington

  
PROF. JEFFERSSON ALEX DOS SANTOS  
Departamento de Ciência da Computação - UFMG

  
PROFA JULIANA GAMBINI  
Departamento de Ingeniería Informática - Instituto Tecnológico de Buenos Aires

Belo Horizonte, 9 de Março de 2021.



# Acknowledgments

I am most grateful to my grandparents, Dimas and Sileide, who were present at all stages of this important moment in my trajectory. I am grateful not only for all the sacrifices made to provide me with a solid education, but also for all the support and encouragement to overcome the various challenges that have surfaced over the past two years. You are my greatest inspiration in life.

I would also like to offer a special thanks to my fiancée, Juliana. Thank you for being part of my life and especially for being by my side during this important moment. Her company, love and patience were essential in this work. Her smile was and is my biggest motivation. Thanks for everything.

I also take the opportunity to register my sincere thanks to my advisor Prof. Dr. Heitor Ramos. I am extremely grateful for your orientation, advice and all the opportunities offered. I feel that thanks to you I have evolved as a student and researcher in those years.

I would also like to thank my co-advisor and mentor Prof. Dr. Alejandro Frery. I feel extremely lucky after leaving each of our conversations and meetings, where even from a distance, I feel motivated with his love for science. I deeply hope that our current partnership is only the beginning of a great history of friendship.

To my friends, Glauber Braga, Bruno Rafael and David Alexandre, who endured all my crises and outbursts and were present in the best and worst moments of the last few years. Love you!

To my new "brother friend", Pedro Barros, who embarked on this journey with me from my hometown, Maceió. Thank you for all of our valued conversations and scientific debates present at Laccan House every day.

I would like to thank the professors André Aquino, André Lages, Osvaldo Rosso and Eliana Almeida for having believed on my academic potential. The support provided by you was crucial. Thank you!

Finally, and no less important, I would like to thank the examining board for their attentive reading, questions and suggestions.

*“Não há alegria mais intensa que a de encontrar um fato que não pode ser entendido nos termos das ideias atualmente aceitas”*

*(Cecilia Payne-Gaposchkin)*



# Resumo

Nos últimos anos observamos um crescimento expressivo no número de aplicações inteligentes envolvendo análise, mineração e classificação de dados. Com o aumento da complexidade das investigações a necessidade de abordagens simples, rápidas e com baixo custo computacional tornou-se fundamental. No contexto de análise não-paramétrica de séries temporais, o uso da metodologia de simbolização de Bandt-Pompe tornou-se relevante. Tendo como base o uso de padrões ordinais formados por meio dos elementos da série analisada, quando unido ao uso de descritores causais da teoria da informação mostrou-se apresentar um alto poder de caracterização da dinâmica geradora do processo subjacente aos dados.

Dentre os descritores, dois destes por apresentarem definições complementares vem recebendo um grande destaque na literatura: a entropia de Shannon, que neste contexto mensura o grau de desordem da distribuição dos padrões ordinais e a complexidade estatística, que por outro lado, representa o grau de dependência estrutural entre os elementos da sequência. Em conjunto, tais features formam o plano Complexidade-Entropia, cujo o presente trabalho possui como objetivo evidenciar as suas principais lacunas, são elas: (i) a ausência de métodos para construção de regiões de confiança e (ii) a ambiguidade na formação dos símbolos provocada pela ausência de informações da amplitude de seus elementos. Visando apresentar métodos alternativos para os problemas relatados, propomos duas soluções: uma modificação no grafo de transição de padrões ordinais, o Weighted Amplitude Transition Graph, que realiza o cálculo do peso de suas arestas usando informações de variação de amplitude entre os símbolos, e o HC-PCA, um método de geração de regiões de confiança empíricas sobre o plano. Para validar nossas propostas, aplicações no contexto de sensoriamento remoto e análise de sequências de ruídos brancos foram desenvolvidas.

**Palavras-chave:** Simbolização de Bandt-Pompe, Padrões ordinais, Plano Complexidade-Entropia, Teoria da Informação.

# Abstract

In recent years we have seen significant growth in the number of intelligent applications involving analysis, data mining, and classification. With the increase in the complexity of the investigations, the need for simple, fast, and low computational approaches has become essential. In the context of non-parametric analysis of time series, the use of the Bandt-Pompe symbolization methodology has become relevant. The use of ordinal patterns formed by time-series elements when combined with the use of information theory descriptors proved to have a high power of characterization of the process underlying the dynamics of the data.

Among the descriptors, two of these for presenting complementary definitions have received a great prominence in the literature: Shannon's entropy, which in this context measures the degree of disorder in the distribution of ordinal patterns formed through the time series, and the statistical complexity, which on the other hand, represents the degree of structural dependence between the elements of the sequence. Together, these features form the Complexity-Entropy plane, whose present work aims to highlight and solve its main gaps: (i) the absence of methods to build confidence regions and (ii) the ambiguity in the formation of symbols caused by the lack of information on the amplitude of the elements. In order to present alternative methods for the reported problems, we propose two solutions: a modification in the transition graph of ordinal patterns, the Weighted Amplitude Transition Graph, which performs the calculation of the weight of its edges using amplitude variation information between the symbols, and the HC-PCA, a method of generating empirical confidence regions on the plane. To validate our proposals, applications in the context of remote sensing and analysis of white noise sequences were developed.

**Keywords::** Bandt-Pompe Symbolization, Ordinal Patterns, Complexity-entropy Plane, Information theory.

# List of Figures

2.1	Illustration of the Bandt and Pompe coding. Here the gray line represents the analyzed sequence $\mathcal{X} = (1.8, 1.2, 3.2, 4.8, 4.2, 4.5, 2.3, 3.7, 1.2, 0.5)$ , the sequence illustrated by the dotted line shows the path taken when applying $\tau = 2$ , and the sequence illustrated by the black line shows the elements of the pattern $\mathbf{X}_1^{(5,2)} = (1.8, 3.2, 4.2, 2.3, 1.2)$ . . . . .	10
2.2	<b>Rank permutation mapping:</b> The complete alphabet for $D = 3$ of the rank mapping technique obtained by permuting all possible ranks. . . . .	11
2.3	<b>Chronological index permutation mapping:</b> The complete alphabet for $D = 3$ of the chronological mapping technique obtained by permuting all of its indexes. . . . .	12
2.4	Boundaries of the $H \times C$ plane for $D = 3, 4, 5$ . . . . .	16
2.5	Patterns histograms of selected time series with $D = 6$ , $\tau = 1$ and $T = 10^4$ . . . . .	18
2.6	Eleven systems and their points in the $H \times C$ plane when we apply $D = 6$ , $\tau = 1$ and $T = 10^4$ . . . . .	19
3.1	Two examples of possible $D$ -dimensional vectors corresponding to the same ordinal patterns when using $D = 3$ . . . . .	22
3.2	Schematic of the composition of the weighted amplitude graph (WATG). . . . .	26
4.1	Hilbert-Peano curves in areas of: (a) $8 \times 8$ , (b) $16 \times 16$ , and (c) $32 \times 32$ pixels. . . . .	31
4.2	Outline of the methodology used for the classification of textures. . . . .	32
4.3	Types of regions (Guatemala forest, Canaveral ocean types 1 and 2, Munich urban area, and pasture area) and their signal representation. . . . .	34
4.4	Ground truth, speckled, and speckled transposed versions. . . . .	35
4.5	Modifications to the $H \times C$ Plane features by adding different multiplicative noises. . . . .	36

4.6	Difference of edges weights between the transition graph and the weighted graph of ordinal patterns transitions; urban area, with dimension 3 and delay 1. . . . .	37
4.7	Evaluation of the sliding window parameters using ROC curve . . . . .	38
4.8	Characterization resulting in $H \times C$ Plane from the application of the Hilbert-Peano curve in WATG on textures of different regions: Guatemala (forest), Cape Canaveral (ocean) and Munich (urban). The continuous curves correspond to the maximum and minimum values of $C$ as a function of $H$ . . . . .	39
4.9	Bi-dimensional visualization of the descriptors obtained by our proposal and other methods of texture characterization and classification. For those who provide more than two features, we apply the PCA algorithm to obtain such a visualization. . . . .	41
5.1	Outline of the methodology used for the construction of the confidence regions.	49
5.2	Outline of the methodology used to calculate the $p$ -value. The new point is denoted as a crossed circle. . . . .	53
6.1	White noise samples considered during the construction of the proposed confidence regions. . . . .	58
6.2	Representation of true random white noise sequences of length $T = 50000$ in the PCA space for $D = 3$ and $D = 6$ , and the quantiles of 90 %, 95 %, 99 %, and 99.9 %. . . . .	59
6.3	Histograms of the first principal component for $D = 3, 4, 5, 6$ . . . . .	60
6.4	Results of the analysis behavior of true random noises in the regions of confidence built. . . . .	62
6.5	Analysis of the test power with correlated $f^{-k}$ noise. . . . .	63

# List of Tables

4.1	Experimental results using $k$ -NN . . . . .	42
5.1	Result of the main works of white noise sequences analysis in the $H \times C$ plane. . . . .	47
6.1	Coordinates in the $H \times C$ plane of the emblematic series and the points that define the confidence regions at 90 %, 95 %, 99 %, and 99.9 % for $D = 3, 4, 5, 6$ and $N = 1000$ . . . . .	60
6.2	Coordinates in the $H \times C$ plane of the emblematic series and the points that define the confidence regions at 90 %, 95 %, 99 %, and 99.9 % for $D = 3, 4, 5, 6$ and $N = 50000$ . . . . .	61
6.3	Results of the sequences generated by the main PRNGs in the literature. The sequences have length $T = 5 \times 10^4$ . . . . .	64

# Contents

Acknowledgments	xi
Resumo	xv
Abstract	xvii
List of Figures	xix
List of Tables	xxi
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Goals . . . . .	3
1.3 Contributions . . . . .	4
1.4 Publications . . . . .	4
1.5 Work Organization . . . . .	6
<b>2 Bandt-Pompe Symbolization: Concepts and definitions</b>	<b>7</b>
2.1 Ordinal Patterns Representation . . . . .	8
2.1.1 Rank Permutation Mapping . . . . .	10
2.1.2 Chronological Index Permutation Mapping . . . . .	11
2.2 Ordinal Patterns Transition Graphs . . . . .	12
2.3 Information-Theoretic Descriptors . . . . .	14
2.3.1 Permutation Entropy . . . . .	14
2.3.2 Statistical Complexity . . . . .	14
2.3.3 Causality Complexity-Entropy Plane . . . . .	15
<b>3 Weighted Amplitude Transition Graph</b>	<b>21</b>
3.1 Weighted Ordinal Patterns Methods in Literature . . . . .	22
3.1.1 Weighted Permutation Entropy . . . . .	22

3.1.2	Fine-Grained Permutation Entropy . . . . .	23
3.1.3	Amplitude-Aware Permutation Entropy . . . . .	24
3.2	Proposed Method . . . . .	24
<b>4</b>	<b>Case study: Classification of SAR Textures</b>	<b>27</b>
4.1	Introduction . . . . .	27
4.2	Linearization of image patches . . . . .	30
4.3	Methodology . . . . .	31
4.4	Image Dataset . . . . .	33
4.5	Experimental Results and Analysis . . . . .	34
4.5.1	Properties . . . . .	34
4.5.2	Analysis of ordinal patterns methods . . . . .	35
4.5.3	Experiments on sliding window selection . . . . .	37
4.5.4	Quantitative Evaluation . . . . .	38
4.6	Conclusions . . . . .	43
4.7	Reproducibility and Replicability . . . . .	44
<b>5</b>	<b>HC-PCA Confidence Regions</b>	<b>45</b>
5.1	Complexity-Entropy plane in the literature . . . . .	46
5.2	Proposed Method . . . . .	47
5.2.1	Overall Framework . . . . .	47
5.2.2	Empirical Confidence Regions and $p$ -values . . . . .	48
<b>6</b>	<b>Case study: White Noise Confidence Regions</b>	<b>55</b>
6.1	Introduction . . . . .	55
6.2	Experimental Settings . . . . .	56
6.2.1	Parameters Settings and Dataset . . . . .	56
6.2.2	True Random Numbers . . . . .	57
6.3	Experimental Results and Analysis . . . . .	59
6.3.1	Descriptive analysis of empirical confidence regions . . . . .	59
6.3.2	Test Size . . . . .	60
6.3.3	Test Power . . . . .	61
6.3.4	Revisiting the White Noise Hypothesis in the Literature . . . . .	63
6.4	Conclusions . . . . .	63
6.5	Reproducibility and Replicability . . . . .	64
<b>7</b>	<b>Conclusions and Future Work</b>	<b>65</b>





# Chapter 1

## Introduction

### 1.1 Motivation

In the last decades we have dealt with a drastic increase in the number of applications coming from data mining, 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 were also increased. Thus, the study of simple approaches, inexpensive computationally and independent of the type of data for the extraction and characterization of patterns has become fundamental.

One of the fields of study in this area is the application of information theory descriptors. The information theory developed by Claude Shannon emerged as an interdisciplinary branch, producing countless results both in the theoretical point of view and in the applications in information extraction in signals, covering in its solutions concepts present in Probability, Statistics and Telecommunications. It currently consists of a powerful tool for the quantification of different levels of order and complexity present in the processes that generate the data.

In the context of non-parametric analysis of time series, a new methodology was proposed by Bandt and Pompe (2002a) for data analysis. Obtaining a representation of the time series in Bandt-Pompe ordinal patterns, two approaches are currently being applied to acquire a non-parametric probability distribution of the data: the use of frequency histograms or transition graphs. When opting for this new representation of the data, the resulting distribution becomes less sensitive to outliers and, as it does not depend on any model, it can be applied to a variety of situations. Despite its simplicity, this method is robust to noise and yields good results in assessing the randomness of a sequence, characterization and classification of signals.

The Bandt-Pompe methodology and its variants have been used successfully in the analysis of many types of dynamics, receiving so far more than 1800 citations, according to the Web of Science. We found works using this approach in several areas of scientific knowledge such as, for example: distinguishing noise from chaos (Rosso et al., 2007a); the study of electroencephalography signals using wavelet decomposition (Baravalle et al., 2018a,b); description of El Niño/Southern Oscillation during the Holocene (Saco et al., 2010); the characterization of household appliances through their energy consumption (Aquino et al., 2017); detecting and quantifying stochastic and coherence resonance (Rosso and Masoller, 2009a,b); analysis and characterization of economic time series, e.g., stock market, sovereign bonds, credit rating, commodities, and cryptocurrencies (Zunino et al., 2010, 2012a; Bariviera et al., 2013, 2018; de Araujo et al., 2019); online signature classification and verification (Rosso et al., 2016). Schieber et al. (2016) verified the effect of attacks on complex networks by the displacement of points in the  $H \times C$  plane. Aquino et al. (2015) described vehicles' behavior depending on the topology of cities, and Chagas et al. (2020) succeeded in expanding the use of such techniques for analyzing textured images corrupted by speckle noise. Bariviera et al. (2015) identified spurious interventions in the Libor market using the  $H \times C$  plane representation. Echegoyen et al. (2020) were able to discriminate between individuals with mild cognitive impairment from those diagnosed with Alzheimer's disease using magnetoencephalography recordings.

Given the probability distribution of the patterns, each time series is then described by a point in the  $\mathbb{R}^2$  subinterval, the Complexity-Entropy plane. Two points are well known in this plane:

1. **White noises**, that is, random sequences without any spatial structure, where the entropy presents its maximum value while the complexity is minimal - we can describe statistically the systems present in such locations of the plane by a random variable taken from some distribution; and
2. **Deterministic data**, that is, sequences with a periodic structure, where entropy and statistical complexity have their minimum values - we can reconstruct the patterns of such data with only a small portion of patterns.

Through these references, we can characterize time series according to the dynamics of their generation process. Studies with different applications have managed to obtain relevant results from time series through information on the nature of the data provided by the  $H \times C$  plane. Examples include: (i) Echegoyen et al. (2020) analysis of permutation in magnetoencephalography recordings of individuals suffering

from mild cognitive impairment and individuals diagnosed with Alzheimer’s disease by trajectories in the  $H \times C$  plane, (ii) the study of Schieber et al. (2016) of the effect of attacks on complex networks by shifting their points on the  $H \times C$  plane, and (iii) the description of vehicle behavior based on the topology of cities carried out by Aquino et al. (2015).

However, we can see from the examples above that Bandt-Pompe symbolization has a wider range of applications for one-dimensional signals, such as time series. Thus, it still presents challenges in applications with higher dimensions, such as reducing the loss of spatial correlation between the pixels of an analyzed image.

The aforementioned studies illustrate the importance of the symbolization method in the most diverse areas of knowledge. This dissertation works in this context and advances in the state of the art with a focus on investigating the main problems still present in the literature: (i) the application of the Bandt-Pompe approach in images, (ii) the absence of amplitude weighting methods in transition graphs of ordinal patterns, and (iii) the lack of proposals to build confidence regions.

By proposing an expansion of the applicability of Bandt-Pompe in the image context, this work presents a new perspective of extracting texture characteristics from SAR (Synthetic Aperture Radar) images. In this way, by linearizing the image samples, we proposed the use of time series analysis tools for a new context: where exists dependence and spatial correlation between the elements.

## 1.2 Goals

The aim of this work is to develop solutions for the main gaps in the Bandt-Pompe symbolization methodology. Thus, we propose two solutions: the first proposal to incorporate amplitude information in transition graphs and the first approach to build empirical confidence regions in the Complexity-Entropy plane. We started the work by presenting a modification of the traditional transition graphs of ordinal patterns, which when applied to the context of analysis of SAR image textures proved to be the first approach to be able to characterize and classify images with results directly comparable to the techniques present in the state of the art. On the other hand, one of the major problems involving the direct use of causal descriptors, Shannon entropy and statistical complexity, for classification activities is the lack of an appropriate distance metric. Due to the curvilinear shape of the plane, we verified the need for work focused on building confidence regions and test statistics for such tooling. Thus, to minimize the impacts caused by the absence of a specific metric, we propose the HC-PCA. Based

on this study, we were able to increase the power of characterization of different classes of time series, where as a use case, we present how the proposed approach manages to determine regions of pure randomness present in the plane.

## 1.3 Contributions

The main contributions of this dissertation work are:

**Weighted Amplitude Transition Graph:** We propose the first approach of transition graphs of weighted ordinal patterns using amplitude information of the analyzed sequences. In this way, we were able to reduce the ordinal ambiguity present in transition graphs, thus increasing its characterization power.

**Analysis and Classification of SAR Textures using Information Theory:** Through WATG, we propose a new representation of SAR textures, which allows a low-dimension characterization useful for, among other applications, its classification. Our approach is robust for rotations and the presence of speckle noise. In addition to perfect separation between urban areas, pastures, ocean and forest, the proposed descriptors are interpretable in terms of the degree and structure of the spatial dependence between the observations.

**HC-PCA:** We provide the first contribution in the construction of confidence regions in the Complexity-Entropy Plane, and to measure the similarity of new data sequences with the empirical points we propose the construction of a new test statistic.

**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 analysis of white noise. We were able to capture the random behavior of short sequences of PRNGs already analyzed in the literature and we found that in the scenario presented, our technique is robust to correlation structures.

## 1.4 Publications

The first year of the master's degree was dedicated to research an extensive literature review work, where a comprehensive study was review focusing on the classification of SAR textures, alternatives for using Bandt-Pompe symbolization in two-dimensional

data, and ordinal pattern weighting approaches by amplitude. Once have the state-of-the-art, we use the knowledge of several fields of research, such as machine learning, information theory, and graphs to prepare our first proposal, which resulted in some works listed below in chronological order of publication/presentation:

- WATG: Incorporating amplitude in ordinal pattern decomposition for time series analysis. Poster presentation at Khipu, Nov 2019, Montevideo, Uruguay. Latin American Meeting In Artificial Intelligence, 2019.
- Characterization Of SAR Images With Weighted Amplitude Transition Graphs. 2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LA-GIRS).
- Analysis and Classification of SAR Textures using Information Theory. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2020).
- An exploratory study of the transition graphs application in different resolutions and polarizations of SAR images. 2021 China International SAR Symposium (CISS). (In writing and submission process).

During the second year, we plan to focus on developing our proposal for building confidence regions in the Complexity-Entropy plane. Although they presented unlivable results, detailed studies were carried out involving the use of classical bi-variate analysis, linear regression, and generalized linear models. After defining the methodology, we focus on writing the article below that is currently in the submission process:

- Confidence Regions for Information-Theoretic Descriptors of Time Series. International Statistical Review.

A work related to the general area studied was also carried out in collaboration with other researchers:

- Detecção de eventos no Twitter através de Grafos de visibilidade natural. III Workshop de Computação Urbana. SBC, 2019.
- Supervised Distance Metric learning Encoder with Similarity Space for malware classification through image representation. Computer Networks (Under submission).

## 1.5 Work Organization

This document is organized as follows. Chapter 2 introduces concepts and definitions of the Bandt-Pompe symbolization process and the Complexity-Entropy plane. Chapter 3 provides an overview of ordinal pattern weighting techniques. We also propose the first algorithm for incorporating amplitude information into transition graphs, the WATG. Chapter 4 presents a use case for WATG, providing the experimental results obtained in characterizing and classifying homogeneous textures of SAR images. Chapter 5 discusses the state of the art analysis of white noise confidence regions in the plane. Here we also propose the HC-PCA region of empirical confidence and the construction of a specific statistical test. Chapter 6 presents the main results obtained in our case study with white noise samples. Finally, in Chapter 7, we provide our final thoughts, a perspective on this work, and directions for future work.

## Chapter 2

# Bandt-Pompe Symbolization: Concepts and definitions

Dynamic systems describe the relationship in time of a point in a geometric space. Thus, the study of time series assists in the analysis of the dynamics of its generating processes. Whether a scalar time series collected through the observation of natural phenomena or obtained through synthetic simulations, its values will be the result of a function formed by its majority of unobserved and measured variables. Thus, an important research question in data analysis has been:

Given a system and a sample resulting from it whose evolution can be tracked over time, how much information is encoded in this observable about the dynamics of the underlying system and the variables that characterize them?

Traditionally, the study of time series is carried out under two lines of study, in the domains of time and frequency (Brockwell and Davis, 1991). However, both approaches directly use the data resulting from the observational process, which are sensitive to noise and contamination effects. The approach to the use of non-parametric methods appears in the literature as a way to prevent such effects from compromising the analyzes.

To make good inferences in data analysis studies, it is expected that the proposed approach meets the following requirements:

- Be simple, fast and transparent;
- Make little or no assumptions about the underlying process; and
- Be resilient towards outliers.

When analyzing classic probabilistic techniques, we found that they cannot obtain good results without assuming characteristic properties of the data, such as the shape of the probability distribution of the samples. On the other hand, proposals based on machine learning do not mean to guarantee the required simplicity and transparency, as they do not provide a clear view of the observed characteristics (Bandt, 2019).

The 1820 citations received by the seminal paper appeared in 684 journals indexed by the Web of Science. These journals belong to 127 categories, spanning from Multidisciplinary Physics (24% of the publications) to Zoology (only one of the citing articles). There are 17 citing articles from journals that belong to the Statistics & Probability category. Five of these articles appeared in *Stochastic Environmental Research and Risk Assessment*, two in the *Journal of Time Series Analysis*, and each of the remaining ten appeared in a different journal. Most of these articles relate successful applications of the Bandt and Pompe methodology, except Sinn and Keller (2011) that obtained the sample entropy's properties under zero-mean Gaussian processes. It is also noteworthy that, in this category of publications, Abrams et al. (2013) provided a formal and more general proof of the structure of the boundary of the  $H \times C$  manifold than that obtained by Martín et al. (2006). The lack of attention that the Bandt and Pompe approach has received by the Probability & Statistics community confirms that it is a fertile research avenue waiting to explore.

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, visualization and characterization of different dynamic regimes, among other applications.

## 2.1 Ordinal Patterns Representation

Bandt and Pompe (2002b) introduced the representation of time series by ordinal patterns as a transformation resistant to noise, and invariant to nonlinear monotonic transformations. Let  $\mathcal{X} \equiv \{x_t\}_{t=1}^T$  be a real valued time series of length  $T$ , without ties. As stated by Bandt and Pompe (2002b) in their seminal work:

"If the  $\{x_t\}_{t=1}^T$  attain infinitely many values, it is common to replace them by a symbol sequence  $\Pi \equiv \{\pi_j\}$  with finitely many symbols, and calculate source entropy from it".

Also, as stressed by these authors,



“The corresponding symbol sequence must come naturally from the  $\{x_t\}_{t=1}^T$  without former model assumptions”.

Let  $\mathfrak{A}_D$  (with  $D \geq 2$  and  $D \in \mathbb{N}$ ) be the symmetric group of order  $D!$  formed by all possible permutation of order  $D$ , and the symbol component vector  $\boldsymbol{\pi}^{(D)} = (\pi_1, \pi_2, \dots, \pi_D)$  so every element  $\boldsymbol{\pi}^{(D)}$  is unique ( $\pi_j \neq \pi_k$  for every  $j \neq k$ ). Consider for the time series  $\mathcal{X} \equiv \{x_t\}_{t=1}^T$  its time delay embedding representation, with embedding dimension  $D \geq 2$  and time delay  $\tau \geq 1$  ( $\tau \in \mathbb{N}$ , also called “embedding time,” “time delay,” or “delay”):

$$\mathbf{X}_t^{(D,\tau)} = (x_t, x_{t+\tau}, \dots, x_{t+(D-1)\tau}), \quad (2.1)$$

for  $t = 1, 2, \dots, N$  with  $N = T - (D - 1)\tau$ . Then the vector  $\mathbf{X}_t^{(D,\tau)}$  can be mapped to a symbol vector  $\tilde{\boldsymbol{\pi}}_t^D \in \mathfrak{A}_D$ . This mapping is such that preserves the desired relation 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 to the same  $\tilde{\boldsymbol{\pi}}_t^D$ .

We define the mapping  $\mathbf{X}_t^{(D,\tau)} \mapsto \tilde{\boldsymbol{\pi}}_t^D$  by ordering the observations  $x_t \in \mathbf{X}_t^{(D,\tau)}$  in increasing order. Consider the time series:

$$\mathcal{X} = (1.8, 1.2, 3.2, 4.8, 4.2, 4.5, 2.3, 3.7, 1.2, 0.5) \quad (2.2)$$

depicted in Fig. 2.1. Assume we are using patterns of length  $D = 5$  with unitary time lag  $\tau = 1$ . The code associated to  $\mathbf{X}_3^{(5,1)} = (x_3, \dots, x_7) = (3.2, 4.8, 4.2, 4.5, 2.3)$ , shown in black, is formed by the indexes in  $\boldsymbol{\pi}_3^5 = (1, 2, 3, 4, 5)$  which sort the elements of  $\mathbf{X}_3^{(5,1)}$  in increasing order: 51342. With this,  $\tilde{\boldsymbol{\pi}}_3^5 = 51342$ , and we increase the counting related to this motif in the histogram of all possible patterns of size  $D = 5$ .

The dash-dot line in Fig. 2.1 illustrates  $\mathbf{X}_1^{(5,2)}$ , i.e. the sequence of length  $D = 5$  starting at  $x_1$  with lag  $\tau = 2$ . In this case,  $\mathbf{X}_1^{(5,2)} = (1.8, 3.2, 4.2, 2.3, 1.2)$ , and the corresponding motif is  $\tilde{\boldsymbol{\pi}}_1^5 = 51423$ .

The classic approach to calculating the probability distribution of ordinal patterns is through the frequency histogram. Denote  $\Pi$  the sequence of symbols obtained by a given series  $\mathbf{X}_t^{(D,\tau)}$ . The Bandt-Pompe probability distribution is the relative frequency of symbols in the series against the  $D!$  possible patterns  $\{\tilde{\boldsymbol{\pi}}_t^D\}_{t=1}^{D!}$ :

$$p(\tilde{\boldsymbol{\pi}}_t^D) = \frac{\#\left\{\mathbf{X}_t^{(D,\tau)} \text{ is of type } \tilde{\boldsymbol{\pi}}_t^D\right\}}{T - (D - 1)\tau}, \quad (2.3)$$

where  $t \in \{1, \dots, T - (D - 1)\tau\}$ . These probabilities meet the conditions  $p(\tilde{\boldsymbol{\pi}}_t^D) \geq 0$  and  $\sum_{i=1}^{D!} p(\tilde{\boldsymbol{\pi}}_t^D) = 1$ , and are invariant before monotonic transformations of the time

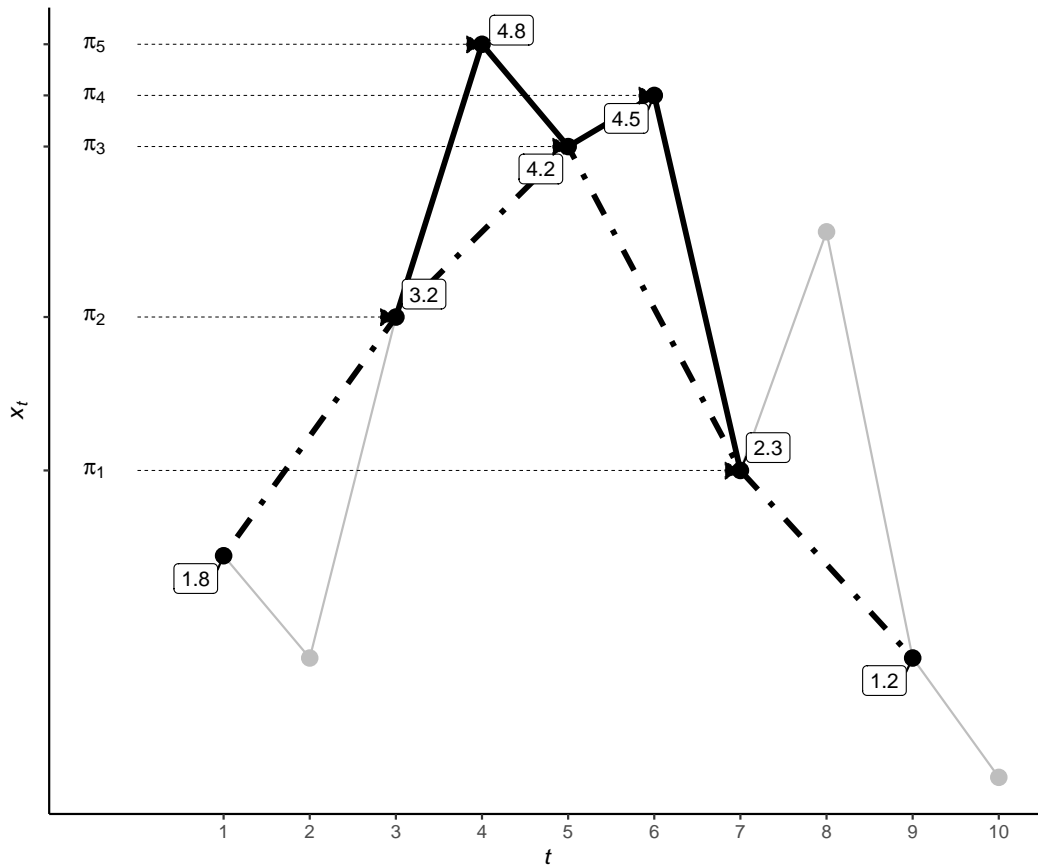


Figure 2.1. Illustration of the Bandt and Pompe coding. Here the gray line represents the analyzed sequence  $\mathcal{X} = (1.8, 1.2, 3.2, 4.8, 4.2, 4.5, 2.3, 3.7, 1.2, 0.5)$ , the sequence illustrated by the dotted line shows the path taken when applying  $\tau = 2$ , and the sequence illustrated by the black line shows the elements of the pattern  $\mathbf{X}_1^{(5,2)} = (1.8, 3.2, 4.2, 2.3, 1.2)$ .

series values. For example, the presence of  $\alpha$  multiplicative noise in  $\mathcal{X}$  does not change the results of the patterns produced.

In the literature, there are two ways to perform the mapping  $\mathbf{X}_t^{(D,\tau)} \mapsto \pi_t^D$  in the symbolization of Bandt-Pompe: (i) Sorting the positions of the elements  $x_t \in \mathbf{X}_t^{(D,\tau)}$  in chronological order (Permutation of Classification), and (ii) Sorting the time indexes of the elements  $x_t \in \mathbf{X}_t^{(D,\tau)}$  (Chronological Index Permutation).

### 2.1.1 Rank Permutation Mapping

For an arbitrary  $t$ , the vector of real values  $\mathcal{X}_t^{(D,\tau)} = (x_t, x_{t+1}, \dots, x_{t+D-1})$  with time delay  $\tau \geq 1$  ( $\tau \in \mathbb{Z}$ ), and embedding dimension  $D \geq 2$  ( $D \in \mathbb{Z}$ ) are mapped to the vector symbol  $\pi^D = (\mathbf{R}[x_t], \mathbf{R}[x_{t+1}], \dots, \mathbf{R}[x_{t+D-1}]) \in \mathfrak{A}_D$  formed by the rank of its

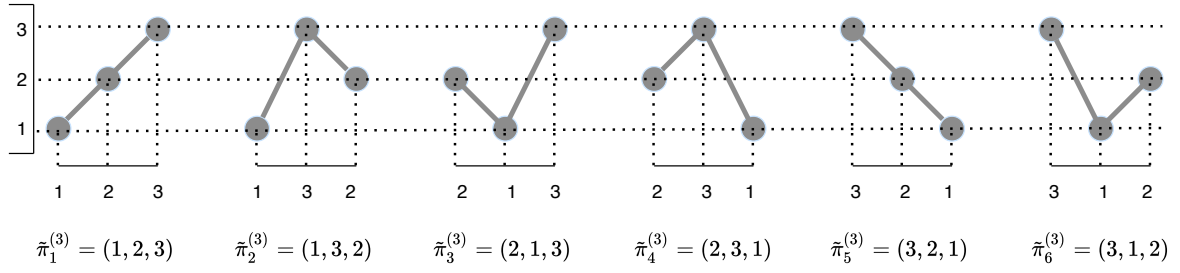


Figure 2.2. **Rank permutation mapping:** The complete alphabet for  $D = 3$  of the rank mapping technique obtained by permuting all possible ranks.

components, defined as the following function:

$$R[x_{t+n}] = \sum_{k=0}^{D-1} \mathbf{1}(x_{t+k} \leq x_{t+n}) \text{ for } n = 0, \dots, D-1 \quad (2.4)$$

where  $x_{t+n} \in X_t^{(D,\tau)}$ ,  $1 \leq R(x_{t+n}) \leq D$ , and  $\mathbf{1}$  is the indicator function:  $\mathbf{1}(Z) = 1$  if  $Z$  is true and 0 otherwise. So the maximum and minimum values of rank are  $R(\min(x_{t+k})) = 1$  and  $R(\max(x_{t+k})) = D$ . The complete alphabet is all the possible permutation of the ranks. Examples of its application can be seen in Riedl et al. (2013) and Bandt and Shiha (2007).

For example, let us take the series 2.2, time delay  $\tau = 1$  and embedding dimension  $D = 3$ , the eight vectors and their corresponding patterns obtained by rank permutation mapping are:

$$\begin{aligned} \mathbf{X}_1^{(3,1)} &= (1.8, 1.2, 3.2) \mapsto (2, 1, 3), & \mathbf{X}_5^{(3,1)} &= (4.2, 4.5, 2.3) \mapsto (2, 3, 1), \\ \mathbf{X}_2^{(3,1)} &= (1.2, 3.2, 4.8) \mapsto (1, 2, 3), & \mathbf{X}_6^{(3,1)} &= (4.5, 2.3, 3.7) \mapsto (3, 1, 2), \\ \mathbf{X}_3^{(3,1)} &= (3.2, 4.8, 4.2) \mapsto (1, 3, 2), & \mathbf{X}_7^{(3,1)} &= (2.3, 3.7, 1.2) \mapsto (2, 3, 1), \\ \mathbf{X}_4^{(3,1)} &= (4.8, 4.2, 4.5) \mapsto (3, 1, 2), & \mathbf{X}_8^{(3,1)} &= (3.7, 1.2, 0.5) \mapsto (3, 2, 1), \end{aligned}$$

In Fig. 2.2, we present an illustrative drawing of this mapping for all alternatives when we have  $D = 3$ . As we can see, the vertical axis indexes are fixed and the resulting pattern is obtained through the time axis labels in chronological order.

### 2.1.2 Chronological Index Permutation Mapping

Again, for an arbitrary  $t$ , the vector of real values  $\mathcal{X}_t^{(D,\tau)} = (x_t, x_{t+1}, \dots, x_{t+D-1})$  with time delay  $\tau \geq 1$  ( $\tau \in \mathbb{Z}$ ), and embedding dimension  $D \geq 2$  ( $D \in \mathbb{Z}$ ) are mapped to the vector symbol  $\boldsymbol{\pi}^D = (\mathbf{i}_1, \mathbf{i}_2, \dots, \mathbf{i}_D) \in \mathfrak{A}_D$  formed by the time indexes ordered according to their amplitudes. So, the sequence must comply  $x_{t+i_1} < x_{t+i_2} < \dots < x_{t+i_D}$ .

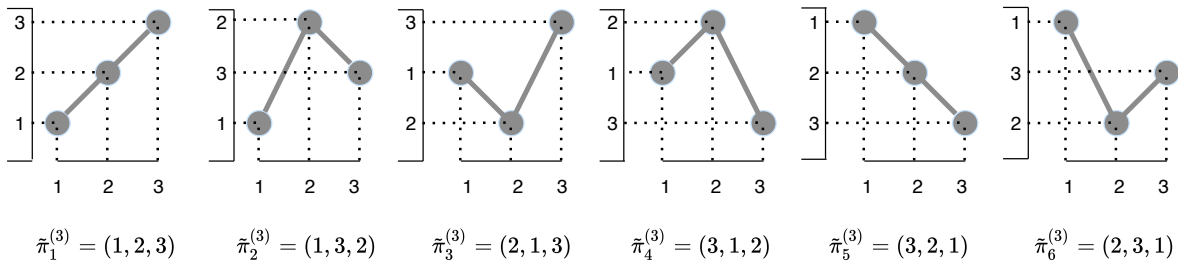


Figure 2.3. **Chronological index permutation mapping:** The complete alphabet for  $D = 3$  of the chronological mapping technique obtained by permuting all of its indexes.

Examples of its application can be seen in Bandt and Pompe (2002b), Ravetti et al. (2014), Parlitz et al. (2012), and Bian et al. (2012).

For example, let us take the series 2.2, time delay  $\tau = 1$  and embedding dimension  $D = 3$ , the eight vectors and their corresponding patterns obtained by rank permutation mapping are:

$$\begin{aligned}
\mathbf{X}_1^{(3,1)} &= (1.8, 1.2, 3.2) \mapsto (2, 1, 3), & \mathbf{X}_5^{(3,1)} &= (4.2, 4.5, 2.3) \mapsto (3, 1, 2), \\
\mathbf{X}_2^{(3,1)} &= (1.2, 3.2, 4.8) \mapsto (1, 2, 3), & \mathbf{X}_6^{(3,1)} &= (4.5, 2.3, 3.7) \mapsto (2, 3, 1), \\
\mathbf{X}_3^{(3,1)} &= (3.2, 4.8, 4.2) \mapsto (1, 3, 2), & \mathbf{X}_7^{(3,1)} &= (2.3, 3.7, 1.2) \mapsto (3, 1, 2), \\
\mathbf{X}_4^{(3,1)} &= (4.8, 4.2, 4.5) \mapsto (2, 3, 1), & \mathbf{X}_8^{(3,1)} &= (3.7, 1.2, 0.5) \mapsto (3, 2, 1),
\end{aligned}$$

In Fig. 2.3, we present an illustrative drawing of this mapping for all alternatives when we have  $D = 3$ . In this mapping, the time indexes are fixed in chronological order. The patterns are chosen by the vertical axis labels in the increasing direction of the amplitude.

## 2.2 Ordinal Patterns Transition Graphs

Alternatively, given the sequence of ordinal patterns  $\Pi$  one may form an oriented graph with the transitions from  $\tilde{\pi}_t^D$  to  $\tilde{\pi}_{t+1}^D$ . The Ordinal Pattern Transition Graph  $G = (V, E)$  represents the transitions between two consecutive ordinal patterns over time  $t$ . The vertices are the  $D!$  possible patterns for an embedding dimension  $D$ , and the edges the transitions between them:  $V = \{v_{\tilde{\pi}_t^D}\}$ , and  $E = \{(v_{\tilde{\pi}_t^D}, v_{\tilde{\pi}_{t+1}^D}) : v_{\tilde{\pi}_t^D}, v_{\tilde{\pi}_{t+1}^D} \in V\}$  (Borges et al., 2019).

The literature reports two approaches to compute the weight of edges. Some authors employ unweighted edges (McCullough et al., 2015; Kulp et al., 2016), which

represent only the existence of transitions, while others apply the frequency of transitions (Sorrentino et al., 2015; Zhang et al., 2017). The weights  $\mathbb{W} = \{w_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} : v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D} \in V\}$  assigned to each edge describe the chance of transitions between the patterns  $(v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D})$ . The weights are calculated as the relative frequency of each transition, i.e.:

$$w_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} = \frac{|\Pi_{\tilde{\pi}_i^D, \tilde{\pi}_j^D}|}{T - (D - 1)\tau - 1}, \quad (2.5)$$

where  $|\Pi_{\tilde{\pi}_i^D, \tilde{\pi}_j^D}|$  is the number of transitions from pattern  $\tilde{\pi}_i^D$  to pattern  $\tilde{\pi}_j^D$ ,  $\sum_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} w_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} = 1$ , and the denominator is the number of transitions between sequential patterns in the series of motifs of length  $T - (D - 1)\tau$ .

When comparing the transition graph with other classical time series representations in graphs, we can highlight some advantageous properties (Borges et al., 2019):

- (i) **Speed and simplicity:** the construction of the graph depends only on two parameters, the time delay  $\tau$  and the embedding dimension  $D$ . On the other hand, the time series symbolization process depends on the length  $N$  of the analyzed series and the dimension of incorporation  $D$ , having a complexity  $\mathcal{O}(n \cdot D^2)$  when we consider a simple algorithm for sorting elements with  $\mathcal{O}(D^2)$  complexity.
- (ii) **Scalability:** Regardless of the number of elements in the time series, the number of vertices of the transition graph will always be limited by  $D!$ , Where  $3 \leq D \leq 7$ .
- (iii) **Robustness:** Since the Bandt-Pompe symbolization produces robust ordinal patterns to the presence of monotonous non-linear noises and transformations, we can conclude that the transformation proposed by the transition graphs also guarantees such specificity.
- (iv) **Probability of self-transition:** The self-transition of the graph represent the proportion of ordinal patterns sequentially repeated and are defined as

$$p_{st} = p(\tilde{\pi}_i^D, \tilde{\pi}_i^D) = \sum_{i \in \{1, \dots, D!\}} w_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_i^D}}. \quad (2.6)$$

Thus, the presence of auto-loops, when evaluated according to the  $\tau$  incorporation delay, proved to be an important characteristic of the underlying dynamics of time series, as they are directly associated with the temporal correlation of the elements.

## 2.3 Information-Theoretic Descriptors

Let  $\mathbb{P} = \{p_{\tilde{\pi}_1^D}, p_{\tilde{\pi}_2^D}, \dots, p_{\tilde{\pi}_{D!}^D}\} = \{p_1, \dots, p_{D!}\}$  be the probability function obtained from the 1-D signal  $\mathcal{X}$  by Bandt-Pompe symbolization. The last step of the characterization process consists of calculating the Information Theory descriptors: Shannon Entropy and Statistical Complexity. Through these features we were able to obtain the point in the  $H \times C$  plane.

### 2.3.1 Permutation Entropy

Entropy measures the disorder or unpredictability of a system characterized by a probability measure  $\mathbb{P}$  and is measured by:

$$H(\mathbb{P}) = - \sum_{i=1}^{D!} p_{\tilde{\pi}_i^D} \log p_{\tilde{\pi}_i^D}. \quad (2.7)$$

Its minimum value occurs when  $H(\mathbb{P}) = H_{min} = 0$ , in this particular case we will have maximum knowledge about the system and we will be able to predict with absolute certainty what will be the next ordinal pattern generated by the data, indicating that the time series is deterministic. On the other hand, when the behavior of the system is described by a uniform distribution, that is, when all the possibilities have the same probability of occurrence and its probability is determined by  $\mathbb{P} = \{1/D!, \dots, 1/D!\}$ , we will have minimal knowledge of the analyzed data, obtaining  $H(\mathbb{P}) = H_{max} = \log D!$ . In that case, the time series would be determined as a completely random system (Bandt and Pompe, 2002b).

However, in the literature it is opted to use the normalized Shannon entropy defined by Martin et al. (2006a), given by:

$$H_S(\mathbb{P}) = \frac{H(\mathbb{P})}{H_{max}} = - \frac{1}{\log D!} \sum_{i=1}^{D!} p_{\tilde{\pi}_i^D} \log p_{\tilde{\pi}_i^D}, \quad (2.8)$$

where  $0 \leq H_S(\mathbb{P}) \leq 1$ .

### 2.3.2 Statistical Complexity

The entropy's ability to capture system properties is limited, so it is necessary to use it in conjunction with other descriptors to obtain a complete analysis. Other interesting measures are the distances between  $\mathbb{P}$  and a probability measure that describes a non-informative process, typically the uniform distribution.

The Jensen-Shannon divergence to the uniform distribution  $\mathbb{U} = (\frac{1}{D!}, \dots, \frac{1}{D!})$  is a measure of how similar the underlying dynamics is to a non-informative process. It is calculated as:

$$JS(\mathbb{P}, \mathbb{U}) = \sum_{i=1}^{D!} \left( p_{\tilde{\pi}_i} \log \frac{p_{\tilde{\pi}_i}}{u_{\tilde{\pi}_i}} + u_{\tilde{\pi}_i} \log \frac{u_{\tilde{\pi}_i}}{p_{\tilde{\pi}_i}} \right). \quad (2.9)$$

Conversely to entropy, the statistical complexity seeks to find interaction and dependence structures among the elements of a given series, being an extremely important factor in the study of dynamic systems. The Statistical Complexity is defined as Lamberti et al. (2004):

$$C_{JS}(\mathbb{P}, \mathbb{U}) = H_S(\mathbb{P}) \cdot Q_{JS}(\mathbb{P}, \mathbb{U}). \quad (2.10)$$

The ‘‘disequilibrium’’  $Q_{JS}$  is calculated by:

$$Q_{JS}(\mathbb{P}, \mathbb{U}) = Q_0 \cdot JS(\mathbb{P}, \mathbb{U}) \quad (2.11)$$

$$= Q_0 \cdot \left\{ H \left[ \frac{\mathbb{P} + \mathbb{U}}{2} \right] - \frac{H[\mathbb{P} + \mathbb{U}]}{2} \right\} \quad (2.12)$$

The normalization constant is equal to the inverse of the maximum value of  $JS$  (Rosso et al., 2007b), so:

$$Q_0 = -2 \left\{ \left( \frac{D! + 1}{D!} \right) \ln(D! + 1) - 2 \ln(2D!) + \ln(D!) \right\}^{-1}, \quad (2.13)$$

where  $0 \leq Q_0 \leq 1$ .

This descriptor quantifies, in addition to randomness, the presence of correlational structures between patterns, reflecting the architecture of the systems, being different from zero if there are more likely states than the others. In this way, different degrees of structure can be quantified, reflecting properties revealed by the probability distribution of the underlying process.

### 2.3.3 Causality Complexity-Entropy Plane

The set of all pairs  $(H(\mathbb{P}), C(\mathbb{P}, \mathbb{U}))$  for any time series described by patterns of length  $D$  lies in a compact subset of  $\mathbb{R}^2$ : the Complexity-Entropy plane (or Entropy-Complexity plane). Through this tool it is possible to discover the nature of the series just by checking its region of location on the plane, its associated values help

to determine whether it corresponds to a chaotic (or other deterministic dynamics) or stochastic sequence.

Anteneodo and Plastino (1996) proved that, for a fixed value of entropy, there are two extreme values of complexity. Martin et al. (2006b), using geometrical arguments on the space of configurations, found expressions for such boundaries. The lower boundary  $C_{\min}$  is continuous, while the upper  $C_{\max}$  is defined by  $D! - 1$  pieces.

Fig. 2.4 shows the boundaries of the  $H \times C$  plane for the embedding dimensions  $D = 3$  (red)  $D = 4$  (green), and  $D = 5$  (blue). The jagged structure of  $C_{\max}$  increases the difficulty of finding distributions for the points in the  $H \times C$  plane.

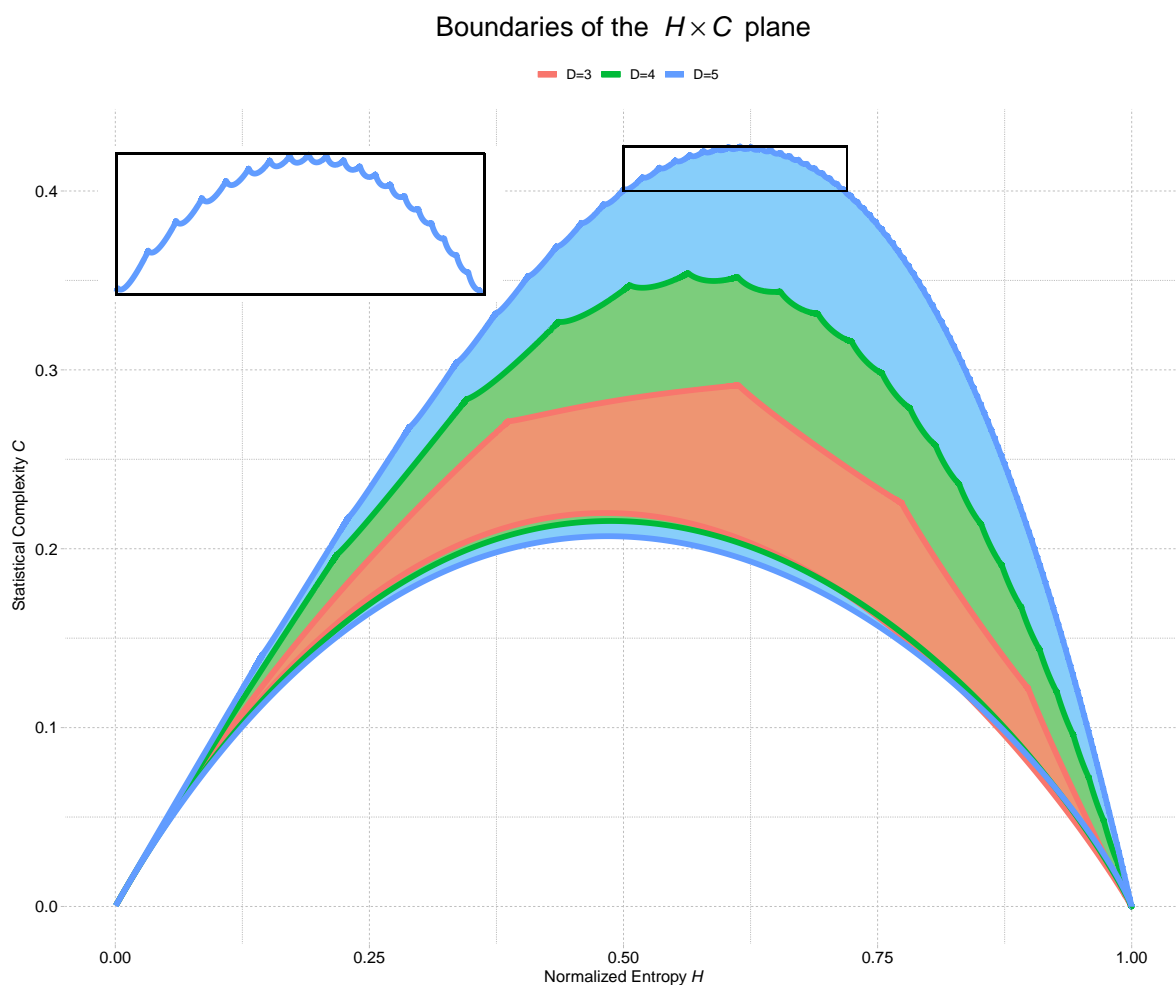


Figure 2.4. Boundaries of the  $H \times C$  plane for  $D = 3, 4, 5$ .

We illustrate the use of the Complexity-Entropy ( $H \times C$ ) with the following time series:

- $f^{-k}$  **noises**. Correspond to colored noise sequences generated synthetically using the power spectrum  $f^{-k}$  (Ravetti et al., 2014). Here, we opt to use the following



configurations of the spectrum: white ( $k = 0$ ),  $k = -1/2$ , pink ( $k = 1$ ),  $k = 3/2$ , red ( $k = 2$ ),  $k = 5/2$ , and  $k = 3$ ;

- **Logistic maps.** The classical chaotic map defined by the formula  $x_t = rx_{t-1}(1 - x_{t-1})$  (Peitgen et al., 2006). In this work, we use this chaotic logistic series with the parameters  $r = 3.6$  and 4;
- **Deterministic series.** Monotonic increasing ( $\log(x_t + 0.1)$ ,  $x_t = \{1, 2, \dots, 10^4\}$ ) and periodic ( $\sin(2x_t) \cos(2x_t)$ , with  $0 \leq x_t \leq 2\pi$  over ten thousand equally spaced points).

In all cases, we used  $D = 6$  and  $\tau = 1$ . Fig. 2.5 shows nine of the histograms produced by these series using the Mersenne-Twister pseudorandom number generator; we omitted those corresponding to the deterministic series, as they produce one and two nonzero bins. As we can see, as we add more correlation structures,  $f^{-k}$  noises have a less uniform histogram, making it more evident that some ordinal patterns stand out and are more frequent than others. Therefore, we have an indication that there is a structure of dependence between such elements, making them more deterministic.

Fig. 2.6 shows the  $H \times C$  plane with the bounds for  $D = 6$ , the time series, and the points they were mapped onto. The points due to  $f^{-k}$  noises appear joined by dotted segments. It is noticeable that deterministic patterns have more complexity than random ones. Also, points related to  $f^{-k}$  noises tend to clutter for  $k < 1$ , having the highest entropy values.

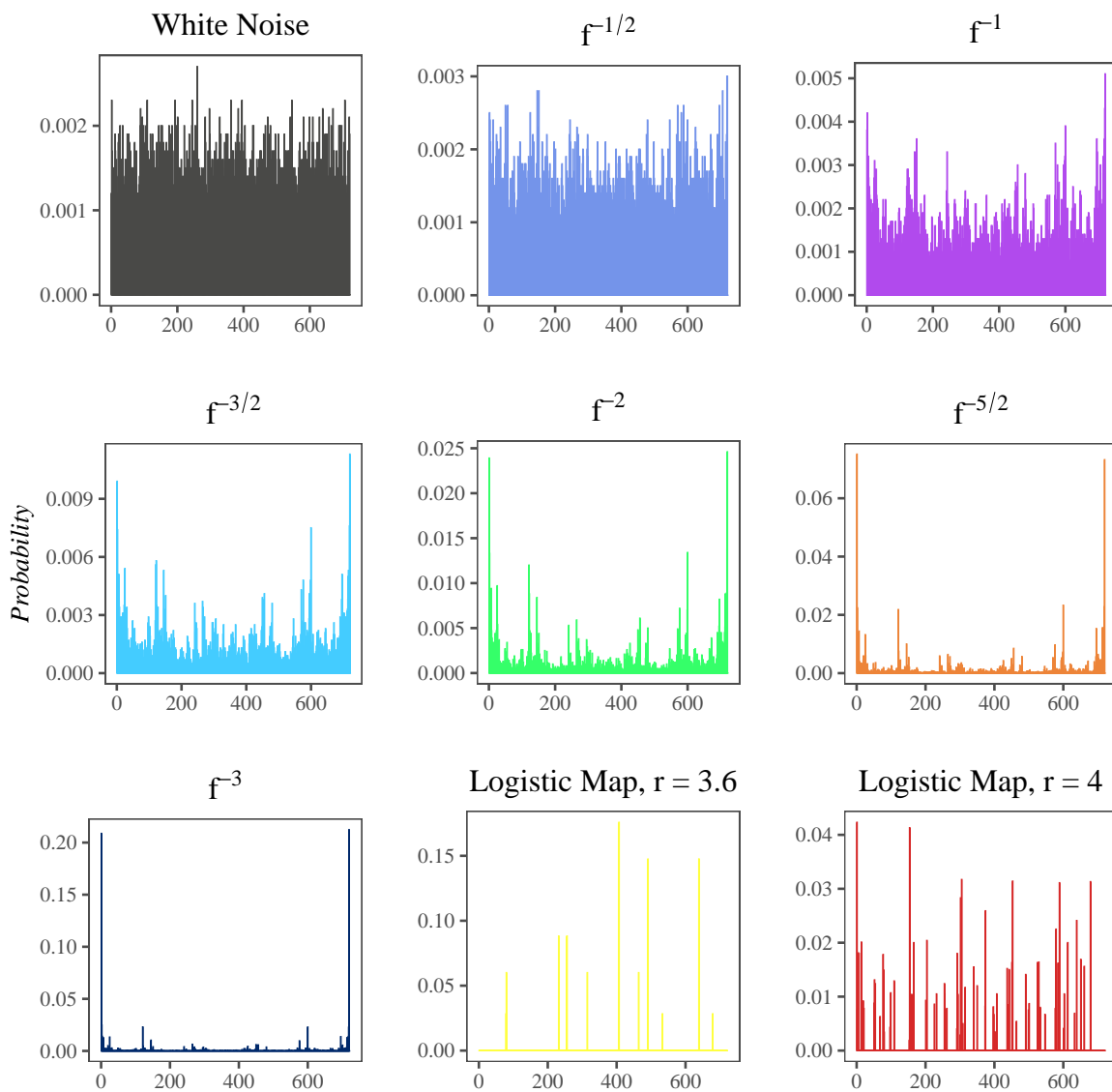


Figure 2.5. Patterns histograms of selected time series with  $D = 6$ ,  $\tau = 1$  and  $T = 10^4$ .

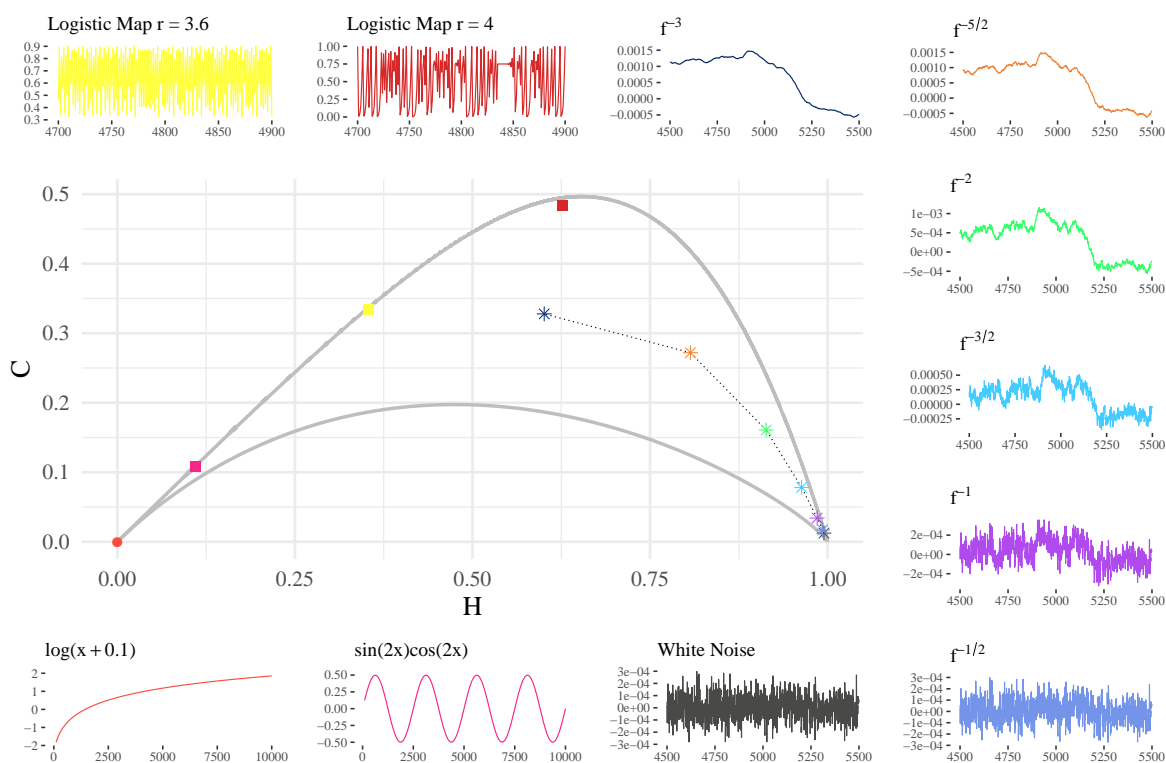


Figure 2.6. Eleven systems and their points in the  $H \times C$  plane when we apply  $D = 6$ ,  $\tau = 1$  and  $T = 10^4$ .

## Chapter 3

# Weighted Amplitude Transition Graph

Although the use of Bandt-Pompe symbolization has several advantages over other feature extraction algorithms, it has two major gaps in its seminal definition:

- (i) The ordinal ambiguity present when we have equal values in the same sub-sequence (Traversaro et al., 2018, 2017), and
- (ii) The lack of information related to differences in sample amplitude (Azami and Escudero, 2016; Fadlallah et al., 2013; Cuesta-Frau, 2019), that is, the mean value of the amplitudes and the differences between neighboring samples are not considered by the original methodology.

Figure 3.1 shows two examples of how sub-sequences with different structural characteristics can be mapped to the same ordinal pattern, thus decreasing the characterization power of the approach.

By adding amplitude information, we were able to attribute less complexity in sequences with greater regularity and locate abrupt changes along the signal, reducing the impact of a possible noise degradation in relation to the final value of the descriptors. In this context, several modifications were proposed in the calculation of the Bandt-Pompe distribution with the aim of neutralizing the limitations discussed and maintaining most of the properties presented in the technique initially.

To analyze the different methods present in the literature, we will apply the following sequence as a numerical example:

$$\mathcal{X} = \{-3.7, -3.5, 2, 1.3, 0.8, -2.3, 1.8, 1.7, 1.3, 2.6, 1.7, 0.9, 0, -0.4, -0.5, 7\} \quad (3.1)$$

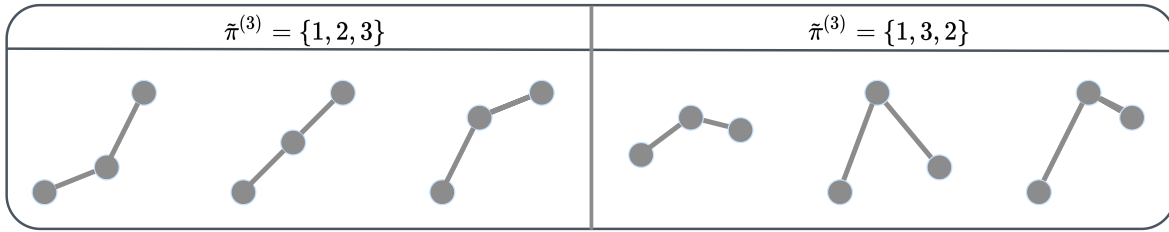


Figure 3.1. Two examples of possible  $D$ -dimensional vectors corresponding to the same ordinal patterns when using  $D = 3$ .

The Chronological Index Permutation was used to generate the ordinal patterns, where for  $D = 3$  and  $\tau = 1$ , we will have  $M = 15$  vectors  $\mathbf{X}^{(D=3)}$ . Its probability distribution obtained by the traditional Bandt-Pompe symbolization method is  $\mathbb{P} = \{0, 0.2, 0.2, 0.0666667, 0.0666667, 0.4666667\}$ .

### 3.1 Weighted Ordinal Patterns Methods in Literature

There are basically two strategies for incorporating amplitude information into ordinal structures. The first assumes that the greater the variation in amplitude in a given ordinal pattern, the greater its weight in calculating the probability distribution. Thus, correction factors are proposed to carry out such weighting. The second proposes an extension in the alphabet of symbols, mapping vectors with different amplitudes to different ordinal patterns. Thus, the total set of patterns has more than  $D!$  possible components and vary according to the mapping method considered (rank permutation or chronological index permutation).

#### 3.1.1 Weighted Permutation Entropy

The Weighted Permutation Entropy (WPE) was proposed by Fadlallah et al. (2013). The proposed idea is to weigh the relative frequencies of the patterns taking into account the variability of the elements in relation to the average value of the analyzed segment. Denote  $\bar{X}_t^{(D,\tau)}$  the arithmetic mean:

$$\bar{X}_t^{(D,\tau)} = \frac{1}{D} \sum_{k=1}^D x_{t+(k-1)}. \quad (3.2)$$

The weight  $w_t$  is the sample variance of each vector  $X_t^{(D,\tau)}$ :

$$w_t = \frac{1}{D} \sum_{k=1}^D [x_{t+(k-1)} - \bar{X}_t^{(D,\tau)}]^2. \quad (3.3)$$

Then, the probability distribution is given from the weighted relative frequencies:

$$p(\tilde{\pi}_t^D) = \frac{\sum_{i:\{\mathbf{X}_i^{(D,\tau)} \mapsto \tilde{\pi}_t^D\}} w_i}{\sum_{i=1}^{T-(D-1)\tau} w_i}. \quad (3.4)$$

Using the time series 3.1 as an example, the weight shown by WPE for each subsequence (the variance of each vector) would be as follows:  $w_1 = 23.389$ ,  $w_2 = 18.389$ ,  $w_3 = 5.976$ ,  $w_4 = 0.242$ ,  $w_5 = 2.536$ ,  $w_6 = 3.047$ ,  $w_7 = 3.647$ ,  $w_8 = 0.047$ ,  $w_9 = 0.296$ ,  $w_{10} = 0.296$ ,  $w_{11} = 0.482$ ,  $w_{12} = 0.482$ ,  $w_{13} = 0.296$ ,  $w_{14} = 0.047$ ,  $w_{15} = 12.336$ . Being described in the  $H \times C$  plane by the following point  $(\mathbf{h}, \mathbf{c}) = (0.965, 0.0322)$ .

### 3.1.2 Fine-Grained Permutation Entropy

The Fine-Grained Permutation Entropy (FGPE) was introduced in Ref. Xiao-Feng and Yue (2009). Let  $\beta_t$  be the difference series:

$$\beta_t = \{|x_{t+1} - x_t|, \dots, |x_{t+(D-1)} - x_{t+(D-2)}|\}. \quad (3.5)$$

The weight  $w_t$  quantifies such differences:

$$w_t = \left\lfloor \frac{\max\{\beta_t\}}{\alpha s(\beta_t)} \right\rfloor, \quad (3.6)$$

where  $s$  is the sample standard deviation,  $\alpha$  is a user-defined parameter, and  $\lfloor \cdot \rfloor$  is the floor function. Then,  $w_t$  is added as a symbol at the end of the corresponding pattern to quantify the difference between neighboring elements in the vector, leading to an update of  $\Pi$ :

$$\pi_t^D = \{\tilde{\pi}_t^D \cup w_t\}. \quad (3.7)$$

Finally, the probability distribution is calculated as:

$$p(\pi_t^D) = \frac{\#\{\mathbf{X}_t^{(D,\tau)} \text{ is of type } \pi_t^D\}}{T - (D - 1)\tau}. \quad (3.8)$$

Using the time series 3.1 as an example, when applying FGPE we will have the formation of the following symbols:

$$\begin{aligned}
\mathbf{X}_1^{(3)} &\mapsto (2, 0, 1, 59), & \mathbf{X}_4^{(3)} &\mapsto (2, 1, 0, 9), & \mathbf{X}_7^{(3)} &\mapsto (0, 2, 1, 0), & \mathbf{X}_{10}^{(3)} &\mapsto (0, 2, 1, 9), & \mathbf{X}_{13}^{(3)} &\mapsto (2, 1, 0, 5), \\
\mathbf{X}_2^{(3)} &\mapsto (1, 2, 0, 5), & \mathbf{X}_5^{(3)} &\mapsto (2, 1, 0, 3), & \mathbf{X}_8^{(3)} &\mapsto (2, 1, 0, 3), & \mathbf{X}_{11}^{(3)} &\mapsto (2, 1, 0, 25), & \mathbf{X}_{14}^{(3)} &\mapsto (2, 1, 0, 3), \\
\mathbf{X}_3^{(3)} &\mapsto (0, 2, 1, 3), & \mathbf{X}_6^{(3)} &\mapsto (1, 0, 2, 11), & \mathbf{X}_9^{(3)} &\mapsto (1, 0, 2, 4), & \mathbf{X}_{12}^{(3)} &\mapsto (2, 1, 0, 25), & \mathbf{X}_{15}^{(3)} &\mapsto (1, 0, 2, 2)
\end{aligned}$$

### 3.1.3 Amplitude-Aware Permutation Entropy

The Amplitude-Aware Permutation Entropy (AAPE) was proposed by Azami and Escudero (2016) and presents a solution to the absence of amplitude information and also quantifies the possible effect of repeated values during symbolization. Here we will report only its simplified version presented in Cuesta-Frau et al. (2018). It consists of weighting the amplitude of ordinal patterns by both the mean and the differences of the elements. For this, only an additional parameter  $A \in [0, 1]$  is required:

$$w_t = \frac{A|x_t|}{D} + \sum_{k=1}^{D-1} \left( \frac{A|x_{t+k}|}{D} + \frac{(1-A)|x_{t+k} - x_{t+k-1}|}{D-1} \right). \quad (3.9)$$

The AAPE consists of one of the methods present in the first approach mentioned, that is, it will also apply a correction factor to the ordinal patterns obtained by the analyzed sequence. In this way, the probability histogram, instead of being updated each time a match is found, will be updated using the weighted relative frequency:

$$p(\tilde{\pi}_t^D) = \frac{\sum_{i:\{\mathbf{X}_i^{(D,\tau)} \mapsto \tilde{\pi}_t^D\}} w_i}{\sum_{i=1}^{T-(D-1)\tau} w_i}. \quad (3.10)$$

Using the time series 3.1 as an example, given the ordinal patterns obtained, the AAPE algorithm acquires the following weights for each subsequence:  $w_1 = 2.683$ ,  $w_2 = 2.016$ ,  $w_3 = 1.483$ ,  $w_4 = 2.616$ ,  $w_5 = 1.358$ ,  $w_6 = 3.216$ ,  $w_7 = 6.083$ ,  $w_8 = 7.375$ ,  $w_9 = 0.983$ ,  $w_{10} = 1.633$ ,  $w_{11} = 0.925$ ,  $w_{12} = 1.291$ ,  $w_{13} = 0.858$ ,  $w_{14} = 0.541$ ,  $w_{15} = 0.275$ .

## 3.2 Proposed Method

Recent approaches to extracting knowledge in time series have been developed using information from transition graphs of ordinal patterns. However, such tooling comes up against the same problem as the traditional symbolization of Bandt-Pompe, the lack of information of amplitude between different sequences. In order to be the first

variation of transition graphs to incorporate this type of information, we propose in this present work the Weighted Amplitude Transition Graph (WATG).

Since we are interested in a comparable metric for different data sets, the first step is to normalize each time series  $\mathcal{X}$  to the interval  $[0, 1]$ :

$$\frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \mapsto x_i, \quad (3.11)$$

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.

Assuming that a variation in amplitude in a given ordinal pattern should directly impact the resulting probability distribution. Thus, we assign for each vector  $\mathbf{X}_t^{(D,\tau)}$  an associated weight  $\beta_t$  that measures the biggest difference between its elements:

$$\beta_t = \max\{|x_i - x_j|\}, \quad (3.12)$$

where  $x_i, x_j \in \mathbf{X}_t^{(D,\tau)}$ .

Thus, each edge will have its value no longer associated with the number of transitions made between ordinal patterns, but with the difference in amplitude observed during the respective transition:

$$w_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} = \sum_{i:\{\mathbf{X}_t^{(D,\tau)} \mapsto \tilde{\pi}_i^D\}} \sum_{j:\{\mathbf{X}_t^{(D,\tau)} \mapsto \tilde{\pi}_j^D\}} |\beta_i - \beta_j|. \quad (3.13)$$

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

$$\begin{cases} \lambda_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} = 1, & \text{if } (v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}) \in E, \\ \lambda_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} = 0, & \text{otherwise.} \end{cases}, \text{ and} \quad (3.14)$$

$$p(\tilde{\pi}_i^D, \tilde{\pi}_j^D) = \frac{\lambda_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}} \cdot w_{v_{\tilde{\pi}_i^D}, v_{\tilde{\pi}_j^D}}}{\sum_{v_{\tilde{\pi}_a^D}, v_{\tilde{\pi}_b^D}} w_{v_{\tilde{\pi}_a^D}, v_{\tilde{\pi}_b^D}}}. \quad (3.15)$$

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.

We were able to observe the impact of adding amplitude information to transition graphs on two different occasions:

- If the signal 1-D shows a low variation in amplitude and peak intensity between



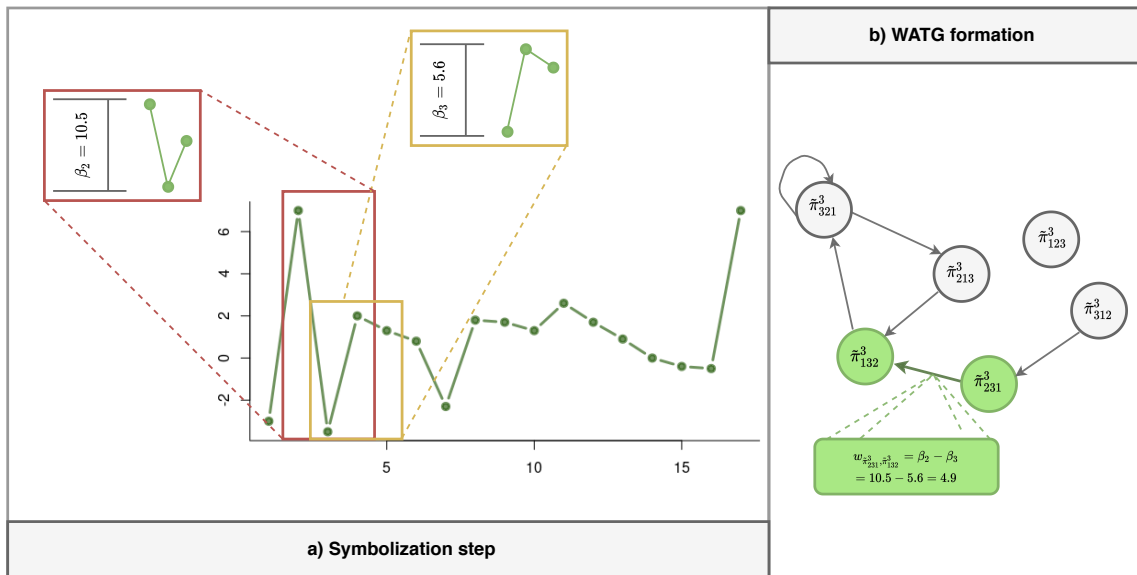


Figure 3.2. Schematic of the composition of the weighted amplitude graph (WATG).

them, the transitions of the ordinal patterns that represent the latter have greater weights. This contributes to the probability distribution becoming less uniform, since it will be more concentrated on these edges. Another characteristic observed is a drop in the entropy values of these sequences when compared to the traditional method.

- If the signal 1-D shows a uniform amplitude variation, the weights assigned to the edges will be well distributed, thus also giving rise to a uniform probability distribution and consequently a greater associated entropy.

The formation schematic of our weighted graph proposal is illustrated in Figure 3.2, where we show how we perform the calculation of the weights of ordinal patterns.

# Chapter 4

## Case study: Classification of SAR Textures

We present a newly proposed technique for texture analysis and classification based on the Bandt-Pompe symbolization for SAR data. It consists of (i) linearize a 2-D patch of the image using the Hilbert-Peano curve, (ii) build the Ordinal Pattern Transition Graph; (iii) obtain a probability distribution function derived from this graph; (iv) compute Information Theory descriptors (Permutation Entropy and Statistical Complexity) from this distribution and use them as features to feed a classifier. The ordinal pattern graph we propose considers that the edges' weight is related to the absolute difference of observations, which encodes the information about the data amplitude. This modification considers the unfavorable signal-to-noise ratio of SAR images and leads to the characterization of several types of textures. Experiments with data from Munich urban areas, Guatemala forest regions, and Cape Canaveral ocean samples show the effectiveness of our technique in homogeneous areas, achieving satisfactory separability levels. The two descriptors chosen in this work are easy and quick to calculate and are used as input for a  $k$ -nearest neighbor classifier. Experiments show that this technique presents results similar to state-of-the-art techniques that employ a much larger number of features and, consequently, impose a higher computational cost.

### 4.1 Introduction

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.

Yue et al. (2020) provide a comprehensive account of how the physical properties of the target are translated into first- and second-order statistical properties of SAR intensity data. There is general agreement that non-deterministic textures are encoded in the second-order features, i.e., in the spatial correlation structure. Therefore is frequent the use of covariance matrix and other measures that assume that a linear dependence, namely the Pearson correlation coefficient, suffices to characterize natural textures. However, in SAR imagery, texture is often visible only over large areas, and the multiplicative and non-Gaussian nature of speckle antagonizes with the additive assumption that underlies classical approaches, making complex the process of characterizing such data.

Surface classification and land use are among the most critical applications of the Synthetic Aperture Radar (SAR) image (Lee et al., 2004). In recent years, handcrafted features and representation learning (supervised and unsupervised) algorithms have been proposed (Han et al., 2020; Huang et al., 2020; Xie et al., 2020). Algorithms of the unsupervised generative adversarial network (GAN) have revolutionized the classification of SAR images, improving performance in small sample problems, and helping the interpretability of such data (Liu et al., 2019). Among the supervised algorithms, support vector machine (SVM) (Sukawattanavijit et al., 2017), random forest (RF) (McNairn et al., 2014), and neural network (NN) (Lin et al., 2017) have been frequently used in remote sensing. The Principle Component Analysis (PCA) (Ressel et al., 2015), autoencoder (Wang and Wang, 2019) and the Boltzmann machine (Qin et al., 2017) can to extract non-local resources and classify non-labeled PolSAR pixels using an unsupervised approach. However, methods such as graph-based semi-supervised deep learning algorithms (Bi et al., 2018) can improve classification accuracy in problems with few labeled samples.

Handcrafted features in SAR textures can be studied following two complementary approaches, namely analyzing the marginal properties of the data (first-order

statistics), and observing their spatial structure (Yue et al., 2020; Numbisi et al., 2018). In this work, we focus on the second approach, which shows relevant results using techniques from the image processing literature, such as co-occurrence matrices and Haralick’s descriptors (Yu et al., 2019). Through the gray-level co-occurrence matrices (GLCM), we can extract features that reflect statistical relationships of the pixel intensity values. On the other hand, Haralick’s descriptors can capture information on intensity and amplitude based on global statistics of SAR images. Radford et al. (2018) used textural information derived from GLCM, along with Random Forests, for geological mapping of remote and inaccessible localities; the authors obtained a classification accuracy of  $\approx 90\%$ , even when using limited training data ( $\approx 0.15\%$  of the total data). Hagensieker and Waske (2018) evaluated the synergistic contribution of multi-temporal L-, C-, and X-band data to tropical land cover mapping, comparing classification outcomes of ALOS-2 (Kankaku et al., 2013), RADARSAT-2 (Morena et al., 2004), and TerraSAR-X (Breit et al., 2009) datasets for a study site in the Brazilian Amazon using a wrapper approach. The wrapper utilizes the gray-level co-occurrence matrix texture information and a Random Forest classifier to estimate scene importance. Storie (2018) proposed an open-source workflow for detecting and delineating the urban-rural boundary using Sentinel-1A SAR data. The author used a combination of GLCM information and a k-means classifier to produce a three-category map that distinguishes urban from rural areas. In higher resolution image classification activities, it is necessary to obtain more granular information from the data by extracting local characteristics such as scale and orientation. In this scenario, techniques such as Fourier power spectrum (Florindo and Bruno, 2012), random fields (Zhu et al., 2016), Gabor filter (Dumitru et al., 2014) and wavelet transform (Akbarizadeh, 2012) are usually applied.

In our approach, we opt to analyze the 1-D signals resulting from the linearization of the image samples, using non-parametric time series analysis techniques. 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 Entropy and the Statistical Complexity, which are easy to obtain and are interpretable. They reveal important features of the underlying process.

The following question guides us:

What is the best representation of a texture patch that allows extracting

expressive Information Theory descriptors to characterize textures in the presence of speckle?

Thus, we use the Weighted Amplitude Transition Graph (WATG), incorporating the absolute difference among observations as weights of the edges between nodes transitions. Such weights take part in the computation of the probabilities and, thus, influence both Entropy and Statistical Complexity.

This work's main contribution is the proposal of a new representation of SAR textures, which allows a low-dimensional characterization useful for, among other applications, their classification. We compare its performance with the classical histograms of Bandt-Pompe ordinal patterns and the regular transition graph. Since the proposed approach has a low computational cost, the results obtained suggest that this technique has good potential in other applications, such as texture segmentation tools of SAR images.

## 4.2 Linearization of image patches

We perform a data dimensionality reduction by turning the 2-D patch into a 1-D signal. This could be accomplished by reading the data by lines, columns, or any transformation of 2-D indexes into a sequence of integers. In this work, we chose to use the Hilbert-Peano (Lee and Hsueh, 1994) curve, due to its low computational cost and its ability to preserve relevant properties of pixel spatial correlation.

Nguyen and Quinqueton (1982) firstly employed Space-filling curves, to map texture into a one-dimensional signal. Carincotte et al. (2006) used the Hilbert-Peano curve in the problem of change detection in pairs of SAR images. The authors noted that this transformation exploits the spatial locality and that its pseudo-randomness of direction changes work well for a large family of images, especially natural ones.

Assuming an image patch is supported by an  $M \times N$  grid, we have the following definition.

**Definition 1** *An image scan is a bijective function  $f: \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$  in the ordered pair set  $\{(i, j) : 1 \leq i \leq M, 1 \leq j \leq N\}$ , which denotes the points in the domain, for the closed range of integers  $\{1, \dots, MN\}$ . A scan rule is  $\{f^{-1}(1), \dots, f^{-1}(MN)\}$ .*

This Definition imposes that each pixel is visited only once and that all pixels are visited.

Space-filling curves, such as raster-1, raster-2, and Hilbert-Peano scanning techniques, stipulate a proper function  $f$ . Hilbert-Peano curves scan an array of pixels

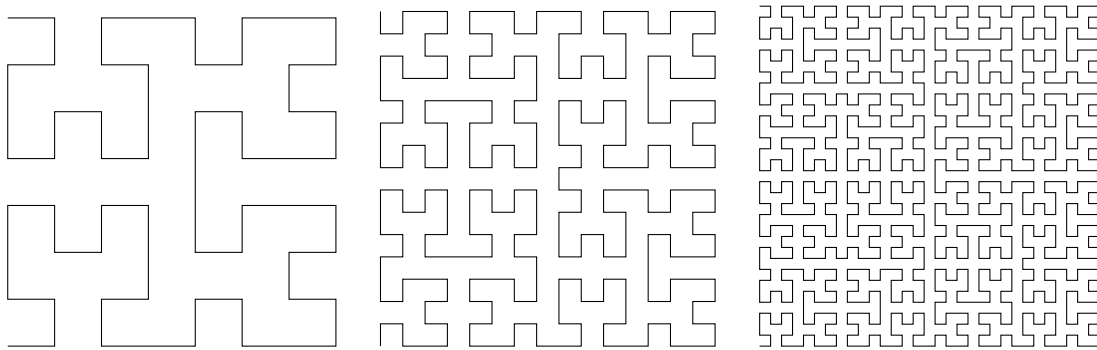


Figure 4.1. Hilbert-Peano curves in areas of: (a)  $8 \times 8$ , (b)  $16 \times 16$ , and (c)  $32 \times 32$  pixels.

of dimension  $2^k \times 2^k$ ,  $k \in \mathbb{N}$ , never keeping the same direction for more than three consecutive points, as shown in Fig. 4.1. Using the Hilbert-Peano curve, we reduce the data dimensionality while maintaining the patch’s spatial dependence information. In this work, we use Hilbert-Peano patches of size  $128 \times 128$ .

Figs. 4.3(a), 4.3(b), 4.3(c), 4.3(d), and 4.3(e) show five image patches with different textures. Figs. 4.3(f), 4.3(g), 4.3(h), 4.3(i), and 4.3(j) present their 1-D representation as signals.

### 4.3 Methodology

Algorithm 1 outlines our methodology. Line 1 transforms the input texture patch  $P$  in a 1-D signal with a Hilbert-Peano sequence. With this, the spatial information is encoded into a one-dimensional signal. Line 2 computes the probability distribution of the weighted transition graph induced by the 1-D signal. The WATG function is detailed in Lines 7–10. Lines 3 and 4 compute the two descriptors of the patch.

The WATG function consists of three steps: (i) each sub-sequence of size (dimension)  $D$  of observations at delay  $\tau$  is transformed into an ordinal pattern using the Bandt-Pompe symbolization (function `BPSymbolization`, Line 7); (ii) function `transitions` (Line 8) calculates the sequence of alternations of the ordinal patterns; and (iii) function `weightGraph` (Line 9) generates the incidence matrix of the graph using the as weights the amplitude differences between the time series elements. Finally, the probability distribution is obtained by turning the transition matrix into a vector (Line 10). These steps are also depicted in Fig. 4.2.

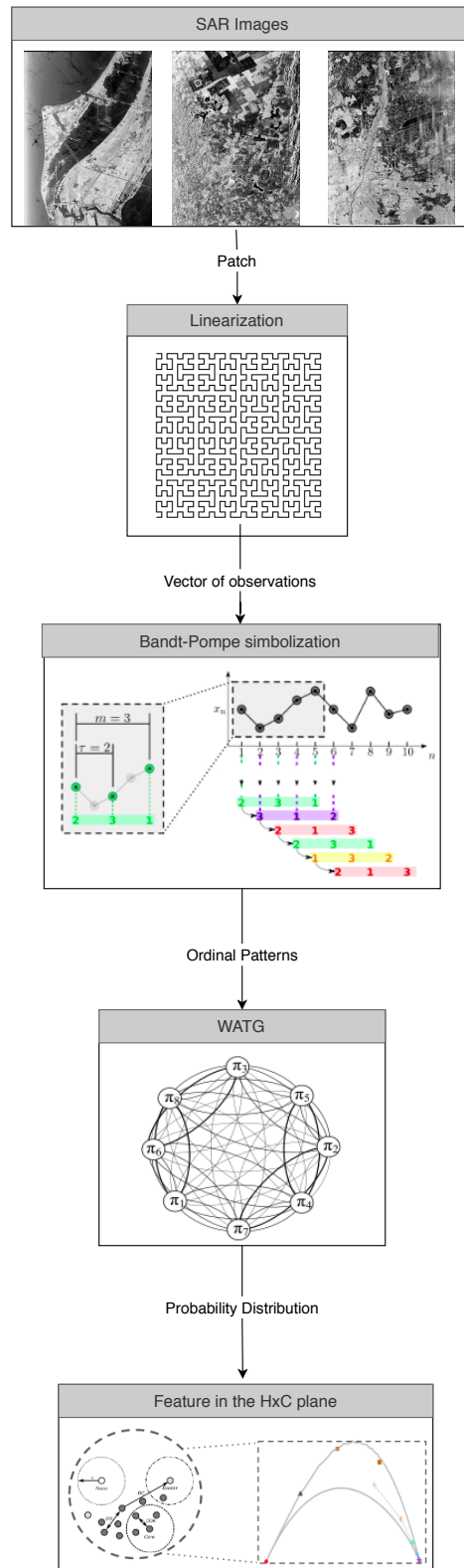


Figure 4.2. Outline of the methodology used for the classification of textures.

```

1 Input: Patch of texture  $P$ , dimension  $D$  and time delay  $\tau$ 
2 Output:  $H \times C$  feature
  1: signal.1-D  $\leftarrow$  hilbertCurve( $P$ )
  2: Probs  $\leftarrow$  WATG(signal.1-D,  $D$ ,  $\tau$ )
  3: H  $\leftarrow$  ShannonEntropy(Probs)
  4: C  $\leftarrow$  StatisticalComplexity(Probs)
  5: return  $H \times C$ 
6: function WATG(signal.1-D,  $D$ ,  $\tau$ )
7:   patterns  $\leftarrow$  BPSymbolization(signal.1-D,  $D$ ,  $\tau$ )
8:   transitions  $\leftarrow$  transitions(patterns)
9:   graph  $\leftarrow$  weighthGraph(signal.1-D, transitions)
10:  Probs  $\leftarrow$  as.vector(graph)
11:  return Probs
12: end function

```

**Algorithm 1:**  $H \times C$  point from a patch using WATG

## 4.4 Image Dataset

We used the HH backscatter magnitudes of three quad-polarimetric L-band SAR images from the NASA Jet Propulsion Laboratory’s (JPL’s) uninhabited aerial vehicle synthetic aperture radar (UAVSAR) sensor with  $L = 36$  nominal looks:

- Forest and pasture region of Sierra del Lacandón National Park, Guatemala, (acquired on April 10, 2015)<sup>1</sup>. The image has  $8917 \times 3300$  pixels with  $10 \text{ m} \times 2 \text{ m}$  resolution.
- Ocean regions from Cape Canaveral Ocean (acquired on September 22, 2016). The image has  $7038 \times 3300$  pixels with  $10 \text{ m} \times 2 \text{ m}$  resolution;
- Urban area of the city of Munich, Germany (acquired on June 5, 2015)<sup>2</sup>. The image has  $5773 \times 3300$  pixels with  $10 \text{ m} \times 3 \text{ m}$  resolution.

We manually selected 200 samples of size  $128 \times 128$  to compose the dataset used in the experiments. It is organized as follows: 40 samples from Guatemalan forest regions; 40 samples from Guatemalan pasture regions; 80 samples from the oceanic regions of Cape Canaveral, divided into two types with different contrast; and 40 samples of urban regions of the city of Munich. Fig. 4.3 shows examples of each. In our analysis, both types of ocean images are grouped.

<sup>1</sup>[https://uavsar.jpl.nasa.gov/cgi-bin/product.pl?jobName=Lacand\\_30202\\_15043\\_006\\_150410\\_L090\\_CX\\_01#dados](https://uavsar.jpl.nasa.gov/cgi-bin/product.pl?jobName=Lacand_30202_15043_006_150410_L090_CX_01#dados)

<sup>2</sup>[https://uavsar.jpl.nasa.gov/cgi-bin/product.pl?jobName=munich\\_19417\\_15088\\_002\\_150605\\_L090\\_CX\\_01#data](https://uavsar.jpl.nasa.gov/cgi-bin/product.pl?jobName=munich_19417_15088_002_150605_L090_CX_01#data)



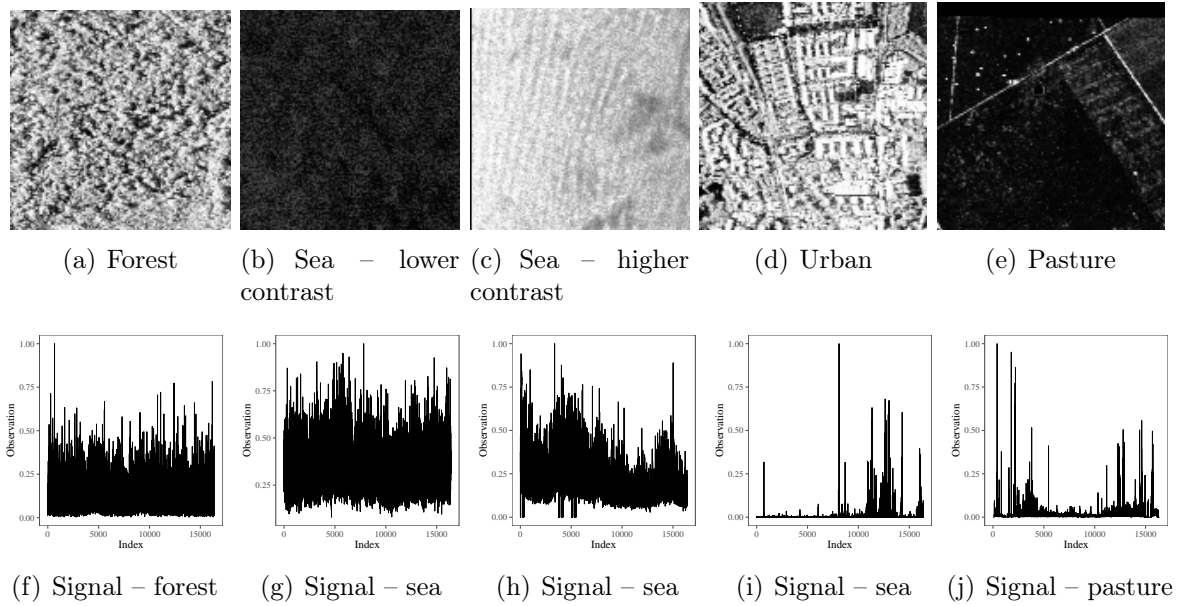


Figure 4.3. Types of regions (Guatemala forest, Canaveral ocean types 1 and 2, Munich urban area, and pasture area) and their signal representation.

We randomly split the samples in training (85%) and test (15%) sets. We used the first set to train a  $k$ -nearest neighbor classifier algorithm with tenfold cross-validation.

## 4.5 Experimental Results and Analysis

In this section, we describe the classification process, and the results of applying WATG. To assess the performance of the technique here proposed, we analyze the impact of its parameters and compare its results in the classification with other methods.

### 4.5.1 Properties

We conducted two experiments to analyze the response of WATG to different noise levels and image rotations. Our truth is the deterministic image generated by the function:

$$z(x, y) = \sin(4x + 0.5y),$$

where  $x, y \in [-2\pi, 2\pi]$ . Fig. 4.4(a) shows this function as a  $128 \times 128$ -pixel patch.

The speckle noise was modeled as outcomes of independent, identically distributed unitary-mean Gamma random variables with shape parameter  $L$  (the number of looks, which controls the signal-to-noise ratio)  $W(L) \sim \Gamma(L, L)$ , with  $L \in$

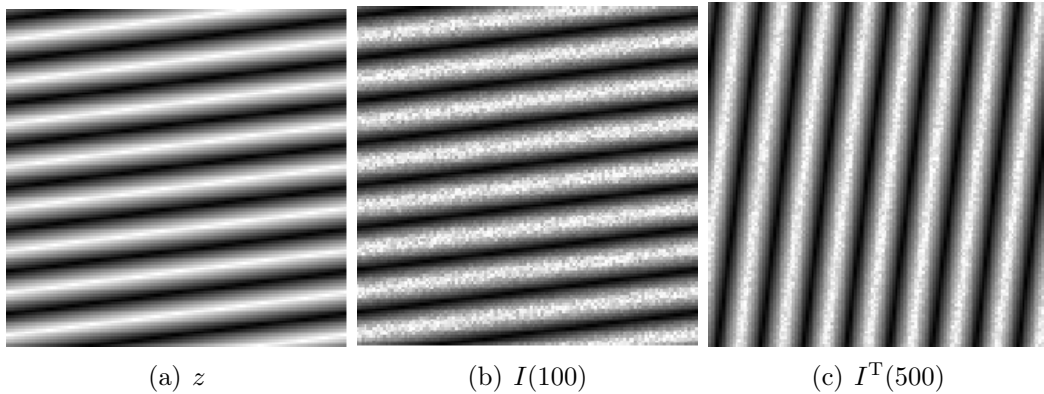


Figure 4.4. Ground truth, speckled, and speckled transposed versions.

$\{1, 100, 150, 200, 250, 300, 350, 400, 450, 500\}$ . The observed images  $I(L)$  are the pixel-wise product of  $z$  and  $w(L)$ . Fig. 4.4(b) shows the product  $I(100)$ .

Fig. 4.5 shows how the point in the  $H \times C$  plane varies according to the level of noise introduced. The ground truth (identified as “0”) has relatively low entropy and is close to the maximum complexity (the continuous line is the upper bound). This behavior is typical of deterministic sequences. Observe that when we inject single-look speckle ( $L = 1$ ), the entropy increases along with the complexity. Thus, the technique is able to identify the deterministic component even when it is embedded in the strongest possible speckle noise. The point “1” shifts towards “0” when the signal-to-noise progressively increases.

We also verified that the points in the  $H \times C$  plane are almost insensitive to rotations. Fig. 4.4(c) shows the transpose of  $I(500)$ . In all cases, the coordinates  $(h, c)$  of the transposed noisy images were equal, up to the fourth decimal place, to those of the original version.

These experiments provide evidence that the WATG mapping is little sensitive to rotations (thanks to the use of Hilbert-Peano curves), and that it can identify the presence of underlying structural information in the presence of varying levels of speckle.

## 4.5.2 Analysis of ordinal patterns methods

Fig. 4.3 shows examples of the ocean, forest, urban, and pasture both as image patches and as 1-D sequences after the linearization process.

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

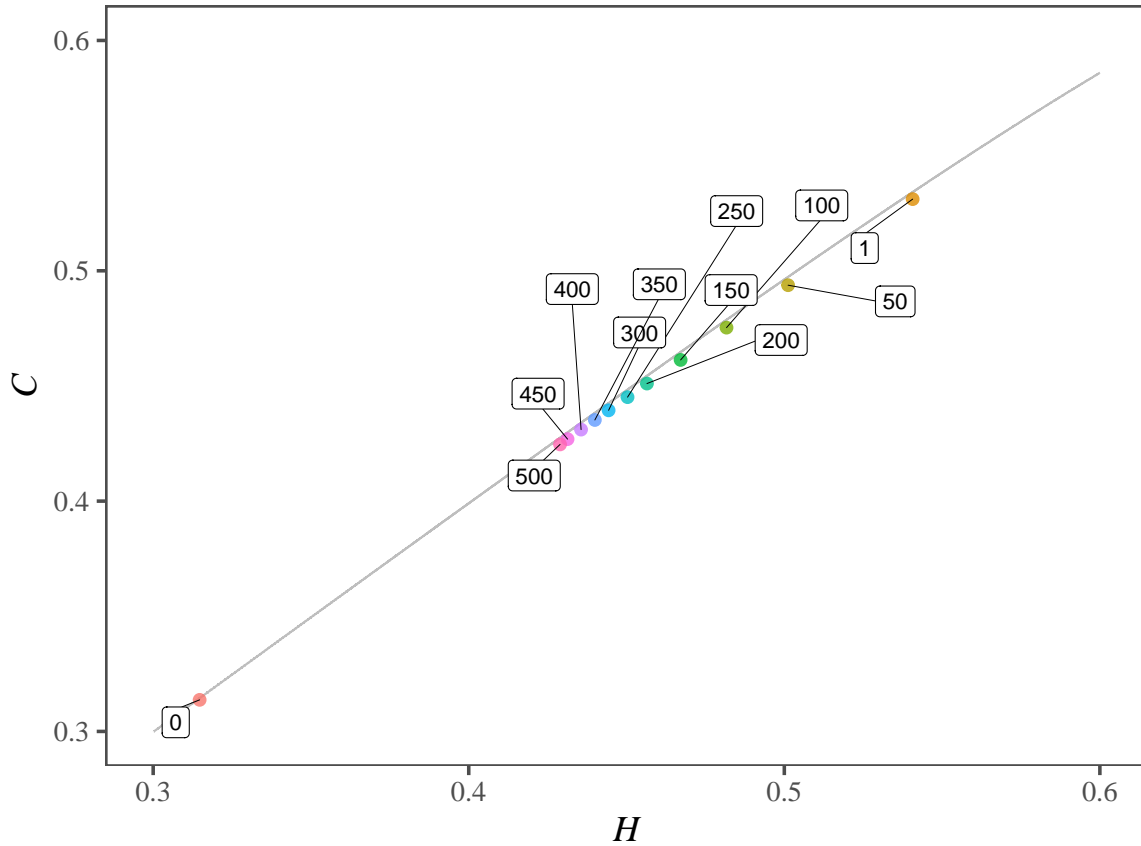


Figure 4.5. Modifications to the  $H \times C$  Plane features by adding different multiplicative noises.

amplitude, the proposed method is able to increase, compared to traditional methods, the granularity of information captured by ordinal patterns.

As already described in Section 3, our proposal weights the edges in terms of the difference of amplitudes. As expected, the greatest impact is observed on the transition graphs obtained from urban areas. The urban area 1-D signal shown in Fig. 4.3 has the largest dynamic range. Fig. 4.6 shows how this information alters the weights of the transition graph. Notice, in particular, that  $(v_{\pi_{123}^3}, v_{\pi_{123}^3})$  almost doubled, while  $(v_{\pi_{312}^3}, v_{\pi_{231}^3})$  and  $(v_{\pi_{213}^3}, v_{\pi_{132}^3})$  became negligible.

Fig. 4.9 shows the impact of using the data amplitude information on the weights of the transition graphs. Bandt-Pompe symbolization was the first method based on ordinal patterns proposed in the literature. As shown in Fig. 4.9(f), it provides limited separation of textures. The transition graphs (Fig. 4.9(g)) improve the spread of the features, but with some amount of confusion. Our proposal, shown in Fig. 4.9(k), 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.

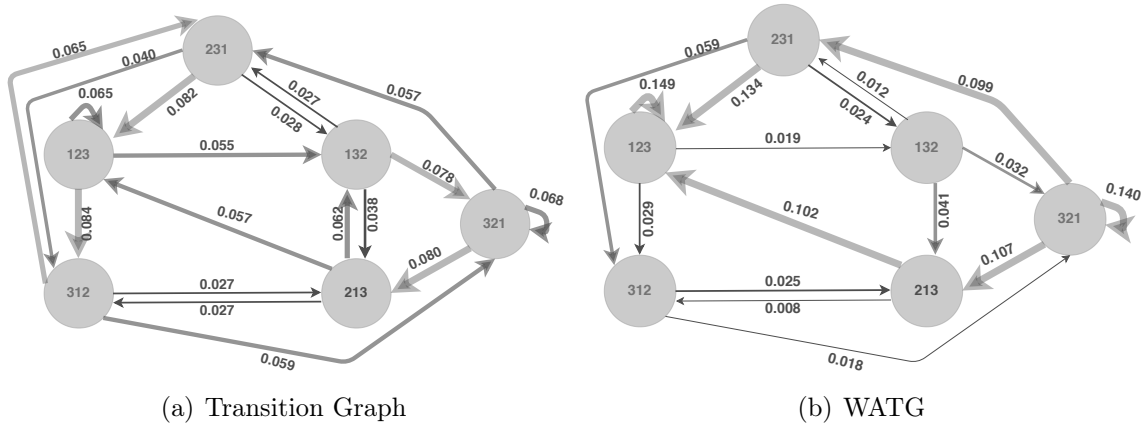


Figure 4.6. Difference of edges weights between the transition graph and the weighted graph of ordinal patterns transitions; urban area, with dimension 3 and delay 1.

### 4.5.3 Experiments on sliding window selection

In this section, we analyze the parameters of the proposed method and its impact on textures classification. McCullough et al. (2015) report that inadequate values may hinder important characteristics of the phenomenon under analysis. The two parameters of the transition graph are the dimension  $D$ , and the delay  $\tau$ . In the experiments below, we present the results in the classification using different values of these parameters.

The classification method's performance based on ordinal patterns is sensitive to window size, the embedding dimension, and the delay. In techniques based in Bandt-Pompe symbolization, for a fixed signal, as the size of the embedding dimension decreases, more ordinal patterns are produced. Therefore, we acquire a higher granularity of information about the dynamics of the system and, consequently, we capture more spatial dependencies between the elements.

Fig. 4.7 shows the ROC plane for different values of  $D \in \{3, 4, 5, 6\}$  and  $\tau \in \{1, 2, 3, 4, 5\}$  to select the best configuration. The configurations that extracted most information from the 1-D signal and, thus, that presented the best results in the experiments, are  $(D = 3, \tau = 1)$  and  $(D = 4, \tau = 1)$ . The technique, thus, shows its best performance choosing the parameters with the lowest computational cost.

Figure 4.8 shows the points in the  $H \times C$  produced by the same samples with all the parameters mentioned above. The spatial distribution of the points changes with the parameters and specific configurations promote better separation. This figure shows that the discrimination ability decreases with increasing  $\tau$ . Larger values of delay dilute the spatial dependence, as neighboring points in the sample tend to be

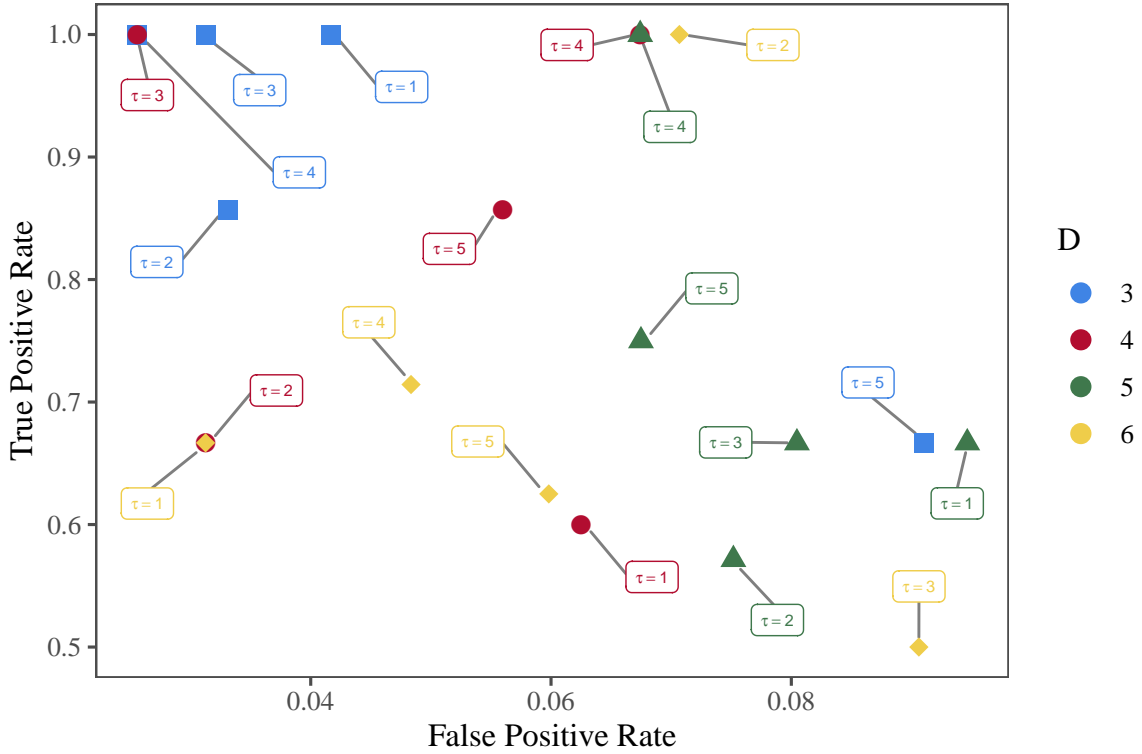


Figure 4.7. Evaluation of the sliding window parameters using ROC curve

more distant in the image. For this reason, we use  $\tau = 1$ . Figure 4.8 suggests that only one feature ( $H$  or  $C$ ) is sufficient to discriminate the classes studied. Although this is true for the experiments herein conducted, we opt to preserve the most common ordinal pattern analysis, which uses both features. As we studied only homogeneous patches, we still do not know how this approach performs with heterogeneous patches. For this last situation, we may need both features.

Considering  $\tau = 1$  (first column of Fig. 4.8), we also notice that  $D = 3$  produces the best separation among classes. Increasing  $D$  also increases the Statistical Complexity; this is noticeable for the Forest class. The other effect of considering larger values of  $D$  is an increased Entropy of Ocean and an undesirable overlap with Urban samples.

#### 4.5.4 Quantitative Evaluation

We present a comparison between our proposal and other methods for texture characterization and classification. We use the following ten methods: Gabor filters (Weldon et al., 1996), Histogram of oriented gradients (HOG) (Dalal and Triggs, 2005), Gray-level co-occurrence matrices (GLCM) (Kourgli et al., 2012), Speeded-Up Robust

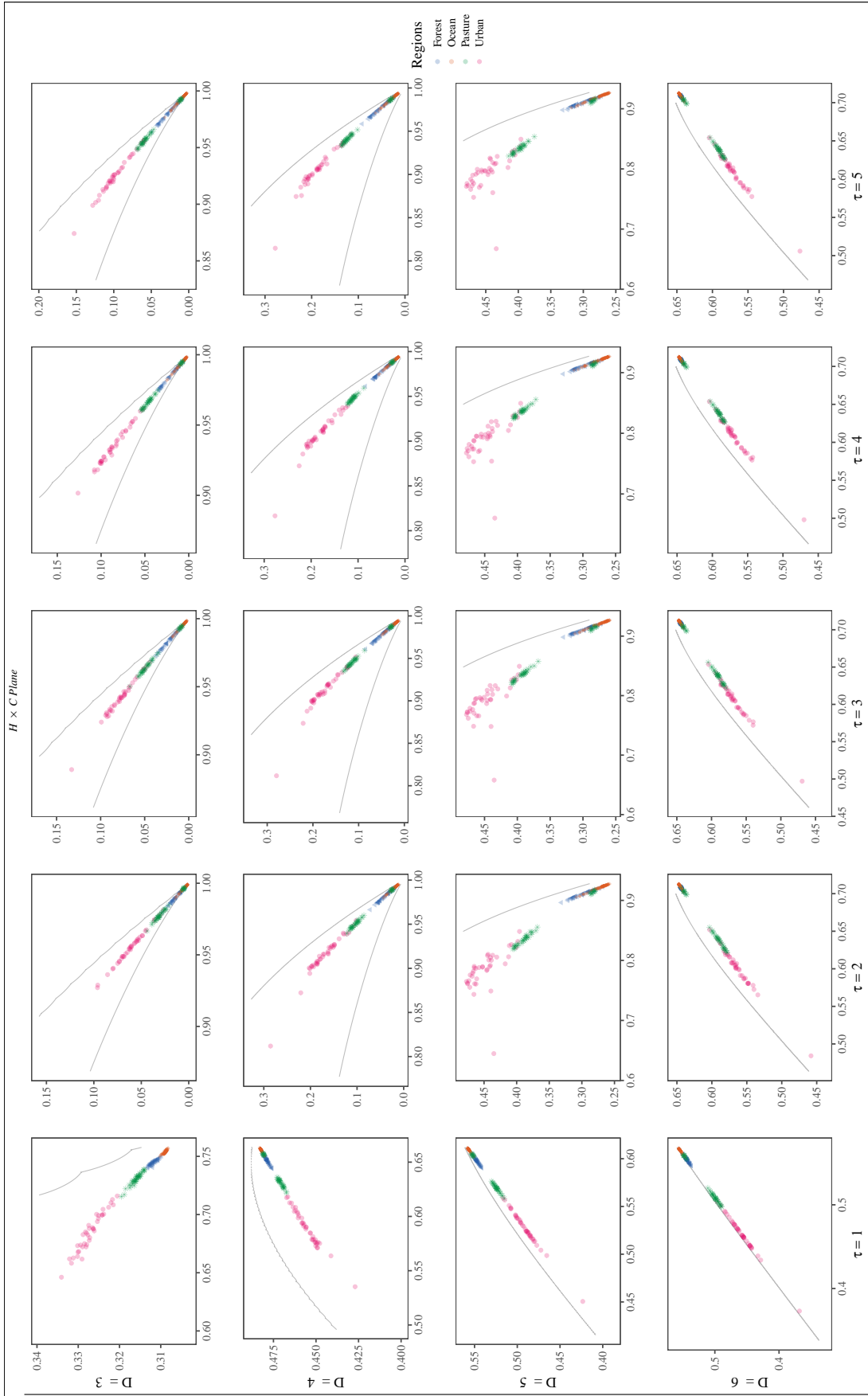


Figure 4.8. Characterization resulting in  $H \times C$  Plane from the application of the Hilbert-Peano curve in WATG on textures of different regions: Guatemala (forest), Cape Canaveral (ocean) and Munich (urban). The continuous curves correspond to the maximum and minimum values of  $C$  as a function of  $H$ .

Features (SURF) (Bay et al., 2006), Short Time Fourier Transform (STFT) (Portnoff, 1980) with SURF, Bandt-Pompe probability distribution (Bandt and Pompe, 2002b), Ordinal patterns transition graphs (Borges et al., 2019), Weighted Permutation Entropy (WPE) (Fadlallah et al., 2013), Fine-Grained Permutation Entropy (FGPE) (Xiao-Feng and Yue, 2009) with  $\alpha = 0.5$ , and Amplitude-Aware Permutation Entropy (AAPE) (Azami and Escudero, 2016) with  $A = 0.5$ . As in Guan et al. (2019), we computed four statistics from co-occurrence matrices: contrast, correlation, energy, and homogeneity. Likewise, we implemented the Gabor filters in five scales and eight orientations; using the energy, we obtained an 80-dimensional feature vector for each patch. For the HOG technique, we used image pixels divided into equal cells of  $3 \times 3$  pixels, and for each cell, we computed 6-bin histograms ranging from  $0^\circ$  to  $180^\circ$  or  $0^\circ$  to  $360^\circ$ .

The results of visualizing the features of these descriptors can be seen in Fig. 4.9. As we can see, the algorithms that combine ordinal characteristics with amplitude information (WPE, AAPE, and FGPE) reduce the intra-class distance, but not are good discriminators. Although the algorithms that have the largest descriptors (Gabor and HOG) do not present a good visualization when applying the dimensionality reduction technique, they presented excellent results in the quantitative evaluation.

We classified the features using the  $k$ -nearest neighbor algorithm with Euclidean distance, selecting the value of  $k$  with the automatic grid search method of the Caret R package (Kuhn, 2008). For validation, we used 10-fold cross-validation. More details about the classifier and the sampling can be seen in Mitchell (1997).

Table 4.1 presents the number of features each method produces, as well as its performance at classifying the 200 samples. We assessed the effectiveness of each approach using the following metrics. We used the first two metrics (Recall and Precision) to evaluate classifiers' per class performance and the last three metrics (Average Accuracy, Micro F1-score, and Macro F1-score) to evaluate the overall performance of the multi-class classifiers. We denote  $TP_i$ ,  $TN_i$ ,  $FP_i$ , and  $FN_i$  as the true positives, true negatives, false positives, and false negatives counts of a given class  $i$ , among a set of  $K$  classes, respectively.

- Recall or True Positive Rate of the class  $i$  ( $TPR_i$ ):

$$TPR_i = \frac{TP_i}{TP_i + FN_i}.$$

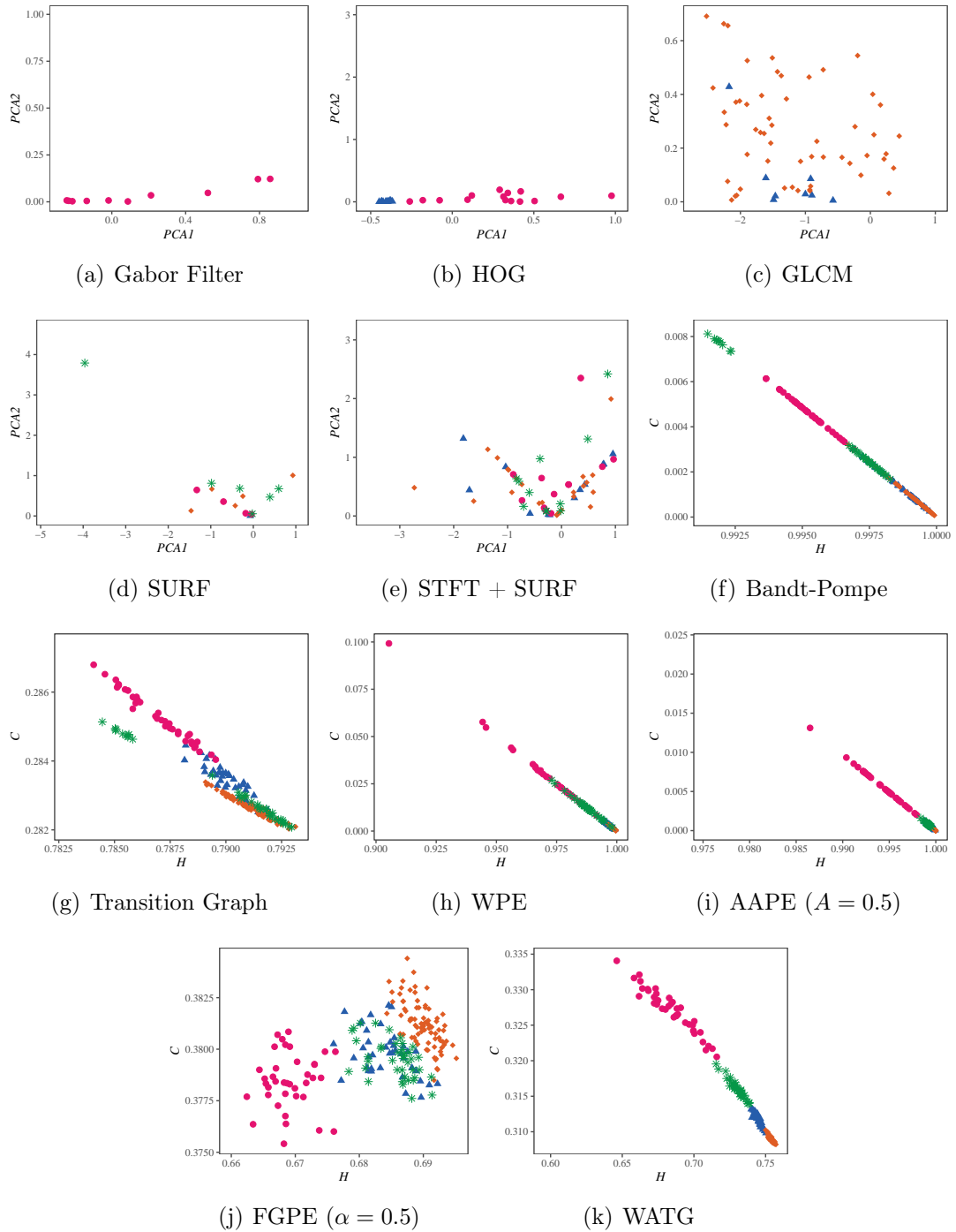


Figure 4.9. Bi-dimensional visualization of the descriptors obtained by our proposal and other methods of texture characterization and classification. For those who provide more than two features, we apply the PCA algorithm to obtain such a visualization.



Table 4.1. Experimental results using  $k$ -NN

Method	# features	TPR				PPV				AA	F1-Score $_{\mu}$	F1-Score $_M$
		Forest	Pasture	Ocean	Urban	Forest	Pasture	Ocean	Urban			
Gabor	80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
HOG	54	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
GLCM	32	0.833	1.000	1.000	0.833	1.000	0.857	0.923	1.000	0.967	0.980	0.970
SURF	1856	0.500	0.000	1.000	0.000	1.000	0.000	0.444	0.000	0.467	0.666	0.572
STFT + SURF	1856	0.166	0.000	0.833	0.166	0.250	0.000	0.416	0.500	0.300	0.462	0.292
Bandt-Pompe	2	0.333	1.000	0.750	1.000	0.500	0.857	0.750	0.857	0.600	0.776	0.633
Transition Graph	2	0.833	0.666	0.833	1.000	0.833	0.800	0.769	1.000	0.767	0.929	0.875
WPE	2	1.000	0.833	1.000	0.833	0.857	0.833	1.000	1.000	0.933	0.868	0.779
AAPE	2	0.666	1.000	1.000	1.000	1.000	0.857	0.923	1.000	0.833	0.947	0.896
FGPE	2	0.666	0.666	1.000	1.000	0.800	0.666	0.923	1.000	0.767	0.868	0.711
WATG	2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

- Precision or Positive Predictive Value of the class  $i$  (PPV $_i$ ):

$$\text{PPV}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i}.$$

- Average Accuracy (AA):

$$\text{AA} = \sum_{i=1}^K \left\{ \frac{\text{TP}_i + \text{TN}_i}{\text{TP}_i + \text{TN}_i + \text{FP}_i + \text{FN}_i} \right\}$$

- Micro F1-score:

$$\text{PPV}_{\mu} = \frac{\sum_{i=1}^K \text{TP}_i}{\sum_{i=1}^K \text{TP}_i + \sum_{i=1}^K \text{FP}_i}$$

$$\text{TPR}_{\mu} = \frac{\sum_{i=1}^K \text{TP}_i}{\sum_{i=1}^K \text{TP}_i + \sum_{i=1}^K \text{FN}_i}$$

$$\text{F1-score}_{\mu} = 2 \frac{\text{PPV}_{\mu} \times \text{TPR}_{\mu}}{\text{PPV}_{\mu} + \text{TPR}_{\mu}}.$$

- Macro F1-score:

$$\text{PPV}_M = \frac{1}{K} \sum_{i=1}^K \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i}$$

$$\text{TPR}_M = \frac{1}{K} \sum_{i=1}^K \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i}$$

$$\text{F1-score}_M = 2 \frac{\text{PPV}_M \times \text{TPR}_M}{\text{PPV}_M + \text{TPR}_M}.$$

Table 4.1 shows that, among the methods of weighting ordinal patterns, FGPE produced the worst results: AA = 76.7%, F1-score $_{\mu}$  = 86.8%, and F1-score $_M$  =

71.1%. WPE also produced a low F1-score, but it produced consistently better results in the other metrics, presenting  $AA = 93.3\%$ . AAPE achieved one of the best F1-score results:  $AA = 83.3\%$ ,  $F1\text{-score}_\mu = 94.7\%$ , and  $F1\text{-score}_M = 89.6\%$ . WATG, considering the transition graph of ordinal patterns, can better describe the textures presented, as it achieves the best performance achievable in all metrics: 100%.

STFT + SURF produced the worst results:  $AA = 30.0\%$ ,  $F1\text{-score}_\mu = 46.2\%$ , and  $F1\text{-score}_M = 29.2\%$ . SURF alone provided a better performance:  $AA = 46.7\%$ ,  $F1\text{-score}_\mu = 66.6\%$ , and  $F1\text{-score}_M = 57.2\%$ . HOG and Gabor filters achieved the highest success rates among all the handcrafted methods here considered:  $AA = 100\%$ ,  $F1\text{-score}_\mu = 100\%$ , and  $F1\text{-score}_M = 100\%$ . However, WATG achieves that same performance using only 2 features. This reduction implies less computational power requirement and avoids the curse of dimensionality (Altman and Krzywinski, 2018). Moreover, the features it is based upon are fully interpretable.

## 4.6 Conclusions

We presented and assessed a new method of analysis and classification of SAR image textures. This method consists of three steps: (1) linearization, (2) computing the Weighted Ordinal Pattern Transition Graph, and (3) obtaining Information Theory descriptors. A simple  $k$ -NN algorithm applied to the pairs Entropy-Statistical Complexity classifies the data with 100% performance. In addition to such perfect separation among urban, pasture, ocean, and forest areas, the proposed descriptors are interpretable in terms of the degree and structure of the spatial dependence among observations.

Experiments using patches from UAVSAR images showed that the proposal performs better than GLCM, Bandt-Pompe, Transition Graphs, SURF, STFT + SURF, and other techniques which also employ amplitude information in the analysis of ordinal patterns. Our approach provides the same quality of results obtained with Gabor filters and HOG. However, while Gabor filters employ 80 features and HOG uses 54 features, our proposal requires only two. This such reduced dimensionality consists of 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 differences between the classes of regions analyzed. Secondly, for machine learning algorithms, the smaller the number of dimensions, the faster the training process is, and the less storage space is required. Thirdly, we managed to avoid overfitting, a recurring problem in data with high dimensionality.

We also observed that only one feature ( $H$  or  $C$ ) is enough to discriminate the classes with the same reported performance. We opted to preserve both features because we consider that this study shed light on a novel way of SAR image analysis. Thus, we preferred to stick to the most common ordinal pattern analysis using the  $H \times C$  plane. As future work, we consider investigating the discriminative power of these features in more complex situations.

Our approach is robust to rotations and the presence of speckle noise. The behavior showed in Figure 4.5 shows that our approach can capture the speckle contamination adequately.

Since the application of this work is limited to texture patches from homogeneous regions, we aim to study the possible impacts of heterogeneous areas, such as mixed culture and urban regions.

## 4.7 Reproducibility and Replicability

Following the guidelines presented in Ref. Frery et al. (2020), the text, source code, and data used in this study are available at the *SAR-WATG* repository <https://github.com/EduardaChagas/SAR-WATG>. The information includes a link to download the 200 labeled samples we employed in the analysis.

# Chapter 5

## HC-PCA Confidence Regions

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.

To investigate the power of representation of the proposed confidence region, we defined as a first application a review of the results obtained in the literature using the  $H \times C$  plane in pseudo-random sequences. Thus, the necessary input to our algorithm consists of sequences of true random generated by physical procedures. As a result, we verify how the sequences previously analyzed behaved in this new analysis scenario.

Below we list the main strategies and sequences analyzed by the literature for the study of randomness using the information theory descriptors. Soon after, the test and characterization framework is presented.

## 5.1 Complexity-Entropy plane in the literature

The first works on the characterization of white noises with permutation entropy arose from the need to discriminate them in relation to chaotic maps (Rosso et al., 2013; Xiong et al., 2020).

Stationary time series can be decomposed into two main components:

1. **Deterministic**: it is described by a linear combination of its own past,
2. **Random**: it is a component of the moving average of finite order.

In this way, chaotic systems produce sequences composed of a physical structure, easily captured by measures of complexity. Thus, it was found that the measure of statistical complexity was able to efficiently quantify the performance of pseudorandom number generators, expanding the possibilities of using information theory descriptors with the Bandt-Pompe symbolization (Larrondo et al., 2002; González et al., 2005; Larrondo et al., 2006).

Table 5.1 presents a summary of the main works in the literature that perform analysis of non-chaotic algorithmic generators, according to their features in  $H \times C$ . We also provide the length  $T$  and embedding dimension  $D$  of the time series under scrutiny. The following algorithmic generators were analyzed:

- Mother RNG, available in Marsaglia website (Marsaglia, 1994) (MOT);
- Multiple with carry RNG (MWC) (Marsaglia, 1994);
- Combo RNG (COM) (Marsaglia, 1994);
- Lehmer RNG (LEH) (Payne et al., 1969);
- Fractional Gaussian noise with  $\alpha = 0$  (fGn) (Bardet et al., 2003);
- $f^{-k}$  noise with  $k = 0$  (Larrondo, 2012);
- Linear Congruential Generator (LCG) (Knuth, 1997).

None of these works provide  $p$ -values or hypothesis tests for their analysis. The focus of the articles is finding the set of descriptors that best discriminates chaos from noise. The authors make assessments about the randomness of sequence on ad hoc visual inspection of the point's location in the  $H \times C$  plane. Our work fills such a gap for finite sequences of white noise and proposes a methodology that can be extended to any other situation.

Table 5.1. Result of the main works of white noise sequences analysis in the  $H \times C$  plane.

Reference	PRNG	$T$	$D$	$H$	$C$	Considered random?
Larrondo et al. (2013)	MOT	NA	6	$\cong 0.9969$	$\cong 0$	no
González et al. (2005)	MWC	65536	NA	$\cong 1$	0.3	yes
	MOT	65536	NA	$\cong 1$	0.3	yes
	COM	65536	NA	$\cong 1$	0.05	yes
Larrondo et al. (2006)	LEH	$5 \times 10^6$	5	$\cong 1$	$10^{-4}$	yes
	MOT	$5 \times 10^6$	5	$\cong 1$	$10^{-4}$	yes
	MWC	$5 \times 10^6$	5	$\cong 1$	$10^{-4}$	yes
Rosso et al. (2013)	LCG	$1 \times 10^7$	6	0.997871	0.005101	no
Xiong et al. (2020)	fGn	$2 \times 10^{17}$	6	$\cong 1$	$\cong 0$	yes
	$f^{-k}$	$2 \times 10^{17}$	6	$\cong 1$	$\cong 0$	yes

## 5.2 Proposed Method

In this section, we formalize the task of building a confidence region in the Entropy-Complexity manifold. Then, we present our proposal to change space through the algorithm of the principal components analysis. Our goal is to find a latent space representative of the data, without the restrictions of a curvilinear space. Through this new representation of the data, we calculate empirical regions with different levels of confidence. Finally, after calculating these regions, we build a test statistic that determines the probability that a given sequence belongs to the distribution of the points provided.

### 5.2.1 Overall Framework

The structure of our proposal consists of two steps:

- **Empirical confidence region:** With the data present in a Euclidean plane, we can easily calculate empirical regions that involve the data with a certain level of confidence.
- **Construction of a test statistic:** To measure the similarity of new data sequences with the empirical points, a test statistic was proposed. By acquiring a p-value less than 0.05, we can reject the null hypothesis, which states that such data belong to the empirical probability distribution used for the construction of the confidence region.

## 5.2.2 Empirical Confidence Regions and $p$ -values

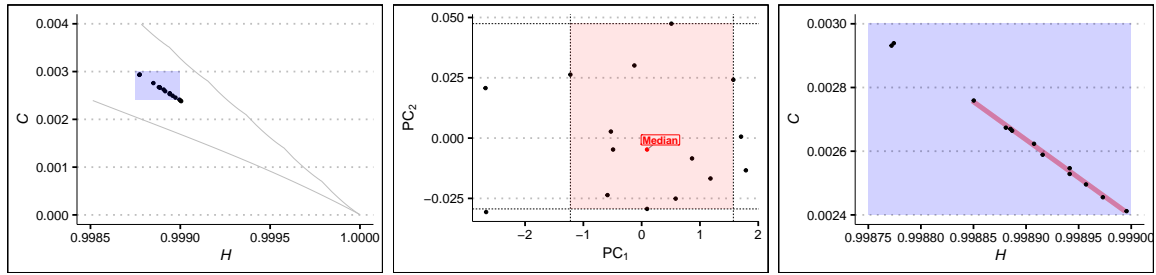
Our first approaches to analyzing sequences of points in the  $H \times C$  plane produced by TRWNS verified that they, and usual transformations, are far from bivariate Gaussian and generalized Hyperbolic distributions (Schmidt et al., 2006). Different types of regression models of  $C$  explained by  $H$  did not produce acceptable results. Thus, we adopted a non-parametric approach and made an empirical analysis of the data obtained from physical sources for using them as our reference in the search for confidence regions and  $p$ -values.

Let  $\underline{x} = (x_1, x_2, \dots, x_N)$  be  $N$  times series of length  $T$ , and define an embedding dimension  $D$ . In the sequel, whenever possible, we will omit  $T$  and  $D$ . For each  $n = 1, 2, \dots, N$ , the time series  $x_n$  is mapped onto the point  $(h_n, c_n)$  in the  $H \times C$  plane, thus  $\underline{hc} = ((h_1, c_1), (h_2, c_2), \dots, (h_N, c_N))$  are the points that correspond to the  $N$  time series. Fig. 5.1(a) illustrates this step. We will obtain confidence regions and  $p$ -values from  $\underline{hc}$ .

The first step consists in finding and applying the principal components transformation to  $\underline{hc}$ . With this, we obtain the set of uncorrelated points  $\underline{uv} = ((u_1, v_1), (u_2, v_2), \dots, (u_N, v_N))$ , in which  $u_n$  and  $v_n$  are the first and second principal components of  $h_n$  and  $c_n$ , respectively. This projection allows us to obtain a “central” point of the data set, around which we will build a rectangular box containing  $100(1 - \alpha)\%$  of the observations, where  $\alpha$  is the significance level analyzed. 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 and Har-Peled (2020).

For simplicity, and without loss of generality, assume  $N$  is odd.

1. Find the indexes that sort the values of the first principal component  $\mathbf{u} = (u_1, u_2, \dots, u_N)$  in ascending order:  $\mathbf{r} = (r_1, r_2, \dots, r_N)$ , 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  $\mathbf{P}' = (h', v')$ . Call the corresponding time series “emblematic time series.”
3. Find the point  $(u, v)$  whose first principal component is the quantile  $\alpha/2$ :  $(u_{r_{\lceil N\alpha/2 \rceil}}, \cdot)$ .
4. Find the point  $(u, v)$  whose first principal component is the quantile  $1 - \alpha/2$ :  $(u_{r_{\lceil N(1-\alpha/2) \rceil}}, \cdot)$ .



(a) Mapping true white noise random sequences onto the  $H \times C$  plane. (b) Transformation of the points in the  $H \times C$  plane by Principal Components, and determination of minimal boxes. (c) Inverse transformation from the Principal Components plane to the  $H \times C$  plane.

Figure 5.1. Outline of the methodology used for the construction of the confidence regions.

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  $\mathbf{P}_1 = (h_{v_1}, c_{v_1})$ ,  $\mathbf{P}_2 = (h_{v_2}, c_{v_2})$ ,  $\mathbf{P}_3 = (h_{v_3}, c_{v_3})$  and  $\mathbf{P}_4 = (h_{v_4}, c_{v_4})$ .

Fig. 5.1 illustrates these steps. Fig. 5.1(a) shows the points produced by TRWNS in the  $H \times C$  plane. The blue box includes a certain percentage of points, with sides parallel to the  $H$  and  $C$  axes. The area in the  $H \times C$  plane overestimates the desired proportion and may include “unacceptable” points. Fig. 5.1(b) shows the previous points projected onto the principal components space (steps 2 to 7). The red box includes the same percentage of desired points, with axes parallel to the first and second principal components. We highlighted in red the point whose first principal component is the median of the observed values. Fig. 5.1(c) shows the result of projecting back the red box from the principal components space to the  $H \times C$  plane (step 9). The comparison of the red and blue boxes shows that the area has been reduced, thus improving the test’s power.



Algorithm 2 provides details on how we obtain the confidence regions, defined by a set of points  $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4$ , for each  $D \in \mathcal{D}$ , each  $T \in \mathcal{T}$ , and each significance level  $\alpha$ . We also obtain the “emblematic point”  $\mathbf{P}'$ , a kind of median point in the  $H \times C$  plane for each situation.

```

input : A data base of true random values
input : The desired values of embedding dimension  $\mathcal{D}$ , sequence length  $\mathcal{T}$ ,
        and confidence levels  $\mathcal{A}$ 
output: Confidence regions as points in the  $H \times C$  plane
1 for each  $D \in \mathcal{D}$  do
2   for each  $T \in \mathcal{T}$  do
3     for each  $n = 1, 2, \dots, N$  do
4       build the time series  $p_n$  with unused values from the data base;
5       compute the point  $(h_n, c_n)$  in the  $H \times C$  plane that corresponds to
         $p_n$ ;
6     end
7     obtain  $PC(D, T)$ , the principal components transformation based on
        the points  $(h_1, c_1), (h_2, c_2), \dots, (h_N, c_N)$ , and its inverse  $PC^{-1}(D, T)$ ;
8     apply  $PC(D, T)$  to the points  $(h_1, c_1), (h_2, c_2), \dots, (h_N, c_N)$ , and
        obtain  $(u_1, v_1), (u_2, v_2), \dots, (u_N, v_N)$ ;
9     find the indexes  $\mathbf{r} = (r_1, r_2, \dots, r_N)$  that sort the values of the first
        principal component  $\mathbf{u} = (u_1, u_2, \dots, u_N)$  in ascending order;
10    find the point  $(u, v)$  whose first principal component is the median:
         $(u_{r_{(N+1)/2}}, \cdot)$ ;
11    apply the inverse principal components transformation  $PC^{-1}(D, T)$  to
         $(u, v)$ , and obtain  $\mathbf{P}' = (h', v')$ ; call the corresponding time series
        “emblematic time series”;
12    return  $\mathbf{P}'$ ;
13    for each confidence level  $\alpha \in \mathcal{A}$  do
14      find the point  $(u, v)$  whose first principal component is the
        quantile  $\alpha/2$ :  $(u_{r_{\lfloor N\alpha/2 \rfloor}}, \cdot)$ ;
15      find the point  $(u, v)$  whose first principal component is the
        quantile  $1 - \alpha/2$ :  $(u_{r_{\lfloor N(1-\alpha/2) \rfloor}}, \cdot)$ ;
16      the values  $u_{r_{\lfloor N\alpha/2 \rfloor}}$  and  $u_{r_{\lfloor N(1-\alpha/2) \rfloor}}$  are the rightmost and leftmost
        bounds of the box, respectively;
17      the bottom bound of the box is the smallest second principal
        component value whose first principal component is at least
         $u_{r_{\lfloor N\alpha/2 \rfloor}}$ ; denote this value  $v_{\min}$ ;
18      the top bound of the box is the largest second principal value
        whose first principal component is at most  $u_{r_{\lfloor N(1-\alpha/2) \rfloor}}$ ; denote this
        value  $v_{\max}$ ;
19      the corners of the box are  $(u_{r_{\lfloor N\alpha/2 \rfloor}}, v_{\min}), (u_{r_{\lfloor N\alpha/2 \rfloor}}, v_{\max}),$ 
         $(u_{r_{\lfloor N(1-\alpha/2) \rfloor}}, v_{\min})$  and  $(u_{r_{\lfloor N(1-\alpha/2) \rfloor}}, v_{\max})$ ;
20      apply the inverse principal components transformation
         $PC^{-1}(D, T)$  to these corners obtaining  $\mathbf{P}_1 = (h_{v_1}, c_{v_1}),$ 
         $\mathbf{P}_2 = (h_{v_2}, c_{v_2}), \mathbf{P}_3 = (h_{v_3}, c_{v_3})$  and  $\mathbf{P}_4 = (h_{v_4}, c_{v_4})$ ;
21      return  $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4$ ;
22    end
23  end
24 end

```

**Algorithm 2:** Determination of confidence regions and emblematic time series

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

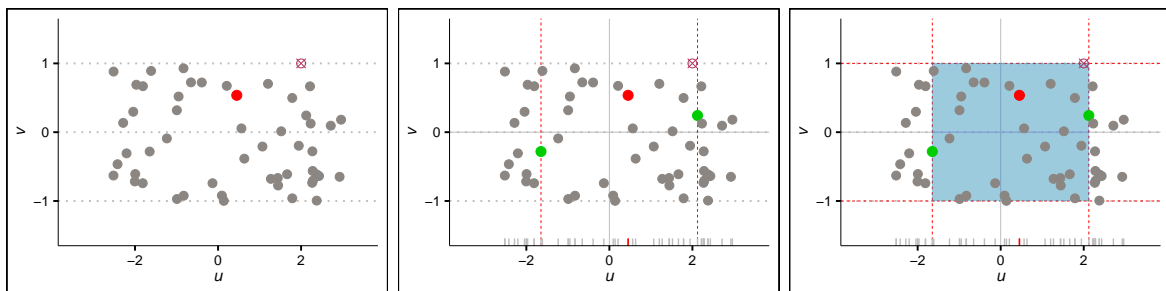
The procedure operates on the principal components space and consists of measuring the closeness between the “emblematic point” and the observed point. We are given a time series  $\mathbf{x}$  of size  $T$ , and we want its  $p$ -value when contrasted with TWNRS of the same size at embedding dimension  $D$ . 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  $\mathbf{x}$ , and obtain a new point  $(u_x, v_x)$ . The closer  $\mathbf{x}$  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 than  $u_x$ ; assume it corresponds to the quantile  $\eta_u$  of  $\underline{u} = (u_1, u_2, \dots, u_N)$ . By definition,  $\eta_u \geq 1/2$ .
2. the left side of the box is the  $1 - \eta_u$  quantile of  $\underline{u}$ .
3. the top side of the box is the smallest  $v_j$  which is larger than  $v_x$ ; assume it corresponds to the quantile  $\eta_v$  of  $\underline{v} = (v_1, v_2, \dots, v_N)$ . By definition,  $\eta_v \geq 1/2$ .
4. the bottom side of the box is the  $1 - \eta_v$  quantile of  $\underline{v}$ .

Fig. 5.2 illustrates these steps.

The definition of the box for the case  $u_x < u$  follows naturally, and is described in Algorithm 3. With this approach, we obtain the smallest box that (i) contains the new point, and (ii) 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  $H \times C$  plane; cf. Fig. 6.2. Algorithm 3 shows the details.



(a) Points produced by TRWNSs in the space of principal components.  
 (b) Points whose first principal component is the median (red) and side the box (green).  
 (c) The  $p$ -value of the new point is the proportion of points outside the box.

Figure 5.2. Outline of the methodology used to calculate the  $p$ -value. The new point is denoted as a crossed circle.

**input** : The sequence  $\mathbf{x}$  of length  $T$  to be contrasted to the null hypothesis  $\mathcal{H}_0$  that it is adherent to white noise

**input** : The embedding dimension  $D$

**input** :  $N$  points  $(h_1, c_1), (h_2, c_2), \dots, (h_N, c_N)$  of true white noise series of length  $T$ ; the principal components transformation  $\text{PC}(D, T)$  induced by these points; the points projected onto the  $H \times C$  plane:  $(u_1, v_1), (u_2, v_2), \dots, (u_N, v_N)$

**output**: An approximate  $p$ -value

- 1 find the point  $(u, v)$  whose first principal component is the median:  $(u_{r_{(N+1)/2}}, \cdot)$ ;
- 2 find the point  $(h, c)$  of the sequence  $\mathbf{x}$ ;
- 3 find the projection  $(u_x, v_x)$  of  $(h, c)$  onto the principal components space using  $\text{PC}(D, T)$ ;
- 4 **if**  $u_x > u$  **then**
- 5      $u_{r_{[N(1-\alpha/2)]}}$  is defined as the smallest element larger than  $u_x$ ;
- 6      $u_{r_{[N\alpha/2]}} \leftarrow 2u - u_{r_{[N(1-\alpha/2)]}}$ ;
- 7 **else**
- 8     **if**  $u_x < u$  **then**
- 9          $u_{r_{[N\alpha/2]}}$  is the largest minor element of  $u_x$ ;
- 10          $u_{r_{[N(1-\alpha/2)]}} \leftarrow 2u - u_{r_{[N\alpha/2]}}$ ;
- 11     **else**
- 12          $u_{r_{[N\alpha/2]}}$  and  $u_{r_{[N(1-\alpha/2)]}}$  is equal to  $u$ , the median point of the first principal component;
- 13     **end**
- 14 **end**
- 15 obtain the maximum values of the second component whose values of the first principal component are at least  $u_{r_{[N\alpha/2]}}$  and at most  $u_{r_{[N(1-\alpha/2)]}}$  and denote it  $v_{\max}$ ;
- 16 obtain the minimum values of the second component whose values of the first principal component are at least  $u_{r_{[N\alpha/2]}}$  and at most  $u_{r_{[N(1-\alpha/2)]}}$  and denote it  $v_{\min}$ ;
- 17 the corners of the box  $b_\alpha(h, c)$  are  $(u_{r_{[N\alpha/2]}}, v_{\min}), (u_{r_{[N\alpha/2]}}, v_{\max}), (u_{r_{[N(1-\alpha/2)]}}, v_{\min})$  and  $(u_{r_{[N(1-\alpha/2)]}}, v_{\max})$ ;
- 18 count  $n_{\mathbf{x}}$ , the number of points out of the  $N$  points which belong to  $b_\alpha(h, c)$ ;
- 19 **return**  $1 - n_{\mathbf{x}}/N$

**Algorithm 3:** Determination of the  $p$ -value of the sequence  $\mathbf{x}$  under  $\mathcal{H}_0$

# Chapter 6

## Case study: White Noise Confidence Regions

In this section, we will present the results obtained by the first proposal for the construction of empirical confidence regions in the Complexity-Entropy plane for white noise models. Through these regions, we want to check if the randomness present in such PRNGs can be captured by the descriptors even with the use of short sequences. As a consequence, we observed that the proposed methodology proved to be consistent and coherent, managing to capture the randomness of truly random sequences, reproducing the results obtained by generators previously analyzed in the literature, and proving to be robust in the addition of correlation structures.

### 6.1 Introduction

Several works have used deterministic and pseudorandom sequences aiming at understanding the properties of the points they produce in the  $H \times C$  plane. Martin et al. (2006b) analyzed the logistic chaotic map and discuss the boundaries of the  $H \times C$  plane. De Micco et al. (2009) studied chaotic components in pseudorandom number generators. Ravetti et al. (2014) tackled the often hard problem of distinguishing chaos from noise. Zunino et al. (2012b) used a multi-scale approach to analyze the interplay between chaotic and stochastic dynamics.

However, we were able to verify that the reported works carry out statistical studies using as a basis the descriptors of Information Theory. Motivated by previous works, in this work, we advance the state-of-the-art providing the first test for white noise points in the  $H \times C$  plane. In this proposal, the input is a sequence of true random observations generated by a physical-based procedure. We obtain the confidence

regions by performing an orthogonal projection of the data onto the space of principal components, thus eliminating the restrictions imposed by the bounded space of the Complexity-Entropy plane. Our contributions can be summarized as follows:

- We provide the first contribution in constructing a test in the Complexity-Entropy Plane: we provide confidence regions and  $p$ -values.
- We evaluate this test's size by analyzing random sequences generated by physical procedures and pseudorandom generators (PRNGs).
- We verify the test's power contrasting correlated noise time series.

## 6.2 Experimental Settings

We evaluated the performance of the proposed method in relation to a large set of random sequences provided by state-of-the-art pseudo-random number generators. In this section, we present the settings of the parameters that we use as a reference, the true random physical generators used to calculate the empirical distribution, and descriptive analysis of representative points in relation to the confidence regions.

### 6.2.1 Parameters Settings and Dataset

We conducted an ablation study to identify the influence of the parameters  $T$ ,  $D$ , and  $\tau$  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 not consider  $\tau$  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  $T \in \mathcal{T} = \{T = 10^3, 5 \times 10^4\}$ ,
- 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, 2002a).

Using this parametric space, we analyzed the different degrees of information captured by the ordinal patterns formed. For the construction of the confidence regions presented, we used:

- A set of 104 596 points in the  $H \times C$  plane, corresponding to sequences of length  $T = 1000$ , for each value of  $D \in \mathcal{D}$ , and

- a set of 2093 points in the  $H \times C$  plane, corresponding to sequences of length  $T = 50000$ , for each value of  $D \in \mathcal{D}$ .

We used the R platform (R Core Team, 2017, v. 4.0.3) for data generation and analyses, and the `ggplot2` library Wickham (2009) for generating the plots.

## 6.2.2 True Random Numbers

Random numbers are used in many fields, from gambling to cryptography, aiming to guarantee a secure, realistic or unpredictable behavior. Pseudo random results can be achieved by software in a deterministic way. But, some applications need actual random numbers (despite the somewhat elusive nature of actual randomness). Randomness can be observed in unpredictable real world phenomena like cathodic radiation or atmospheric noise.

In this study we used two sources of random numbers, here called true random, both from physical phenomena observation and measurement. The first is based on vacuum states to generate random quantum numbers, the setup consists of an ordinary laser source to generate a local oscillator (LO), a half-wave plate, a polarizing beam-splitter (BPS), and two balanced detectors working together adding or subtracting the photocurrents results in a quadrature measurement of the LO or vacuum state. The probability distribution of the vacuum state is binned into  $2^n$  equal parts (bins of same size), then, assigning a fixed bit combination of length  $n$  to each sample point in a given bin Gabriel et al. (2010). The second one is based on atmospheric noise captured by a cheap radio receiver, started as a gambling engine, the randomness comes from an ordinary radio receiver that has no filter for static unwanted sounds caused by atmospheric noise, but perfect for random purposes, developed over a distributed setup with some radios located at different geographical locations sending random bits to a cloud server who process data and hosts random.org, the history, and some other information could be found at Haahr (2018). We used  $54 \times 10^6$  4B words from each physical generator, which approximately amounts 200 MB of data.

In Fig. 6.1, we show how the samples used to build the empirical confidence regions are arranged in the  $H \times C$  plane.



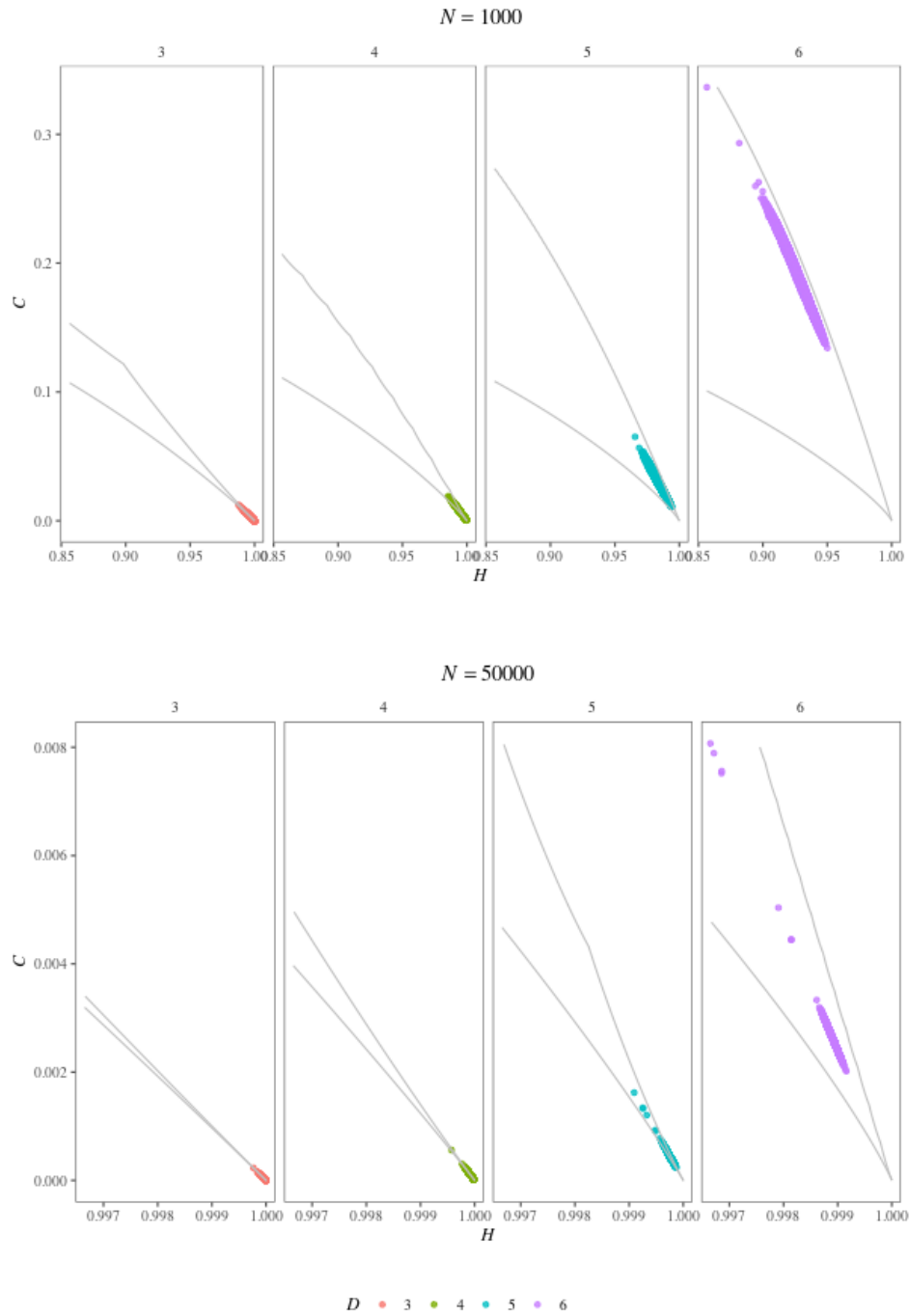
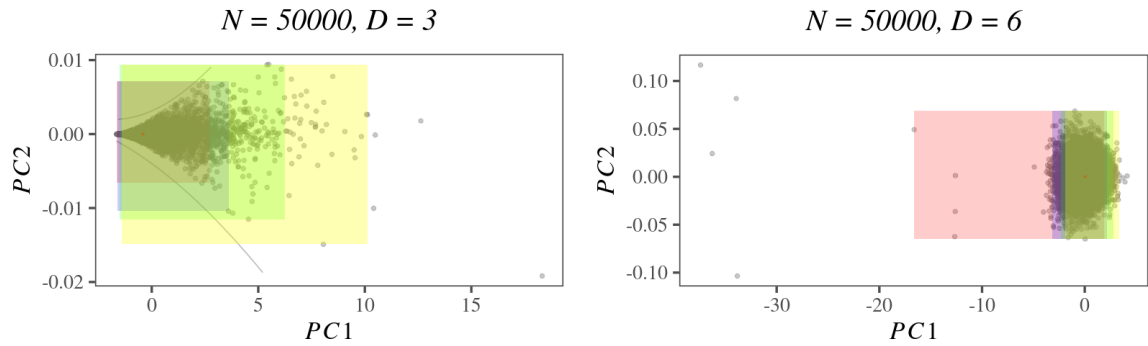


Figure 6.1. White noise samples considered during the construction of the proposed confidence regions.



(a) Points in the Principal Components plane for  $D = 3$  (b) Points in the Principal Components plane for  $D = 6$

Figure 6.2. Representation of true random white noise sequences of length  $T = 50000$  in the PCA space for  $D = 3$  and  $D = 6$ , and the quantiles of 90 %, 95 %, 99 %, and 99.9 %.

## 6.3 Experimental Results and Analysis

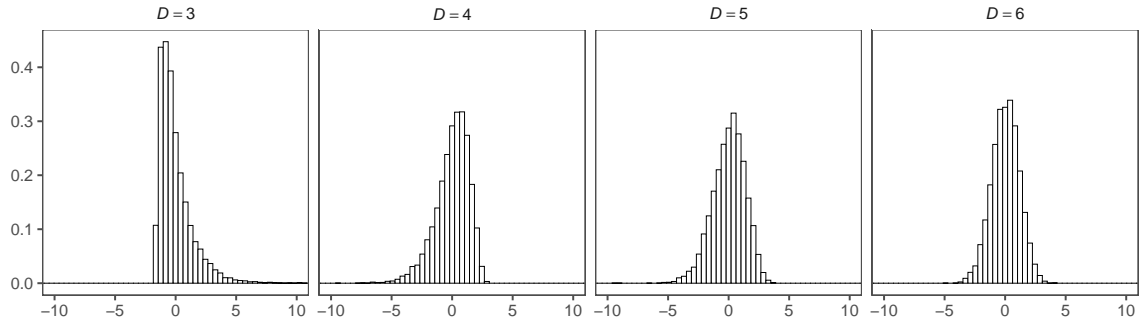
### 6.3.1 Descriptive analysis of empirical confidence regions

The regions used as a reference in this work are obtained through true random sequences, where we extract the empirical distribution of white noises in the Complexity-Entropy plane. Tables 6.1 and 6.2 list the coordinates in the  $H \times C$  plane of the emblematic point  $\mathbf{P}'$  and of the four points  $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4$  that define the confidence regions at 90 %, 95 %, 99 %, and 99.9 %, for  $D = 3, 4, 5, 6$  and  $N = 1000, 50000, \dots$ . The points are presented counterclockwise, starting with the one with the largest complexity.

Fig. 6.2 shows the results for  $T = 50000$  and  $D = 3, 6$  in the new principal components space, along with the quantiles of order 90 %, 95 %, 99 %, and 99.9 %. We also show the projection of the  $H \times C$  plane boundaries in this space, as well the median of each data set, the latter being represented as red dots. The confidence regions exceed the  $H \times C$  boundaries, but this issue does not compromise the test's size since no points can be observed outside such boundaries.

Fig. 6.2 also shows that the data are not evenly distributed among the axes of the first principal component. They tend to concentrate close to the point that corresponds to  $(1, 0)$  in the  $H \times C$  plane. As we use order statistics to define the confidence regions, this issue is also of little relevance for our results. Moreover, Fig. 6.3 shows that such asymmetry diminishes when the embedding dimension  $D$  increases.

As we can see in Fig. 6.3 in the new representation space produced by the PCA, the data are not evenly distributed among the axes of the first main component, maintaining the character presented in the  $H \times C$  plane, since such points tend to be

Figure 6.3. Histograms of the first principal component for  $D = 3, 4, 5, 6$ 

concentrated close to the point  $(1, 0)$ .

Table 6.1. Coordinates in the  $H \times C$  plane of the emblematic series and the points that define the confidence regions at 90 %, 95 %, 99 %, and 99.9 % for  $D = 3, 4, 5, 6$  and  $N = 1000$

		$N = 1000$			
$D$	Point	90 %	95 %	99 %	99.9 %
3	$P'$	(0.9992089, 0.0007800)			
	$P_1$	(0.9973334, 0.0025601)	(0.9967311, 0.0031343)	(0.9953009, 0.0045054)	(0.9931825, 0.0065387)
	$P_2$	(0.9974047, 0.0026304)	(0.9968219, 0.0032238)	(0.9954349, 0.0046375)	(0.9933704, 0.006724)
	$P_3$	(0.9999497, 0)	(0.9999398, 0)	(0.9999203, 0)	(0.9998925, 0)
	$P_4$	$(1, 5.17 \times 10^{-5})$	$(1, 6.12 \times 10^{-5})$	$(1, 8.45 \times 10^{-5})$	$(1, 0.0001104)$
4	$P'$	(0.9967032, 0.0043297)			
	$P_1$	(0.994364, 0.0081246)	(0.9937138, 0.0089796)	(0.9922575, 0.0108947)	(0.9902578, 0.0135243)
	$P_2$	(0.9939234, 0.0075452)	(0.9932534, 0.0083741)	(0.9917308, 0.0102022)	(0.9897312, 0.0128318)
	$P_3$	(0.9994791, 0.0013982)	(0.9991609, 0.0018166)	(0.9987924, 0.0023012)	(0.9985727, 0.0025901)
	$P_4$	(0.9990385, 0.0008188)	(0.9987005, 0.0012111)	(0.9982658, 0.0016087)	(0.9980461, 0.0018976)
5	$P'$	(0.9864873, 0.0245632)			
	$P_1$	(0.9811818, 0.0321294)	(0.9801289, 0.0340045)	(0.977917, 0.0377295)	(0.9753326, 0.0425299)
	$P_2$	(0.9827429, 0.0350291)	(0.9817117, 0.0369446)	(0.9796031, 0.0408613)	(0.9770187, 0.0456617)
	$P_3$	(0.9919707, 0.0120896)	(0.9909376, 0.0139279)	(0.9898161, 0.0156277)	(0.9892599, 0.0166608)
	$P_4$	(0.9935319, 0.0149893)	(0.9925204, 0.016868)	(0.9915021, 0.0187595)	(0.9909459, 0.0197926)
6	$P'$	(0.9296429, 0.1841438)			
	$P_1$	(0.9121895, 0.2201993)	(0.9105951, 0.2239294)	(0.9105951, 0.2239294)	(0.9077672, 0.2305874)
	$P_2$	(0.9146048, 0.2260776)	(0.9130413, 0.2298829)	(0.9130413, 0.2298829)	(0.9102595, 0.2366531)
	$P_3$	(0.9443868, 0.1418373)	(0.9419202, 0.1476904)	(0.9396577, 0.1531967)	(0.9383611, 0.1561279)
	$P_4$	(0.9468021, 0.1477156)	(0.9443663, 0.1536439)	(0.9421039, 0.1591502)	(0.9408534, 0.1621937)

### 6.3.2 Test Size

To analyze the efficiency of the confidence region calculated, we tested its applicability on a set of true random data generated physically not used by the algorithm during its construction. We assessed the size of the test by contrasting 100 new TWNRS for

Table 6.2. Coordinates in the  $H \times C$  plane of the emblematic series and the points that define the confidence regions at 90 %, 95 %, 99 %, and 99.9 % for  $D = 3, 4, 5, 6$  and  $N = 50000$

		$N = 50000$			
$D$	Point	90 %	95 %	99 %	99.9 %
3	$P'$	(0.9999853, $1.45 \times 10^{-5}$ )			
	$P_1$	(0.9999489, $5.06 \times 10^{-5}$ )	(0.9999384, $6.11 \times 10^{-5}$ )	(0.9999079, $9.11 \times 10^{-5}$ )	(0.9998625, 0.0001361)
	$P_2$	(0.9999487, $5.04 \times 10^{-5}$ )	(0.9999382, $6.09 \times 10^{-5}$ )	(0.9999077, $9.09 \times 10^{-5}$ )	(0.9998622, 0.0001358)
	$P_3$	(0.9999998, $4 \times 10^{-7}$ )	(0.9999994, $9 \times 10^{-7}$ )	(0.9999982, $2 \times 10^{-6}$ )	(0.9999973, $3 \times 10^{-6}$ )
	$P_4$	(0.9999996, $2 \times 10^{-7}$ )	(0.9999991, $7 \times 10^{-7}$ )	(0.999998, $1.8 \times 10^{-6}$ )	(0.999997, $2.7 \times 10^{-6}$ )
4	$P'$	(0.9999394, $7.94 \times 10^{-5}$ )			
	$P_1$	(0.9999684, $3.98 \times 10^{-5}$ )	(0.9999725, $3.44 \times 10^{-5}$ )	(0.9999783, $2.68 \times 10^{-5}$ )	(0.9999833, $2.02 \times 10^{-5}$ )
	$P_2$	(0.9999696, $4.13 \times 10^{-5}$ )	(0.9999737, $3.6 \times 10^{-5}$ )	(0.9999795, $2.83 \times 10^{-5}$ )	(0.9999845, $2.18 \times 10^{-5}$ )
	$P_3$	(0.9998075, 0.0002508)	(0.9998506, 0.0001942)	(0.9998756, 0.0001615)	(0.9998889, 0.000144)
	$P_4$	(0.9998087, 0.0002524)	(0.9998518, 0.0001958)	(0.9998768, 0.000163)	(0.9998901, 0.0001456)
5	$P'$	(0.9997616, 0.0004264)			
	$P_1$	(0.9998172, 0.0003232)	(0.9998259, 0.0003075)	(0.9998428, 0.0002774)	(0.9998573, 0.0002517)
	$P_2$	(0.9998194, 0.0003273)	(0.9998282, 0.0003116)	(0.999845, 0.0002814)	(0.9998593, 0.0002553)
	$P_3$	(0.9994812, 0.0009246)	(0.9996371, 0.0006455)	(0.9996703, 0.0005862)	(0.9996884, 0.000554)
	$P_4$	(0.9994834, 0.0009286)	(0.9996394, 0.0006495)	(0.9996725, 0.0005901)	(0.9996904, 0.0005576)
6	$P'$	(0.9989108, 0.0026093)			
	$P_1$	(0.9990169, 0.002336)	(0.9990368, 0.002288)	(0.9990736, 0.0021997)	(0.9991069, 0.0021197)
	$P_2$	(0.9990249, 0.0023554)	(0.9990449, 0.0023074)	(0.9990817, 0.0022191)	(0.999115, 0.0021392)
	$P_3$	(0.9978983, 0.0050219)	(0.998714, 0.0030633)	(0.998765, 0.0029407)	(0.9987884, 0.0028845)
	$P_4$	(0.9979064, 0.0050413)	(0.998722, 0.0030827)	(0.9987731, 0.0029601)	(0.9987965, 0.0029039)

each situation of  $D = 3, 4, 5, 6$  and of  $\alpha = 0.01, 0.05$ . Table 4.1 and Fig. 6.4 show the results.

On the one hand, long series ( $T = 50000$ ) present a good size for every embedding dimension. On the other hand, short series ( $T = 1000$ ) exhibit only one situation with a noticeable divergence between the expected and the observed size: the test rejects 13 % of the 100 series when  $D = 6$ . In contrast, we expected 1 % of rejection. This might be because, in this case, the condition  $D! \ll T$  is not respected. Notice that the wrongly rejected TWNRS are all close to the point  $(1, 0)$ .

We may then conclude that the test has good empirical size, provided  $D! \ll T$ , a condition that does not hold for  $D = 6$  and  $T = 1000$ .

### 6.3.3 Test Power

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. Several studies in the literature have used this approach for identifying and characterizing randomness.

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

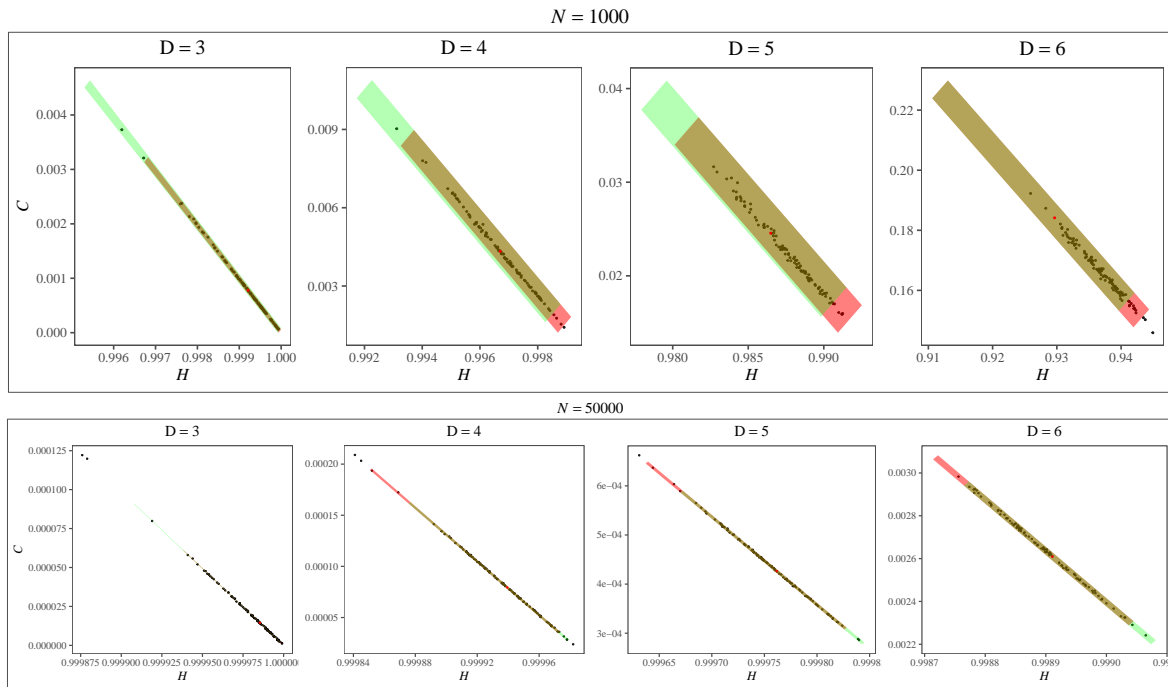


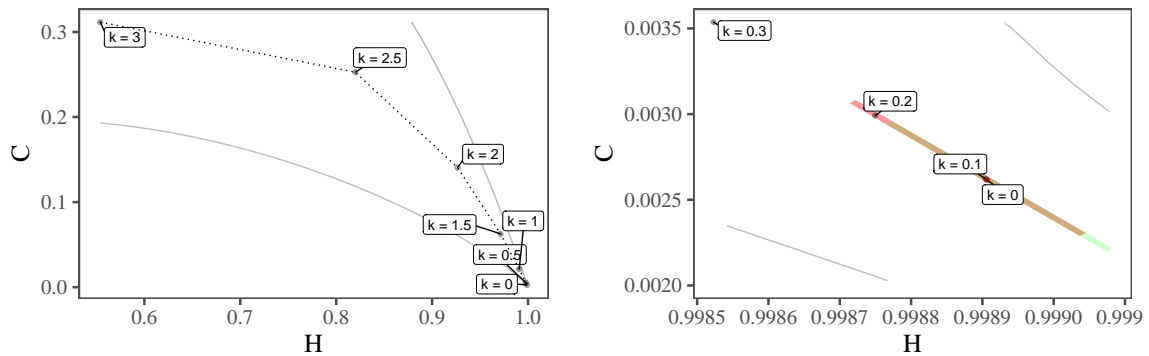
Figure 6.4. Results of the analysis behavior of true random noises in the regions of confidence built.

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. 6.5(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. 6.5(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  $k > 3/10$ .



(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  $k = 1/10, 1/5, 3/10$  along with the confidence regions  $1/2, 1, 3/2, 2, 5/2, 3$ . (b) Points of the emblematic white noise ( $k = 0$ ) and its  $f^{-k}$  correlated noise versions, with  $k = 0.3, 0.2, 0.1$  for white noise.

Figure 6.5. Analysis of the test power with correlated  $f^{-k}$  noise.

### 6.3.4 Revisiting the White Noise Hypothesis in the Literature

In this section, we compare the performance of our test with that of previous analyses that employ the Complexity-Entropy plane. To this aim, we produced 100 sequences of length  $T = 5 \times 10^4$  for each generator and computed the  $p$ -value for each  $D = \{3, 4, 5, 6\}$ . Previous results are shown in Table 5.1, and ours are in Table 6.3. We grouped our results in those that rejected (R) the null hypothesis and those that did not reject it (NR).

Comparing Tables 5.1 and 6.3, we see that our test captures adequately the random dynamics of the sequences produced by most of the analyzed generators. It is noteworthy that the generator Combo RNG sequences only pass our white noise test for  $D = 3$ . In higher embedding dimensions, as we consider longer words, the sequences produced by this generator are not labeled as white noise.

## 6.4 Conclusions

We presented and evaluated the first test for white noise in the Complexity-Entropy plane. Our proposal is based on two stages: (1) building non-parametric empirical confidence regions in the principal components space and mapping these boxes back to the  $H \times C$  plane. (2) computing an approximate  $p$ -value for a given sample by comparing it with points produced by true white noise random sequences (TWNRS). We obtained the TWNRS with data from physical devices.

Our test has a good size, mostly with long TWNRS. We also determined the

Table 6.3. Results of the sequences generated by the main PRNGs in the literature. The sequences have length  $T = 5 \times 10^4$ .

Algorithm	$D$	$p$ -value	HC-PCA	Algorithm	$D$	$p$ -value	HC-PCA
MOT	3	0.305	NR	LEH	5	0.495	NR
	4	0.572	NR		6	0.501	NR
	5	0.455	NR	fGn	3	0.521	NR
	6	0.508	NR		4	0.519	NR
			5		0.498	NR	
MWC	3	0.501	NR	6	0.470	NR	
	4	0.477	NR	$f^{-k}$	3	0.482	NR
	5	0.496	NR		4	0.520	NR
	6	0.496	NR		5	0.513	NR
COM	3	0.123	NR	6	0.508	NR	
	4	0.002	R	LCG	3	0.009	R
	5	$1.11 \times 10^{-16}$	R		4	$1.11 \times 10^{-16}$	R
	6	$1.11 \times 10^{-16}$	R		5	$1.11 \times 10^{-16}$	R
			6		$1.11 \times 10^{-16}$	R	
LEH	3	0.531	NR				
	4	0.515	NR				

power of our test for the alternative hypothesis of correlated  $f^{-k}$  noise and found that it rejects the null hypothesis ( $k = 0$ ) for  $k > 3/10$ .

Although our work focuses on the study of short sequences, we were able to capture the random behavior of well-known pseudorandom number generators already analyzed in the literature. With this, we verified the adequacy of our technique as it is capable of detecting correlation structures.

## 6.5 Reproducibility and Replicability

Following the recommendations provided by Frery et al. (2020), we make the text, source code, and data used in this study available at the *Confidence-Regions* repository <https://github.com/EduardaChagas/ConfidenceRegions>.

# Chapter 7

## Conclusions and Future Work

The main objective of this work was the investigation of problems present in the methodology of symbolization of Bandt-Pompe and its applicability in the characterization of time series and images. Interested in expanding the range of possible applications, we focus on investigating properties of transition graphs and their possible limitations present in the state of the art. Another objective was the study of the joint distribution obtained by the descriptors of the Complexity-Entropy plane, as well as possible linear transformations in this space. Thus, we have advanced in the state of the art by proposing some solutions to deal with scenarios not foreseen in the seminal article by Bandt-Pompe.

The work developed consists of the presentation of two new approaches and their respective applications in data analysis scenarios. The first problem investigated was the characterization and classification of homogeneous patches of SAR image textures. Knowing the limitations that exist in sequences that do not follow a unidimensional structure, the first step was the study of linearization methods, as well as the study of the different properties obtained in each class of data analyzed. Thus, we propose a new approach for classification of remote sensing images, which consists of the following steps: linearization of data using Hilbert curves, generation of the weighted amplitude transition graphs for each sequence, extraction of information theory descriptors, and the classification of these features through the k-nearest neighbors algorithm. As a result, we verified that we were able to obtain the same evaluation metrics of state-of-the-art handcraft algorithms with the use of only two features, which still provides the user with a new way of the general view of the problem, through the Complexity-Entropy plane.

The most important step of the application above was the proposal for a new generation of the ordinal patterns probability distribution, the weighted amplitude



transition graph (WATG). Considering that the magnitude variation of the targets' backscatter signal is an intrinsic characteristic of the analyzed region, we propose a modification in the ordinal patterns transition graph. As can be seen in the literature, one of the limitations of traditional methods of ordinal patterns is the absence of amplitude information during the generation of the distribution. In this way, we propose here the first variation of transition graphs that adds amplitude variation information to its edges. An advantage of using WATG before classical techniques of adding amplitude information to the symbol histogram is in the higher level of information captured about the underlying dynamics of the system, even when using smaller values of patterns dimension. Thus, we can obtain a greater degree of data discrimination more quickly and efficiently.

Another focus of the work was the development of an approach to build confidence regions in the Complexity-Entropy plane. In order to improve the plane characterization capacity, we proposed the HC-PCA, which consists of calculating empirical regions in a new space generated by the linear transformation made by the principal component analysis algorithm. We found that due to the correlation between the descriptors, proposals such as classical bivariate analysis, regressions and generalized linear models cannot describe the system dynamics well. To verify the effectiveness of the proposal, we used these regions to capture the randomness of PRNGs in short sequences and thus characterize generators previously analyzed in the literature. The technique proved to be fast, consistent and robust the addition of correlation structures.

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. Its possible application in the segmentation of classes of regions is a challenging problem.

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. This means that there is a lot of space to conduct further investigations regarding WATG and HC-PCA.

# Bibliography

- Abrams, A., Babson, E., Landau, H., Landau, Z., and Pommersheim, J. (2013). Distributions of order patterns of interval maps. *Combinatorics Probability & Computing*, 22(3):319--341. ISSN 0963-5483.
- Akbarizadeh, G. (2012). A new statistical-based kurtosis wavelet energy feature for texture recognition of SAR images. *IEEE Trans. Geosci. Remote Sens.*, 50(11):4358-4368.
- Altman, N. and Krzywinski, M. (2018). The curse(s) of dimensionality. *Nat. Methods*, 15(6):399--400.
- Anteneodo, C. and Plastino, A. R. (1996). Some features of the López-Ruiz-Mancini-Calbet (LMC) statistical measure of complexity. *Physics Letters A*, 223:348--354.
- Aquino, A. L. L., Cavalcante, T. S. G., Almeida, E. S., Frery, A. C., and Rosso, O. A. (2015). Characterization of vehicle behavior with information theory. *The European Physical Journal B: Condensed Matter and Complex Systems*, 85(10):257--269. ISSN 1434-6036.
- Aquino, A. L. L., Ramos, H. S., Frery, A. C., Viana, L. P., Cavalcante, T. S. G., and Rosso, O. A. (2017). Characterization of electric load with information theory quantifiers. *Physica A*, 465:277--284.
- Azami, H. and Escudero, J. (2016). Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation. *Comput. Methods Programs Biomed.*, 128:40--51.
- Bandt, C. (2019). Small order patterns in big time series: a practical guide. *Entropy*, 21(6):613.
- Bandt, C. and Pompe, B. (2002a). Permutation entropy: A natural complexity measure for time series. *Physical Review Letters*, 88:174102-1--174102-4.

- Bandt, C. and Pompe, B. (2002b). Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.*, 88:174102.
- Bandt, C. and Shiha, F. (2007). Order patterns in time series. *Journal of Time Series Analysis*, 28(5):646--665.
- Baravalle, R., Rosso, O. A., and Montani, F. (2018a). Discriminating imagined and non-imagined tasks in the motor cortex area: Entropy-complexity plane with a wavelet decomposition. *Physica A: Statistical Mechanics and its Applications*, 511:27-39.
- Baravalle, R., Rosso, O. A., and Montani, F. (2018b). Rhythmic activities of the brain: Quantifying the high complexity of beta and gamma oscillations during visuomotor tasks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(7):075513.
- Bardet, J.-M., Lang, G., Oppenheim, G., Philippe, A., and Taqqu, M. S. (2003). Generators of long-range dependent processes: a survey. *Theory and applications of long-range dependence*, pages 579--623.
- Bariviera, A. F., Guercio, M. B., Martinez, L. B., and Rosso, O. A. (2015). The (in)visible hand in the Libor market: an information theory approach. *The European Physical Journal B*, 88(8):208.
- Bariviera, A. F., Zunino, L., Guercio, M. B., Martinez, L. B., and Rosso, O. A. (2013). Efficiency and credit ratings: a permutation-information-theory analysis. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(08):P08007.
- Bariviera, A. F., Zunino, L., and Rosso, O. A. (2018). An analysis of high-frequency cryptocurrencies prices dynamics using permutation-information-theory quantifiers. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(7):075511.
- 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.
- Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In *European Conference on Computer Vision*, pages 404--417. Springer.
- Bi, H., Sun, J., and Xu, Z. (2018). A graph-based semisupervised deep learning model for PolSAR image classification. *IEEE Trans. Geosci. Remote Sens.*, 57(4):2116--2132.

- Bian, C., Qin, C., D Y Ma, Q., and Shen, Q. (2012). Modified permutation-entropy analysis of heartbeat dynamics. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 85:021906.
- Borges, J., Ramos, H., Mini, R., Rosso, O. A., Frery, A. C., and Loureiro, A. A. F. (2019). Learning and distinguishing time series dynamics via ordinal patterns transition graphs. *Appl. Math. Comput.*, 362:UNSP 124554.
- Breit, H., Fritz, T., Balss, U., Lachaise, M., Niedermeier, A., and Vonavka, M. (2009). TerraSAR-X SAR processing and products. *IEEE Trans. Geosci. Remote Sens.*, 48(2):727--740.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer-Verlag, Berlin, 2 edition.
- Carincotte, C., Derrode, S., and Bourennane, S. (2006). Unsupervised change detection on SAR images using fuzzy hidden Markov chains. *IEEE Trans. Geosci. Remote Sens.*, 44(2):432--441. ISSN 1558-0644.
- Chagas, E. T. C., Frery, A. C., Rosso, O. A., and Ramos, H. S. (2020). Characterization of sar images with weighted amplitude transition graphs. In *2020 IEEE Latin American GRSS ISPRS Remote Sensing Conference (LAGIRS)*, pages 264–269.
- Chan, T. M. and Har-Peled, S. (2020). Smallest k-enclosing rectangle revisited. *Discrete & Computational Geometry*.
- Cuesta-Frau, D. (2019). Permutation entropy: Influence of amplitude information on time series classification performance. *Math. Biosci. Eng*, 5:1--16.
- Cuesta-Frau, D., Miró-Martínez, P., Oltra-Crespo, S., Jordán-Núñez, J., Vargas, B., and Vigil, L. (2018). Classification of glucose records from patients at diabetes risk using a combined permutation entropy algorithm. *Computer methods and programs in biomedicine*, 165:197--204.
- Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886--893. IEEE.
- de Araujo, F. H. A., Bejan, L., Rosso, O. A., and Stosic, T. (2019). Permutation entropy and statistical complexity analysis of brazilian agricultural commodities. *Entropy*, 21(12):1220.

- 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. ISSN 0378-4371.
- De Micco, L., Larrondo, H. A., Plastino, A., and Rosso, O. A. (2009). Quantifiers for randomness of chaotic pseudo-random number generators. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1901):3281--3296. ISSN 1471-2962.
- Dumitru, C. O., Cui, S., Schwarz, G., and Datcu, M. (2014). Information content of very-high-resolution SAR images: Semantics, geospatial context, and ontologies. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 8(4):1635--1650.
- Echegoyen, I., López-Sanz, D., Martínez, J. H., Maestú, F., and Buldú, J. M. (2020). Permutation entropy and statistical complexity in mild cognitive impairment and alzheimer's disease: An analysis based on frequency bands. *Entropy*, 22(1):116.
- Fadlallah, B. H., Chen, B., Keil, A., and Príncipe, J. C. (2013). Weighted-permutation entropy: a complexity measure for time series incorporating amplitude information. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 87 2:022911.
- Florindo, J. B. and Bruno, O. M. (2012). Fractal descriptors based on Fourier spectrum applied to texture analysis. *Physica A*, 391(20):4909--4922.
- Frery, A. C., Gomez, L., and Medeiros, A. C. (2020). A badging system for reproducibility and replicability in remote sensing research. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*
- 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. ISSN 1749-4893.
- González, C., Larrondo, H., and Rosso, O. (2005). Statistical complexity measure of pseudorandom bit generators. *Physica A: Statistical Mechanics and its Applications*, 354:281--300.
- Guan, D., Xiang, D., Tang, X., Wang, L., and Kuang, G. (2019). Covariance of textural features: A new feature descriptor for SAR image classification. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 12(10):3932--3942.
- Haahr, M. (1998–2018). RANDOM.ORG: true random number service. <https://www.random.org>. Accessed: 2018-06-01.

- Hagensieker, R. and Waske, B. (2018). Evaluation of multi-frequency SAR images for tropical land cover mapping. *Remote Sensing*, 10(2):257.
- Han, P., Han, B., Lu, X., Cong, R., and Sun, D. (2020). Unsupervised classification for PolSAR images based on multi-level feature extraction. *Int. J. Remote Sens.*, 41(2):534--548.
- Huang, Z., Dumitru, C. O., Pan, Z., Lei, B., and Datcu, M. (2020). Classification of large-scale high-resolution SAR images with deep transfer learning. *arXiv preprint arXiv:2001.01425*.
- Kankaku, Y., Suzuki, S., and Osawa, Y. (2013). Alos-2 mission and development status. In *IGARSS 2013 – IEEE International Geoscience and Remote Sensing Symposium*, pages 2396--2399. IEEE.
- Knuth, D. (1997). Sorting and searching, vol. 3 of the art of computer programming, section 6.2. 2.
- Kourgli, A., Ouarzeddine, M., Oukil, Y., and Belhadj-Aissa, A. (2012). Texture modelling for land cover classification of fully polarimetric SAR images. *International Journal of Image and Data Fusion*, 3(2):129--148.
- Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Stat. Softw.*, 28(5):1--26.
- Kulp, C. W., Chobot, J. M., Freitas, H. R., and Sprechini, G. D. (2016). Using ordinal partition transition networks to analyze ECG data. *Chaos*, 26(7):073114.
- Lamberti, P. W., Martín, M. T., Plastino, A., and Rosso, O. A. (2004). Intensive entropic non-triviality measure. *Physica A*, 334(1-2):119--131. ISSN 0378-4371.
- Larrondo, H. (2012). Matlab program: noisefk. m. <http://www.mathworks.com/matlabcentral/fileexchange/35381>.
- Larrondo, H., Micco, L., González, C., Plastino, A., and Rosso, O. (2013). *Statistical Complexity of Chaotic Pseudorandom Number Generators*, pages 283 -- 308. Bentham e-Books, Bentham Publishers.
- Larrondo, H. A., De Micco, L., and González, C. M. (2002). Statistical complexity of chaotic pseudorandom number generators. *Phys. Rev. Lett*, 88:174102.

- 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. ISSN 0375-9601.
- Lee, J.-H. and Hsueh, Y.-C. (1994). Texture classification method using multiple space filling curves. *Pattern Recognit. Lett.*, 15(12):1241–1244.
- Lee, J.-S., Grunes, M. R., Pottier, E., and Ferro-Famil, L. (2004). Unsupervised terrain classification preserving polarimetric scattering characteristics. *IEEE Trans. Geosci. Remote Sens.*, 42(4):722–731.
- Lin, Z., Ji, K., Kang, M., Leng, X., and Zou, H. (2017). Deep convolutional highway unit network for sar target classification with limited labeled training data. *IEEE Geosci. Remote Sens. Lett.*, 14(7):1091–1095.
- Liu, F., Jiao, L., and Tang, X. (2019). Task-oriented GAN for PolSAR image classification and clustering. *IEEE Trans. Neural Netw. Learn. Syst.*, 30(9):2707–2719.
- Marsaglia, G. (1994). Yet another rng. *Posted to the electronic billboard sci. stat. math, August*, 1.
- Martín, M. T., Plastino, A., and Rosso, O. A. (2006). Generalized statistical complexity measures: Geometrical and analytical properties. *Physica A*, 369(2):439–462.
- Martin, M. T., Plastino, A., and Rosso, O. A. (2006a). Generalized statistical complexity measures: Geometrical and analytical properties. *Physica A: Statistical Mechanics and its Applications*, 369(2):439–462. ISSN 0378-4371.
- Martin, M. T., Plastino, A., and Rosso, O. A. (2006b). Generalized statistical complexity measures: Geometrical and analytical properties. *Physica A*, 369:439–462.
- McCullough, M., Small, M., Stemler, T., and Iu, H. H.-C. (2015). Time lagged ordinal partition networks for capturing dynamics of continuous dynamical systems. *Chaos*, 25(5):053101.
- McNairn, H., Kross, A., Lapen, D., Caves, R., and Shang, J. (2014). Early season monitoring of corn and soybeans with TerraSAR-X and RADARSAT-2. *Int. J. Appl. Earth Obs. Geoinf.*, 28:252–259.
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-hill New York.
- Morena, L., James, K., and Beck, J. (2004). An introduction to the RADARSAT-2 mission. *Canadian Journal of Remote Sensing*, 30(3):221–234.

- Nguyen, P. and Quinqueton, J. (1982). Space filling curves and texture analysis. In *IEEE Intl. Conf. Pattern Recognition*, pages 282--285.
- Numbisi, F. N., Van Coillie, F., and De Wulf, R. (2018). Multi-date Sentinel SAR image textures discriminate perennial agroforests in a tropical forest-savannah transition landscape. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 42(1).
- Parlitz, U., Berg, S., Luther, S., Schirdewan, A., Kurths, J., and Wessel, N. (2012). Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics. *Computers in biology and medicine*, 42(3):319--327.
- Payne, W., Rabung, J. R., and Bogyo, T. (1969). Coding the lehmer pseudo-random number generator. *Communications of the ACM*, 12(2):85--86.
- Peitgen, H.-O., Jürgens, H., and Saupe, D. (2006). *Chaos and fractals: new frontiers of science*. Springer Science & Business Media.
- Portnoff, M. (1980). Time-frequency representation of digital signals and systems based on short-time fourier analysis. *IEEE Trans. Acoust., Speech, Signal Process.*, 28(1):55--69.
- Qin, F., Guo, J., and Sun, W. (2017). Object-oriented ensemble classification for polarimetric sar imagery using restricted boltzmann machines. *Remote Sensing Letters*, 8(3):204--213.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Radford, D. D., Cracknell, M. J., Roach, M. J., and Cumming, G. V. (2018). Geological mapping in Western Tasmania using radar and random forests. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 11(9):3075--3087.
- Ravetti, M. G., Carpi, L. C., Gonçalves, B. A., Frery, A. C., and Rosso, O. A. (2014). Distinguishing noise from chaos: objective versus subjective criteria using Horizontal Visibility Graph. *PLOS One*, 9(9):1--15.
- Ressel, R., Frost, A., and Lehner, S. (2015). A neural network-based classification for sea ice types on X-band SAR images. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 8(7):3672--3680.



- Riedl, M., Müller, A., and Wessel, N. (2013). Practical considerations of permutation entropy. *The European Physical Journal Special Topics*, 222(2):249--262.
- Rosso, O., Larrondo, H., Martin, M., Plastino, A., and Fuentes, M. (2007a). Distinguishing noise from chaos. *Physical Review Letters*, 99:154102.
- Rosso, O. and Masoller, C. (2009a). Detecting and quantifying temporal correlations in stochastic resonance via information theory measures. *The European Physical Journal B*, 69(1):37--43.
- Rosso, O., Olivares, F., Zunino, L., Micco, L., Aquino, A., Plastino, A., and Larrondo, H. (2013). Characterization of chaotic maps using the permutation bandt-pompe probability distribution. *The European Physical Journal E*, 86:116.
- Rosso, O., Zunino, L., Pérez, D., Figliola, A., Larrondo, H., Garavaglia, M., Martín, M., and Plastino, A. (2007b). Extracting features of gaussian self-similar stochastic processes via the bandt-pompe approach. *Physical Review E*, 76(6):061114.
- Rosso, O. A. and Masoller, C. (2009b). Detecting and quantifying stochastic and coherence resonances via information-theory complexity measurements. *Physical Review E*, 79(4):040106.
- Rosso, O. A., Ospina, R., and Frery, A. C. (2016). Classification and verification of handwritten signatures with time causal information theory quantifiers. *PLOS ONE*, 11(12):e0166868.
- Rousseeuw, P. J., Ruts, I., and Tukey, J. W. (1999). The bagplot: A bivariate boxplot. *The American Statistician*, 53(4):382.
- Saco, P. M., Carpi, L. C., Figliola, A., Serrano, E., and Rosso, O. A. (2010). Entropy analysis of the dynamics of el niño/southern oscillation during the holocene. *Physica A: Statistical Mechanics and its Applications*, 389(21):5022--5027.
- Schieber, T. A., Carpi, L., Frery, A. C., Rosso, O. A., Pardalos, P. M., and Ravetti, M. G. (2016). Information theory perspective on network robustness. *Physics Letters A*, 380:359--364.
- Schmidt, R., Hrycej, T., and Stützle, E. (2006). Multivariate distribution models with generalized hyperbolic margins. *Computational Statistics & Data Analysis*, 50(8):2065--2096.

- Sinn, M. and Keller, K. (2011). Estimation of ordinal pattern probabilities in Gaussian processes with stationary increments. *Computational Statistics & Data Analysis*, 55(4):1781--1790.
- Sorrentino, T., Quintero-Quiroz, C., Aragonese, A., Torrent, M., and Masoller, C. (2015). Effects of periodic forcing on the temporally correlated spikes of a semiconductor laser with feedback. *Opt. Express*, 23.
- Storie, C. D. (2018). Urban boundary mapping using Sentinel-1A SAR data. In *IGARSS 2018 - IEEE International Geoscience and Remote Sensing Symposium*, pages 2960--2963.
- Sukawattanavijit, C., Chen, J., and Zhang, H. (2017). GA-SVM algorithm for improving land-cover classification using SAR and optical remote sensing data. *IEEE Geosci. Remote Sens. Lett.*, 14(3):284--288.
- Traversaro, F., Redelico, F., Risk, M., Frery, A. C., and Rosso, O. (2018). Bandt-pompe symbolization dynamics for time series with tied values: a data-driven approach. *Chaos: an Interdisciplinary Journal of Nonlinear Science*. ISSN 1054-1500.
- Traversaro, F., Risk, M., Rosso, O., and Redelico, F. (2017). An empirical evaluation of alternative methods of estimation for permutation entropy in time series with tied values. *arXiv preprint arXiv:1707.01517*.
- Wang, R. and Wang, Y. (2019). Classification of PolSAR image using neural nonlocal stacked sparse autoencoders with virtual adversarial regularization. *Remote Sensing*, 11(9):1038.
- Weldon, T. P., Higgins, W. E., and Dunn, D. F. (1996). Efficient Gabor filter design for texture segmentation. *Pattern Recognit.*, 29(12):2005--2015.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer. ISBN 978-0-387-98140-6.
- Xiao-Feng, L. and Yue, W. (2009). Fine-grained permutation entropy as a measure of natural complexity for time series. *Chin. Phys. B*, 18(7):2690.
- Xie, W., Ma, G., Zhao, F., Liu, H., and Zhang, L. (2020). PolSAR image classification via a novel semi-supervised recurrent complex-valued convolution neural network. *Neurocomputing*.

- Xiong, H., Shang, P., He, J., and Zhang, Y. (2020). Complexity and information measures in planar characterization of chaos and noise. *Nonlinear Dynamics*, 100:1673-1687.
- Yu, Q., Xing, M., Liu, X., Wang, L., Luo, K., and Quan, X. (2019). Detection of land use type using multitemporal SAR images. In *IGARSS 2019 – IEEE International Geoscience and Remote Sensing Symposium*, pages 1534--1537. IEEE.
- Yue, D.-X., Xu, F., Frery, A. C., and Jin, Y.-Q. (2020). A generalized Gaussian coherent scatterer model for correlated SAR texture. *IEEE Trans. Geosci. Remote Sens.*, 58(4):2947--2964.
- Zhang, J., Zhou, J., Tang, M., Guo, H., Small, M., and Zou, Y. (2017). Constructing ordinal partition transition networks from multivariate time series. *Sci. Rep.*, 7:7795.
- Zhu, T., Li, F., Heygster, G., and Zhang, S. (2016). Antarctic sea-ice classification based on conditional random fields from RADARSAT-2 dual-polarization satellite images. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 9(6):2451--2467.
- Zunino, L., Bariviera, A. F., Guercio, M. B., Martinez, L. B., and Rosso, O. A. (2012a). On the efficiency of sovereign bond markets. *Physica A: Statistical Mechanics and its Applications*, 391(18):4342--4349.
- Zunino, L., Soriano, M. C., and Rosso, O. A. (2012b). Distinguishing chaotic and stochastic dynamics from time series by using a multiscale symbolic approach. *Phys. Rev. E*, 86:046210.
- Zunino, L., Zanin, M., Tabak, B. M., Pérez, D. G., and Rosso, O. A. (2010). Complexity-entropy causality plane: A useful approach to quantify the stock market inefficiency. *Physica A: Statistical Mechanics and its Applications*, 389(9):1891--1901.