



Trabalho de Conclusão de Curso

Entropia de Permutação Espacial Multivariada e sua aplicação para Análise da Complexidade de Sistemas Caóticos

Givanildo Lima do Nascimento Junior
givanildo.lima@laccan.ufal.br

Orientador:
Prof. Dr. André Luiz Lins de Aquino

Maceió, Julho de 2023

Givanildo Lima do Nascimento Junior

Entropia de Permutação Espacial Multivariada e sua aplicação para Análise da Complexidade de Sistemas Caóticos

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Engenharia de Computação do Instituto de Computação da Universidade Federal de Alagoas.

Orientador:

Prof. Dr. André Luiz Lins de Aquino

Maceió, Julho de 2023

Catálogo na Fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 - 1767

M414f Nascimento Junior, Givanildo Lima do.

Entropia de permutação espacial multivariada e sua aplicação para análise da complexidade de sistemas caóticos / Givanildo Lima do Nascimento Junior. – 2023.

40 f. : il.

Orientador: André Luiz Lins de Aquino.

Monografia (Trabalho de conclusão de curso em Engenharia de Computação) - Universidade Federal de Alagoas, Instituto de Computação. Maceió, 2023.

Bibliografia: f. 35-40.

1. Sistemas dinâmicos diferenciais. 2. Caos (Sistemas dinâmicos). 3. Entropia (Teoria da informação). 4. Complexidade estatística. 5. Análise multivariada. I. Título.

CDU: 519.722

Agradecimentos

Foi uma jornada longa, mais do que deveria, mas extremamente engrandecedora. Sou muito grato a Deus por tudo que vivi desde o ingresso na UFAL e pela pessoa que me tornei ao término desse ciclo.

Agradeço imensamente a minha vó e a minha mãe, em especial, por compartilharem cada fase desse caminho comigo, essa conquista é muito mais delas do que minha.

Aos meus tios e tia, Helena, Manoel, Geazir e Márcio, que se fizeram tão presentes e ocuparam cada lacuna deixada pela vida.

A minha namorada Cacau, por ter sido o melhor colo ao longo dos últimos anos e me apoiado em cada situação.

Aos amigos da IFAL/aSeita, Felipe, Léo, Maumau, Kevyn, Laurinha, Alan e Pedro. Tenho muito orgulho por ter criado laços tão próximos com pessoas que tanto amo e admiro.

Aos meus amigos de infância, Frank e João, pois nada é mais valioso do que aquilo que resiste ao passar do tempo.

Aos amigos do curso, André e Valdir, sem vocês não teria conseguido. Nossa relação e conversas tornaram as idas a UFAL muito mais leves e fico feliz por poder levar nossa amizade para além da UFAL.

Aos professores, Alla e Fabi, que acreditaram em mim em momentos em que eu mesmo me questionava. Descobri o que mais gostava de fazer graças a Ny e agora sinto que estou na direção correta.

Um abraço e um cheiro calorosos em cada um!

“Eu gosto de onde eu vou e de onde eu vim.”

– Racionais MC’s

Resumo

A análise de séries temporais que modelam sistemas dinâmicos não lineares é um tema bastante explorado no meio científico, em função da ampla gama de aplicações das medidas de complexidade que são utilizadas para caracterização e estudo desses sistemas.

Em particular, o caos é um dos comportamentos que podem ser observados em sistemas não-lineares a depender dos valores que determinados para seus parâmetros e as condições iniciais estabelecidas. Por apresentar grande sensibilidade a essas condições, sua trajetória pode divergir exponencialmente com pequenas mudanças de parâmetros.

Dessa forma, a identificação de regimes caóticos e a quantificação do grau de caoticidade são essenciais para determinar os limites dentro dos quais é possível realizar previsões. Para efetuar essa caracterização, foi introduzido o conceito do Plano Causal de Complexidade-Entropia (CECP), capaz de detectar a complexidade subjacente de sistemas dinâmicos e extrair a estrutura da série temporal de acordo com o função de distribuição de probabilidade construída a partir de seus padrões ordinais. O algoritmo desta técnica é simples e tem a capacidade de distinguir entre sinais periódicos, caóticos e estocásticos.

Entretanto, nessa abordagem, a avaliação é feita sobre séries temporais univariadas, não garantindo uma caracterização adequada para sistemas multivariados, visto que a técnica adotada desconsidera a correlação existente entre as séries componentes do sistema. Além disso, por só considerar os padrões ordinais intrínsecos as séries temporais, essa abordagem implica na perda de alguns detalhes de informações de amplitude da série original.

Neste trabalho, é proposta uma abordagem multivariada do CECP para caracterização de sistemas dinâmicos multidimensionais, por meio do Plano de Causal Multivariado de Complexidade-Entropia (MvCECP). Além disso, é adicionada a função de distribuição de probabilidade a identificação de subpadrões espaciais, tornando a abordagem sensível a mudanças de escala nos conjuntos de dados avaliados. Para assegurar a eficácia do método, é demonstrada sua capacidade de distinguir os comportamentos caótico, estocástico e periódico de sistemas multivariados, sendo ainda mais precisa que a abordagem univariada na caracterização da dinâmica caótica.

Palavras-chave: Sistemas Dinâmicos, Caos, Teoria da Informação, Entropia de Shannon, Complexidade Estatística, Análise Multivariada

Abstract

The analysis of time series that model nonlinear dynamic systems is a topic that has been widely explored in the scientific world, due to the wide range of applications of complexity measures that are used to characterize and study these systems.

In particular, chaos is one of the behaviors that can be observed in nonlinear systems depending on the values that are determined for its parameters and the initial conditions established. Due to its great sensitivity to these conditions, its trajectory can diverge exponentially with small parameter changes.

Thus, the identification of chaotic regimes and the quantification of the degree of chaoticity are essential to determine the limits within which it is possible to make predictions. To carry out this characterization, the concept of the Causal Complexity-Entropy Plan (CECP) was introduced, capable of detecting the underlying complexity of dynamic systems and extracting the structure of the time series according to the probability distribution function constructed from its ordinal patterns. The algorithm of this technique is simple and has the ability to distinguish between periodic, chaotic and stochastic signals.

However, in this approach, the evaluation is performed on univariate time series, not guaranteeing an adequate characterization for multivariate systems, since the technique adopted disregards the existing correlation between the component series of the system. Furthermore, as it only considers the ordinal patterns intrinsic to the time series, this approach implies the loss of some details of amplitude information from the original series.

In this work, a multivariate CECP approach is proposed for the characterization of dynamic multidimensional systems, through the Complexity-Entropy Multivariate Causal Plan (MvCECP). In addition, the probability distribution function is added to the identification of spatial subpatterns, making the approach sensitive to scale changes in the evaluated datasets. To ensure the effectiveness of the method, its ability to distinguish the chaotic, stochastic and periodic behavior of multivariate systems is demonstrated, being even more accurate than the univariate approach in the characterization of chaotic dynamics.

Keywords: Dynamical Systems, Chaos, Information Theory, Shannon Entropy, Statistical Complexity, Multivariate Analysis

Lista de Figuras

3.1	Ilustração gráfica do funcionamento das janelas de dados deslizantes utilizadas nas abordagens univariada (a) e multivariada (b), com $D=\sigma=3$	16
3.2	As 6 configurações de padrões ordinais para o caso em que $\sigma = 3$	17
3.3	Padrões de ordem para dimensão de incorporação $\sigma = 3$ com seus subpadrões π_{18} em um sistema univariado	18
3.4	As 10 possibilidades de padrões de distância existentes para uma dada ordenação dos elementos no caso em que $\sigma = 4$ em um sistema univariado	19
3.5	Plano dividido em $\sigma = 3$ regiões para avaliação de padrões de distância de sistemas bidimensionais	21
3.6	Espaço tridimensional dividido em $\sigma = 3$ regiões para avaliação de padrões de distância	22
4.1	Componentes do Sistema Caótico B_7	26
4.2	Avaliação das séries temporais componentes dos sistemas caóticos, processos estocásticos e dados periódicos no CECP para $D = 6$	27
4.3	Representação Gráfica do Sistema Multivariado B_7 no espaço tridimensional	29
4.4	Avaliação no dos sistemas caóticos, processos estocásticos e dados periódicos no MvCECP para $\sigma = 5$	30

Lista de Tabelas

4.1	Entropia e Complexidade de cada uma das componentes dos Sistemas Caóticos avaliados no CECF	28
4.2	Entropia e Complexidade dos Sistemas Caóticos avaliados no MvCECF	29

Lista de Abreviaturas e Siglas

CS	Sistemas Caóticos
SP	Processos Estocásticos
TS	Séries Temporais
PDF	Função de Densidade de Probabilidade
BP	Bandt-Pompe
PE	Entropia de Permutação
SC	Complexidade Estatística
CECP	Plano Causal de Complexidade-Entropia
MvCECP	Plano Causal Multivariado de Complexidade-Entropia
LFSR	Linear Feedback Shift Register
RSSF	Redes de Sensores Sem Fio
PS	Espectro de Potência

Lista de Símbolos

D	Dimensão de Incorporação de BP
τ	Atraso de Imersão
P	Distribuição de Probabilidade de BP
P_e	Distribuição Uniforme
S	Entropia de Shannon
\mathcal{H}	Entropia de Shannon Normalizada
C	Complexidade Estatística
Q_I	Desequilíbrio entre os Sistemas
J	Divergente de <i>Jensen-Shannon</i>
$\mathcal{H} \times C$	Plano Complexidade-Entropia
P_{Mv}	Distribuição de Probabilidade Multivariada
S_{Mv}	Entropia de Shannon Multivariada
\mathcal{H}_{Mv}	Entropia de Shannon Multivariada Normalizada
C_{Mv}	Complexidade Estatística Multivariada
$\mathcal{H}_{Mv} \times C_{Mv}$	Plano Complexidade-Entropia Multivariado

Conteúdo

Lista de Figuras	iii
Lista de Tabelas	v
Lista de Abreviaturas e Siglas	vi
Lista de Símbolos	vii
1 Introdução	1
1.1 Motivação	1
1.2 Revisão da Literatura	2
1.3 Objetivo	3
1.4 Solução Proposta	4
1.5 Contribuições	4
1.6 Estrutura do texto	6
2 Sistemas Caóticos, Processos Estocásticos e Dados Periódicos	7
2.1 Introdução aos Sistemas Caóticos	7
2.2 Modelagem Matemática	9
3 Teoria da Informação	12
3.1 PDF de Bandt-Pompe	12
3.2 Entropia de Shannon	13
3.3 Complexidade Estatística	14
3.4 Plano de Causalidade Complexidade-Entropia (CECP)	15
3.5 PDF Multivariada	16
3.5.1 Padrões Ordinais	16
3.5.2 Padrões de Distância	17
3.6 Quantificadores de Teoria da Informação Multivariados	24
4 Resultados e Discussões	25
4.1 Construção da Base de Dados	25
4.2 Análise Univariada com o CECP	25
4.3 Análise Multivariada com o MvCECP	27
5 Considerações Finais	32
5.1 Conclusões	32
5.2 Trabalhos Futuros	33

Referências bibliográficas

1

Introdução

1.1 Motivação

A análise da complexidade de séries temporais geradas por sistemas dinâmicos não lineares é um tema de enorme interesse em pesquisas científicas, devido as diversas aplicações práticas das medidas de complexidade que atuam na caracterização dessas séries temporais, sejam elas regulares, caóticas ou baseadas na realidade, tais quais as que são coletadas por redes de sensores. Por exemplo, (Kaffashi et al., 2013) analisaram a complexidade dos sinais de EEG durante o sono, através do contato pele a pele; (Silva et al., 2015) e (Mukherjee et al., 2015) mediram a complexidade da variabilidade da frequência cardíaca e da insuficiência cardíaca, respectivamente; mais recentemente, (Ma et al., 2022) exploraram a estrutura temporal sutil de dados ambientais e sua previsibilidade ao longo das estações do ano.

A medida de complexidade de sistemas caóticos é outro foco de interesse, pois pode descrever elegantemente a dinâmica e a complexidade desses sistemas. (Bandt and Pompe, 2002) foram os primeiros a observar que a maioria dos métodos de cálculo de complexidade dependia de algoritmos específicos e em certos casos, também poderiam existir parâmetros de ajuste. Como efeito direto, os resultados podem ser duvidosos, caso não se conheçam os detalhes dos métodos ou não sejam selecionados os parâmetros adequados.

A fim de solucionar isto, Bandt-Pompe (BP) propuseram a entropia de permutação (PE) por meio da comparação de valores vizinhos que poderiam ser adequados para qualquer série temporal (Bandt and Pompe, 2002). O padrão ordinal é determinado através da reconstrução do espaço de fase subjacente, seguido da comparação de valores consecutivos ou não consecutivos. Além disso, suas Funções de Densidade de Probabilidade (PDF) são invariáveis em termos de transformações monótonas não lineares. O algoritmo desta técnica é simples e tem a capacidade de distinguir entre sinais periódicos, caóticos e estocásticos, mesmo com a presença de ruídos em conjunto com os dados analisados, sendo amplamente utilizado na análise de sistemas dinâmicos não lineares (Yin and Shang, 2014; Zhang and Shang, 2017). Esse recurso

nos permite descobrir detalhes importantes sobre a estrutura ordinal da série temporal e também pode fornecer informações sobre a correlação temporal. Embora este tipo de análise de séries temporais acarrete na perda de alguns detalhes das informações de amplitude da série original, ao referir-se apenas à estrutura intrínseca da série, essa técnica consegue uma significativa redução da dificuldade no que diz respeito à descrição de sistemas complexos.

Posteriormente, (Rosso et al., 2007b) introduziram pela primeira vez o Plano Causal de Complexidade-Entropia (CECP) a partir da combinação da entropia de permutação com a medida de Complexidade Estatística (SC) proposta por (Lopez-Ruiz et al., 1995), a fim de detectar ainda mais precisamente a complexidade subjacente de sistemas dinâmicos fornecendo uma distinção mais robusta entre as dinâmicas determinística e estocástica (Rosso et al., 2007a, 2013). Essa abordagem integra conceitos de entropia e dinâmica não linear e pode extrair a estrutura temporal da série temporal de acordo com a PDF construída a partir dos padrões ordinais. Ela tem sido amplamente utilizada em diversas áreas científicas para distinguir sistemas caóticos e processos estocásticos (Chagas et al., 2022; Borges et al., 2019; Ribeiro et al., 2012; Zunino et al., 2011), além de estudos relativos a amostragem de dados (Nascimento et al., 2021; De Micco et al., 2012).

Entretanto, a avaliação com o CECP é aplicável somente a conjuntos de dados univariados, ainda que a grande maioria dos sistemas existentes seja multivariada. Dessa forma, a caracterização é feita a partir de cada série temporal componente do sistema, desconsiderando a correlação existente entre elas.

Neste trabalho, visando solucionar o problema mencionado, iremos estender o CECP para a análise de sistemas multivariados, adicionando ao processo a análise de subpadrões de distância, além dos padrões ordinais, que permitirão uma caracterização mais detalhada da série original, identificando as diferenças de escala e proximidade entre os dados do sistema. Assim, a PDF multivariada é obtida por meio da reconstrução do espaço de fase subjacente, seguido da comparação de valores consecutivos ou não consecutivos e por fim, a comparação das amplitudes e distâncias euclidianas entre esses elementos. Por fim, para melhor descrever as propriedades intrínsecas e as estruturas físicas dos sistemas caóticos multivariados estudados, introduzimos o Plano Causal Multivariado de Complexidade-Entropia (MvCECP).

1.2 Revisão da Literatura

Na literatura, além da abordagem adotada por Rosso et al. (2007b), existem uma série de outras técnicas para análise univariada do comportamento de sistemas dinâmicos e séries temporais (Rosenstein et al., 1993; Pincus, 1991).

Entretanto, como foi mencionado anteriormente, essas técnicas não conseguem oferecer uma avaliação adequada para sistemas multivariados, pois somente permitem efetuar os estudos componente por componente, desconsiderando a correlação existente entre elas. Então, para

analisar a complexidade de sistemas dinâmicos multivariados, foram propostas técnicas como a entropia de amostra multivariada (MvSampEn) (Ahmed and Mandic, 2011) e a entropia fuzzy multivariada (MvFuzzyEn) (Li et al., 2013). Entretanto, o principal problema compartilhado por essas soluções é sua velocidade de cálculo, pois são bastante lentas, inviabilizando o uso em aplicações que atuem em tempo real. Para esses casos, soluções que utilizam a Entropia de Permutação como referência são mais recomendadas, visto que PE tem velocidade de cálculo mais rápida e é mais precisa para estimativa de complexidade (Li et al., 2014; Unakafova et al., 2013).

Enquanto isso, Ribeiro et al. (2012) aprimoraram o algoritmo para padrões dimensionais superiores e seus resultados provaram que seu método é promissor para reconhecimento entre padrões bidimensionais, mas apresenta algumas deficiências; por exemplo, embora funcione bem para sistemas bidimensionais, não é apropriado para ser aplicado a sistemas tridimensionais ou de dimensões superiores, pois quanto maior a dimensão do sistema em questão, mais complexo é o procedimento e maior o número de variáveis necessárias.

No trabalho de He et al. (2016) é possível encontrar estudos sobre a construção de subpadrões para a análise multivariada, entretanto, seu estudo é aplicável apenas a sistemas tridimensionais. Além disso, suas avaliações consideram somente um instante de tempo a cada rodada, comparando os valores de cada componente naquele instante entre si. Dessa forma, a evolução temporal dos dados em cada componente não é investigada, sendo levada em consideração apenas a correlação entre os dados de componentes distintas.

1.3 Objetivo

Este trabalho tem como objetivo propor uma abordagem multivariada do CECP, intitulada Mv-CECP, por meio da construção de uma função de distribuição de probabilidade capaz de avaliar sistemas por inteiro, considerando todas as suas séries temporais componentes e a correlação existente entre elas.

Além disso, a PDF desenvolvida também foi adaptada para ser capaz de identificar subpadrões de distâncias, avaliando as diferenças de amplitude entre dados consecutivos analisados ao longo das séries temporais. Para atestar a eficácia na caracterização de sistemas caóticos e em distinguir os comportamentos caótico, estocástico e periódico, foram selecionados 8 sistemas caóticos, 2 processos estocásticos e 1 sistema periódico a fim de comparar os estudos univariado e multivariado do comportamento de cada variável ou sistema ao longo do tempo e, por fim, confrontar as dinâmicas mapeadas em cada abordagem.

1.4 Solução Proposta

Inicialmente, foi construída uma base composta por dados de natureza caótica, estocástica e periódica. Cada sistema utilizado foi construído a partir das equações diferenciais que o modelam, sendo indicados os valores utilizados para os coeficientes. As condições iniciais foram baseadas no trabalho de (Sprott and Sprott, 2003).

Para a avaliação desse conjunto de dados, foi desenvolvida uma PDF adaptada a partir da que foi proposta por BP. Em comparação a original, suas diferenças são: (i) uma abordagem adaptada que permite uma avaliação multivariada dos sistemas, considerando assim a correlação existente entre as séries temporais componentes de um sistema multidimensional e (ii) a avaliação de subpadrões de distância, além dos padrões ordinais, a fim de extrair informações sobre padrões de escala entre os elementos consecutivos das séries temporais, assim como na extração de padrões ordinais.

Em seguida, é possível adaptar os cálculos de Entropia de Permutação e Complexidade Estatística para que utilizem a PDF desenvolvida e só então, é construído o Plano Causal Multivariado de Complexidade-Entropia, contando com PE no eixo das abscissas e a SC no eixo das ordenadas.

Então, aplicamos os sistemas a caracterização nas abordagens univariada e multivariada, por meio dos planos CECP e MvCECP, respectivamente, a fim de demonstrar a capacidade da abordagem multivariada em proporcionar uma distinção adequada entre os comportamentos caótico, estocástico e periódico, assim, como na abordagem univariada, com a vantagem de prover uma caracterização mais precisa para diferenciar entre os sistemas caóticos multivariados.

1.5 Contribuições

As contribuições deste trabalho são:

- Construção de PDF para caracterização de sistemas multivariados por meio de seus padrões ordinais;
- PDF desenvolvida também é capaz de detectar padrões de distância e escala entre os elementos avaliados;
- Proposta do Plano Causal Multivariado de Complexidade-Entropia (MvCECP);
- Caracterização de sistemas caóticos em uma abordagem multivariada e comparação com a caracterização univariada.
- Demonstração de que o MvCECP é capaz de fazer distinção entre sistemas caóticos, processos estocásticos e dados periódicos, assim como o CECP;

- MvCECP demonstra ser mais efetivo que o CECP na caracterização de sistemas caóticos multivariados.

Observamos que as contribuições apresentadas são muito relevantes tanto para o domínio geral de análises multivariadas, quanto para os estudos de sistemas caóticos em particular. A utilização do MvCECP para uma caracterização simples, precisa e intuitiva desses sistemas apresenta grandes possibilidades, e sua aplicação na avaliação de dados multivariados pode trazer uma série de benefícios a fim de entender as dinâmicas de diversas variáveis ao longo do tempo sob essa nova perspectiva.

1.6 Estrutura do texto

Os demais capítulos ao longo deste trabalho estão organizados da seguinte forma:

- O capítulo 2 apresenta a modelagem matemática dos três tipos de sistemas multidimensionais utilizados para aplicação e avaliação da técnica multivariada desenvolvida: sistemas caóticos, processos estocásticos e dados periódicos;
- No capítulo 3, apresentamos os quantificadores de teoria da informação de forma detalhada, além da metodologia para construção das medidas de complexidade estatística e entropia de permutação a partir da nova PDF multivariada proposta. Em seguida, integramos essas medidas e apresentamos o MvCECP;
- No capítulo 4 apresentamos a aplicação empírica das análises univariada e multivariada nos sistemas multidimensionais definidos no 2.2 e discutimos os resultados obtidos nas análises efetuadas por meio do CECP e do MvCECP, comparando os resultados;
- O Capítulo 5 encerra as discussões, ressaltando as principais contribuições do trabalho de acordo com os resultados apresentados e delineando as possíveis direções futuras do trabalho;
- Por fim, apresentamos as referências mencionadas ao longo do trabalho.

2

Sistemas Caóticos, Processos Estocásticos e Dados Periódicos

2.1 Introdução aos Sistemas Caóticos

Uma grande variedade de sistemas apresenta comportamentos irregulares e imprevisíveis, em maior ou menor grau, a exemplo do clima, bolsa de valores, o tráfego de carros e até mesmo o aparentemente regular sistema solar ([Laskar, 1989](#)).

Dessa forma, não-linearidades, comuns a uma enorme variedade de sistemas, podem tornar sua evolução não trivial. Em particular, o caos é um dos comportamentos que podem ser observados em sistemas não-lineares a depender dos valores adotados para os parâmetros iniciais. Um sistema é dito caótico quando apresenta grande sensibilidade às condições iniciais, dessa forma, sua trajetória pode divergir exponencialmente apenas com algumas pequenas mudanças de parâmetros.

Assim, mesmo no caso de sistemas que possam ser modelados satisfatoriamente por equações deterministas, ou seja, sem ruídos ou incertezas explícitas, sua evolução pode acontecer de forma aparentemente aleatória e irregular, a depender das condições iniciais. Na presença de caos, a própria incerteza com relação a essas condições, associada à precisão finita de qualquer medição impede uma previsão da evolução futura desses sistemas. Portanto, a identificação de regimes caóticos e a quantificação do grau de caoticidade são essenciais para estabelecer os limites dentro dos quais é possível efetuar previsões ([Hilborn et al., 2000](#)).

Neste trabalho, iremos analisar 8 sistemas caóticos multivariados, sendo 5 deles tridimensionais e outros 3 são bidimensionais, a fim de demonstrar a invariabilidade da técnica desenvolvida quanto a dimensionalidade do sistema avaliado. Os sistemas tridimensionais são B_7 , Burke, Lorenz, Complex Butterfly, Labyrinth e Linear Feedback; enquanto isso, os sistemas bidimensionais são Delayed Logistic, Henon Map e Tinkerbell.

O modelo matemático de Lorenz foi criado como um modelo simplificado para a convecção atmosférica [Lorenz \(1963\)](#), tendo diversas aplicações práticas como na geração de modelos simplificados para lasers [Haken \(1975\)](#), circuitos elétricos ([Cuomo et al., 1993](#)) e motores de corrente contínua sem escova ([Hemati, 1994](#)).

O sistema B_7 foi proposto por ([Chlouverakis and Sprott, 2005](#)) como um campo vetorial caótico abstrato, desenvolvido com o propósito de ser usado no estudo estatístico da relação entre os expoentes de Lyapunov e a dimensão de correlação D_2 .

O atrator caótico de Burke-Shaw foi criado por Bill Burke e Robert Shaw derivado do sistema caótico de Lorenz ([Shaw, 1981](#)). Este sistema é um companheiro do sistema Lorenz, no sentido de pertencer à mesma classe de sistemas. A principal diferença entre o sistema de Burke & Shaw e o sistema de Lorenz não está em suas equações, mas na maneira como são organizadas em torno do eixo z ([Letellier et al., 2005](#)). Ele é caracterizado ser por um modelo com quatro ramificações ([Letellier et al., 1996](#)).

O sistema intitulado Complex Butterfly é também derivado do sistema caótico de Lorenz, mas, ao contrário do sistema de Lorenz original, onde a dinâmica caótica permanece confinada ao semi-espço positivo em relação à variável de estado devido a um efeito limitante, a modelagem do Complex Butterfly permite a oscilação bipolar dessa variável de estado ([Elwakil et al., 2002](#)).

O Labyrinth Chaos foi proposto por [Thomas \(1999\)](#) como um fluxo tridimensional particularmente simples e matematicamente elegante. O sistema é representativo de uma grande classe de modelos autocatalíticos que ocorrem frequentemente em reações químicas ([Rasmussen et al., 1990](#)), ecologia ([Deneubourg and Goss, 1989](#)) e evolução ([Kauffman et al., 1993](#)).

O sistema caótico intitulado Linear Feedback Shift Register (LFSR) é um modelo matemático utilizado na teoria da informação, criptografia e geração de sequências pseudoaleatórias ([Massey, 1969](#)). O LFSR é uma forma especializada de registro de deslocamento que emprega um mecanismo de feedback linear para gerar sequências binárias ([Savir and McAnney, 1990](#)).

O sistema caótico Delayed Logistic é um modelo matemático que exhibe comportamento caótico e é uma extensão do mapa logístico clássico com a inclusão de um atraso temporal ([Wu and Baleanu, 2015](#)). Esse sistema tem sido estudado em diferentes contextos de dinâmica não linear e teoria do caos.

O sistema Hénon Map, também chamado de atrator/mapa de Hénon–Pomeau, é um sistema dinâmico de tempo discreto ([Hsu, 2013](#)). É um dos exemplos mais estudados de sistemas dinâmicos que exibem comportamento caótico. O mapa foi introduzido por Michel Hénon como um modelo simplificado da seção de Poincaré do modelo de Lorenz ([Lozi, 1978](#)).

O sistema caótico Tinkerbell foi nomeado em homenagem à sua semelhança com a forma de uma fada chamada Tinkerbell. Ele forma um atrator caótico com uma estrutura fractal distintiva, que é a assinatura visual do comportamento caótico e tem sido estudado em diversos campos, como física, matemática e dinâmica de sistemas ([dos Santos Coelho and Mariani, 2012](#)).

2.2 Modelagem Matemática

Para este estudo, utilizamos diferentes classes de sistemas para ilustrar a aplicação e efetividade do MvCECP: (a) Sistemas Caóticos, (b) Processos Estocásticos e (c) Dados Periódicos. A seguir, iremos detalhar a modelagem matemática de cada sistema que compõe a base de dados:

1. Sistemas Caóticos:

- O sistema caótico B_7 , dado por

$$\begin{cases} \dot{x} = K + z(x - \alpha y) \\ \dot{y} = z(\alpha x - \varepsilon y) \\ \dot{z} = 1 - x - y^2 \end{cases} \quad (2.1)$$

com $K = 0.5$, $\alpha = 7$ e $\varepsilon = 0.23$.

- O sistema caótico Burke-Shaw, dado por

$$\begin{cases} \dot{x} = -s(x + y) \\ \dot{y} = -y - sxz \\ \dot{z} = sxy + v \end{cases} \quad (2.2)$$

onde foi adotado $s = 10$ e $v = 13$.

- O sistema intitulado Complex Butterfly

$$\begin{cases} \dot{x} = a(y - x) \\ \dot{y} = -Kz \\ \dot{z} = |x| - 1 \end{cases} \quad (2.3)$$

e sendo $a = 0.55$, com

$$K = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases} \quad (2.4)$$

- O sistema intitulado Labyrinth Chaos

$$\begin{cases} \dot{x} = \sin(y) - bx \\ \dot{y} = \sin(z) - by \\ \dot{z} = \sin(x) - bz \end{cases} \quad (2.5)$$

onde foi adotado $b = 0$.

- O sistema intitulado LFSR

$$\begin{cases} \dot{x} = -0.4x + y + 10yz \\ \dot{y} = -x - 0.4y + 5xz \\ \dot{z} = ax - 5xy \end{cases} \quad (2.6)$$

onde foi adotado $a = 0.175$.

- O sistema caótico Lorenz

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = -xz + rx - y \\ \dot{z} = xy - bz \end{cases} \quad (2.7)$$

com $\sigma = 16$, $b = 4$ e $r = 45.92$.

- O sistema caótico Delayed Logistic

$$\begin{cases} x_{k+1} = rx_k(1 - y_k) \\ y_{k+1} = x_k \end{cases} \quad (2.8)$$

adotando-se $r = 2.28$.

- O sistema intitulado Hénon Map

$$\begin{cases} x_{k+1} = 1 - ax_k^2 + y_k \\ y_{k+1} = bx_k \end{cases} \quad (2.9)$$

com $a = 1.7$ e $b = 0.5$.

- O sistema caótico Tinkerbell

$$\begin{cases} x_{k+1} = x_k^2 - y_k^2 + ax_k + by_k \\ y_{k+1} = 2x_k y_k + cx_k + dy_k \end{cases} \quad (2.10)$$

com $a = 0.9$, $b = -0.6$, $c = 2.0$ e $d = 0.5$.

2. Processos Estocásticos:

- O ruído de potência $f^{(-k)}$ é gerado da seguinte forma: (i) A função rand do MATLAB é usada para produzir números pseudoaleatórios. (ii) Então, a partir da Transformada Rápida de Fourier se obtém y_1^k , com y_2^k sendo igual a $y_1^k \cdot pf^{-\frac{k}{2}}$. (iii) y_2^k é simetrizado para obter uma função real, e então, aplica-se a FFT reversa pertinente para obter o k-noise esperado. Para a avaliação multivariada, geramos três séries

temporais $f^{(-k)}$ independentes, y_1^k , y_2^k e y_3^k , e então construímos um sistema tridimensional, $Y = [y_1^k, y_2^k, y_3^k]$.

- Ruído Branco Gaussiano Tridimensional: Aqui geramos um ruído branco Gaussiano trivariado onde três canais de dados são mutuamente independentes, ou seja, $G = [g_1, g_2, g_3]$.

3. Dados Periódicos

- Sistema temporal tridimensional composto por séries geradas pela função $\sin(t)$: Para obter séries temporais periódicas, geramos três sinais independentes $x_i = \sin(0.2\pi t) + b_i$, sendo $i = \{1, 2, 3\}$ com mesmo período e diferentes amplitudes. Define-se $b_1 = 0$, $b_2 = 0.5$ e $b_3 = -0.5$. Finalmente, a série temporal periódica tridimensional é $X = [x_1, x_2, x_3]$.

3

Teoria da Informação

3.1 PDF de Bandt-Pompe

Dada uma série temporal unidimensional qualquer, seu estudo e caracterização se inicia por meio da construção de sua Função de Distribuição de Probabilidade (PDF). [Bandt and Pompe \(2002\)](#) introduziram uma metodologia bem-sucedida para a avaliação da PDF associada a dados escalares de séries temporais usando uma técnica de simbolização.

Os dados simbólicos pertinentes são (i) criados pela classificação dos valores da série e (ii) definidos pela reordenação dos dados incorporados em ordem crescente, o que equivale a uma reconstrução do espaço de fase com dimensão de incorporação (comprimento do padrão) D e atraso de imersão τ ([Zanin et al., 2012](#)). Desta forma é possível quantificar a diversidade dos símbolos de ordenação (padrões) derivados de uma série temporal escalar.

Observe que a sequência de símbolos apropriada surge naturalmente da série temporal e nenhuma suposição baseada em modelo é necessária. Na verdade, as “partições” necessárias são criadas comparando a ordem dos valores relativos vizinhos. Entretanto, fica claro que esse tipo de análise de séries temporais acarreta a perda de alguns detalhes da informação de amplitude da série original.

As vantagens adicionais do método residem em (i) sua simplicidade (são necessários poucos parâmetros: o padrão/dimensão de incorporação D e o atraso de imersão τ e (ii) a natureza extremamente rápida do processo de cálculo pertinente ([Keller and Sinn, 2005](#)).

A distribuição de probabilidade P gerada pela técnica de BP é obtida uma vez que escolhermos a dimensão de incorporação D e o atraso de imersão τ . O primeiro parâmetro desempenha um papel importante na avaliação da distribuição de probabilidade apropriada, uma vez que D determina o número de estados acessíveis, dado por $D!$. Além disso, foi estabelecido que o comprimento N da série temporal deve satisfazer a condição $N \geq D!$ a fim de obter estatísticas confiáveis. Com relação à seleção dos parâmetros, BP sugerem trabalhar com $3 \leq D \leq 7$ e com um intervalo de tempo $\tau = 1$.

No entanto, outros valores de τ podem fornecer informações adicionais, como demonstraram Soriano et al. (2011) e Zunino et al. (2010a). Quando este parâmetro é relevante, ele está fortemente correlacionado com as escalas de tempo intrínsecas do sistema em análise.

Com relação a modelagem matemática, dada uma série temporal unidimensional da forma $\{x(n), n = 1, 2, 3, \dots, N\}$, uma dimensão de incorporação D e um atraso $\tau = 1$, a série reconstruída é denotada por

$$X(i) = \{x(i), x(i+1), x(i+2), \dots, x(i+D-1)\} \quad (3.1)$$

onde $i = 1, 2, 3, \dots, N - D + 1$. Então, $X(i)$ pode ser rearranjado em ordem crescente de modo a determinar um padrão ordinal $\pi = \{r_0, r_1, \dots, r_{D-1}\}$, sendo:

$$x_{i+r_0} < x_{i+r_1} < \dots < x_{i+r_{D-1}} \quad (3.2)$$

Desse modo, como cada padrão é relativo a uma possível ordenação do conjunto de dimensão D , então há $D!$ possíveis padrões. Por exemplo, tomando $D = 3$, há 6 possíveis padrões ordinais: $\{\pi_1, x_1 \leq x_2 \leq x_3\}$, $\{\pi_2, x_1 \leq x_3 \leq x_2\}$, $\{\pi_3, x_2 \leq x_1 \leq x_3\}$, $\{\pi_4, x_2 \leq x_3 \leq x_1\}$, $\{\pi_5, x_3 \leq x_1 \leq x_2\}$ e $\{\pi_6, x_3 \leq x_2 \leq x_1\}$. Se adotarmos $\pi_j = \{j, j = 1, 2, 3, \dots, D!\}$, podemos obter os padrões da série $\{s(i), i = 1, 2, 3, \dots, N - D + 1\}$. Se o padrão ordinal de $X(i)$ é π_j , então $s(i) = j$. Então, temos que a distribuição de probabilidade de Bandt-Pompe $P(\pi_j)$ é dada por

$$P(\pi_j) = \frac{\#\{s | s \leq N - D + 1; s = j\}}{N - D + 1} \quad (3.3)$$

sendo $\#$ o símbolo de cardinalidade do conjunto.

3.2 Entropia de Shannon

Em busca de quantificar aspectos estruturais e dinâmicos de um sinal qualquer, (Shannon, 1948) iniciou os estudos de Teoria da Informação, onde propôs um dos principais quantificadores baseando-se na noção de entropia. Dada uma PDF $f(x)$ com $x \in \Omega \subset \mathbb{R}$ e $\int_{\Omega} f(x) dx = 1$, sua Entropia de Shannon é definida por

$$\mathcal{S}(f) = - \int_{\Omega} f(x) \ln f(x) dx \quad (3.4)$$

Com isso, para uma série temporal $X(t) = \{x(n), n = 1, 2, 3, \dots, D!\}$, sendo $D!$ um conjunto de medidas observáveis de X e sua PDF, dada por $P = \{p_j, j = 1, \dots, N\}$, com $\sum_{j=1}^N p_j = 1$ e sendo N o número de estados possíveis de um sistema estudado, a medida de

informação de (Shannon, 1948) é definida por

$$\mathcal{S}[P] = - \sum_{j=1}^{D!} p_j \ln(p_j) \quad (3.5)$$

Nosso conhecimento de um determinado sistema é máximo quando a Entropia de Shannon é igual a zero, assim podemos prever com total certeza quando o estado j irá de fato acontecer, dada a probabilidade p_j . Ao contrário, o conhecimento do sistema é mínimo quando temos uma distribuição uniforme $P_e = \{p_j = 1/N, \forall j = 1, \dots, N\}$. A distribuição P_e denota a mesma que P para uma apropriada referência ou estado de equilíbrio. Assim, definimos a Entropia de Shannon normalizada $0 \leq \mathcal{H} \leq 1$ como

$$\mathcal{H}[P] = \mathcal{S}[P]/\mathcal{S}[P_e] = \mathcal{S}[P]/\ln N \quad (3.6)$$

Assim, a Entropia \mathcal{H} de uma série temporal pode ser interpretada de acordo com as características de sua distribuição de probabilidade, baseando-se na ordenação de seus elementos.

3.3 Complexidade Estatística

Em busca de compreender a estrutura física do sistema complexo, a Complexidade Estatística \mathcal{C} mede o desequilíbrio do sistema baseado no conteúdo da informação descrita pela distribuição de probabilidade e pela entropia (Lopez-Ruiz et al., 1995).

(Lamberti et al., 2004) desenvolveram uma medida de complexidade estatística \mathcal{C} , capaz de: (i) Detectar detalhes essenciais da dinâmica do sistema avaliado e (ii) diferenciar entre caos e periodicidade (em diferentes graus). Essa medida de complexidade fornece informações adicionais sobre a função de distribuição de probabilidade que não são detectadas pela entropia.

Seguindo a notação apresentada por Lopez-Ruiz et al. (1995), temos que a expressão da complexidade estatística \mathcal{C} é dada por

$$\mathcal{C}[P] = Q_J[P, P_e] \cdot \mathcal{H}[P] \quad (3.7)$$

Medimos o desequilíbrio Q_J através do divergente de *Jensen-Shannon* (Fuglede and Topsoe, 2004), que quantifica a diferença entre duas (ou mais) distribuições de probabilidade, representado por

$$Q_J[P, P_e] = Q_0 \cdot \mathcal{J}[P, P_e], \quad (3.8)$$

com

$$\mathcal{J}[P, P_e] = \mathcal{S}[(P + P_e)/2] - \mathcal{S}[P]/2 - \mathcal{S}[P_e]/2, \quad (3.9)$$

onde Q_0 é uma constante de normalização ($0 \leq Q_j \leq 1$) igual ao inverso do valor máximo de $J[P, P_e]$, dado pela expressão

$$Q_0 = -2 \left\{ \left(\frac{n+1}{n} \right) \log(n+1) - 2\log(2n) + \log(n) \right\}^{-1} \quad (3.10)$$

em que n é o número total de estados acessíveis, nesse caso, $n = D!$.

A complexidade estatística de um sistema é nula em duas situações opostas: no conhecimento perfeito ou na aleatoriedade completa, qualquer sistema mensurado ficará entre essas configurações extremas.

3.4 Plano de Causalidade Complexidade-Entropia (CECP)

Na mecânica estatística, muitas vezes estamos interessados em sistemas isolados caracterizados por uma distribuição de probabilidade inicial, arbitrária e discreta. A evolução em direção ao equilíbrio deve ser descrita como o objetivo primordial. No equilíbrio, podemos pensar, sem perda de generalidade, que este estado é dado pela distribuição uniforme P_e .

A evolução temporal da Complexidade Estatística pode ser analisada usando um diagrama de C versus tempo t . No entanto, é bem conhecido que a segunda lei da termodinâmica afirma que para sistemas isolados a entropia cresce monotonicamente com o tempo ($dHS/dt \geq 0$) (Plastino and Plastino, 1996). Isso implica que a Entropia de Shannon pode ser considerada como um paralelo do tempo, de modo que uma forma equivalente de estudar a evolução temporal da Complexidade Estatística intensivo é por meio da análise do C versus \mathcal{H} . Desta forma, o eixo da entropia normalizado substitui o eixo do tempo. Além disso, foi demonstrado que para um dado valor de \mathcal{H} , a faixa de possíveis valores de complexidade estatística varia entre um mínimo C_{min} e um máximo C_{max} (Martin et al., 2006), restringindo os possíveis valores de complexidade neste plano. Portanto, a avaliação da complexidade fornece informações adicionais sobre os detalhes da distribuição de probabilidade do sistema, que não é discriminada por medidas de aleatoriedade como a entropia (Rosso et al., 2007b; Feldman et al., 2008).

A complexidade também pode ajudar a descobrir informações relacionadas às estruturas correlacionais relacionadas aos componentes do processo físico em estudo (Rosso and Masoller, 2009). O diagrama de Entropia x Complexidade ($\mathcal{H} \times C$), também intitulado como Plano Causal de Complexidade-Entropia (CECP), tem sido utilizado para estudar mudanças na dinâmica de um sistema originadas por modificações de alguns parâmetros característicos (Zunino et al., 2012, 2010b; Rosso et al., 2010).

3.5 PDF Multivariada

3.5.1 Padrões Ordinais

Dado um sistema multivariado qualquer, tomemos sua representação como sendo da forma $\{X_j(n), n = 1, 2, \dots, N \mid j = 1, 2, \dots, \alpha\}$, onde α é o número de componentes que integram o sistema. Como os intervalos de flutuação de cada uma dessas séries podem ser diferentes, iremos normalizá-las, a fim de garantir uma avaliação conjunta utilizando uma única escala de referência. Para este propósito, empregamos a normalização min-max (Patro and Sahu, 2015), que é definida como

$$\tilde{X}_j = \frac{x_j - \min(x_j)}{\max(x_j) - \min(x_j)}, \quad (3.11)$$

onde $\max(\cdot)$ é a função de máximo e $\min(\cdot)$ é a função de mínimo. Após a normalização, todas as séries temporais componentes do sistema terão valores compreendidos dentro do intervalo fechado $[0, 1]$.

Então, dado o sistema multivariado X_j , uma dimensão de incorporação multivariada σ e um atraso de imersão $\tau = 1$, é possível avaliar os padrões ordinais utilizando uma janela deslizante, assim como na abordagem univariada, que cubra um conjunto com σ pontos, sendo eles formados pela junção das componentes do sistema multivariado em um instante de tempo qualquer.

Na figura 3.1 comparamos a forma de construir as janelas deslizantes para percorrer os dados nas abordagens univariada e multivariada para o caso em que $D = \sigma = 3$. Como o R^n não é um corpo ordenável para $n > 1$, utilizaremos como métrica a norma de cada vetor que é determinado pela origem do espaço em direção aos pontos do sistema avaliado.

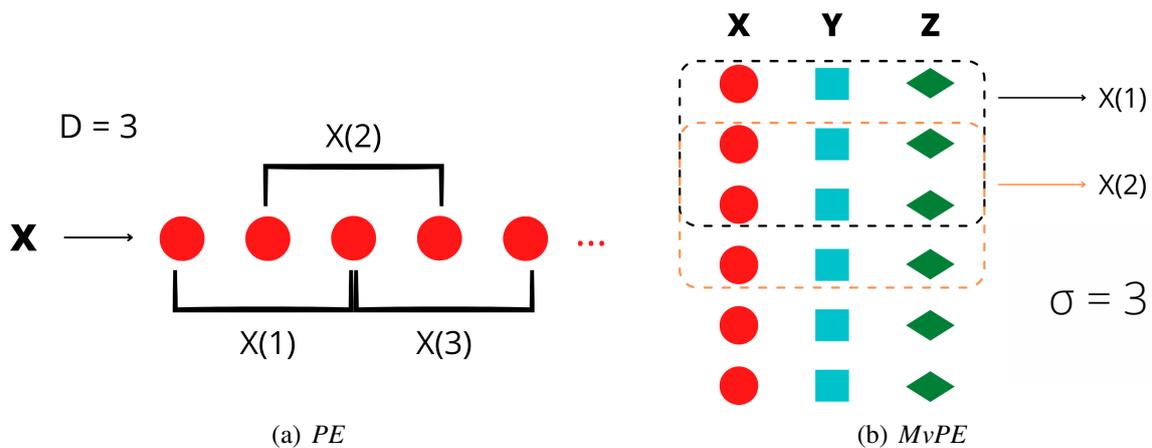


Figura 3.1: Ilustração gráfica do funcionamento das janelas de dados deslizantes utilizadas nas abordagens univariada (a) e multivariada (b), com $D=\sigma=3$

Com essa adaptação, passa a ser possível fazer um estudo dos padrões ordinais de sistemas multivariados, determinando as categorias dos pontos do sistema a partir da ordenação das respectivas normas dos vetores associados. Dessa forma, a avaliação adotada passa a observar a correlação entre as séries temporais componentes do sistema.

Para cálculo de distância, adotaremos neste trabalho a expressão para cálculo da distância euclidiana (Danielsson, 1980). Como os dados foram previamente normalizados, o valor da norma de cada vetor observado é da forma

$$d_n = \sqrt{\tilde{x}_1[n]^2 + \tilde{x}_2[n]^2 + \tilde{x}_3[n]^2 + \dots + \tilde{x}_\sigma[n]^2} \quad (3.12)$$

visto que estamos calculando a distância do ponto até a origem do espaço.

Assim, a partir da ordenação dos módulos desses vetores, poderemos determinar os padrões ordinais associados a cada rearranjo da série reconstruída a partir de subconjuntos com dimensão σ .

3.5.2 Padrões de Distância

Como os padrões ordinais nem sempre podem detectar mudanças nas séries temporais de forma sensível (Yin and Shang, 2014), passamos a projetar subpadrões capazes de detectar tais mudanças (He et al., 2016). Para isso, dividimos cada padrão ordinal em mais situações, comparando as distâncias entre os elementos envolvidos a fim de obter ainda mais detalhes dos sistemas analisados. Inicialmente, iremos analisar o estudo dos subpadrões de distância em um caso univariado, a fim de entender como ele funciona e replicar o processo para estudos multidimensionais.

Tomando $\sigma = 3$, sabemos que existem 6 possíveis ordenações dos elementos de uma sequência $x(n)$, como indicado na figura 3.2.

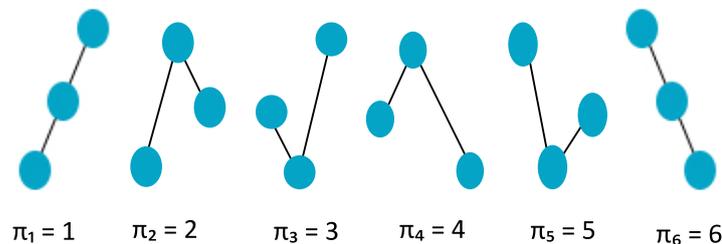


Figura 3.2: As 6 configurações de padrões ordinais para o caso em que $\sigma = 3$

Como os dados sempre são normalizados, neste caso, qualquer elemento se encontra no segmento fechado de $[0, 1]$. A fim de analisar as relações de distância entre esses 3 elementos, dada uma ordenação qualquer entre eles, selecionaremos os elementos de maior e menor valor,

para posicionar sobre o segmento. Então, o elemento de valor intermediário poderá assumir um dos 3 estados dos padrões de distância: (i) ter um valor consideravelmente mais próximo ao máximo; (ii) ter um valor consideravelmente mais próximo ao mínimo ou (iii) ocupar uma região intermediária.

Para delimitar essas regiões, utilizaremos 2 valores de referência, Top e Bot, que irão dividir o segmento existente entre os valores máximo e mínimo em 3 segmentos de igual comprimento, partes do original, que demarcarão os estados baseados nos padrões de distância que foram supracitados.

Na figura 3.3 podemos perceber que a partir do padrão de incorporação $\sigma = 3$, existem 18 possíveis estados acessíveis ($M = 18$). As expressões para determinar os escalares de referência, denominados Top e Bot, a fim de dividir o segmento entre os valores máximo e mínimo em 3 regiões de igual comprimento, são dadas por:

$$\begin{cases} Top_n = \frac{2}{3}max(x(n)) + \frac{1}{3}min(x(n)) \\ Bot_n = \frac{1}{3}max(x(n)) + \frac{2}{3}min(x(n)) \end{cases} \quad (3.13)$$

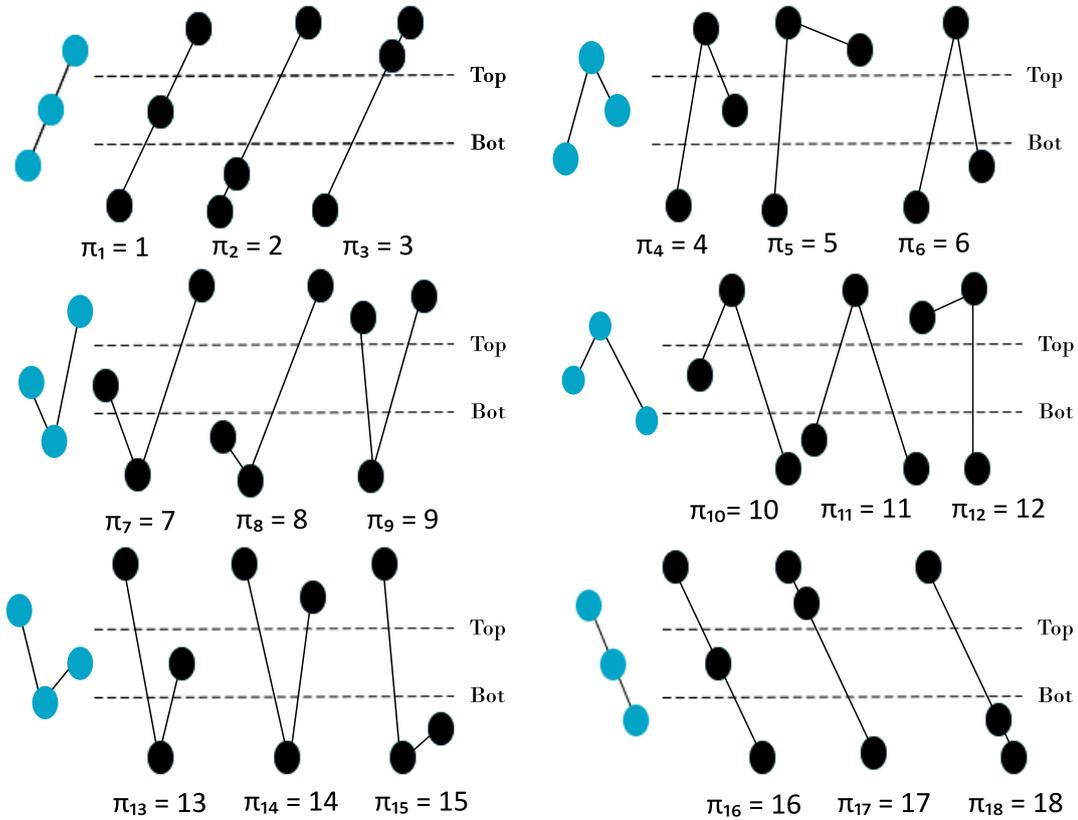


Figura 3.3: Padrões de ordem para dimensão de incorporação $\sigma = 3$ com seus subpadrões π_{18} em um sistema univariado

Podemos perceber observando as expressões que os valores de referência Top e Bot são obtidos a partir de operações envolvendo os valores de máximo e mínimo da sequência e por isso, fixar suas posições inicialmente é fundamental para esta análise.

Analogamente, para o caso em que $\sigma = 4$, os elementos de máximo e mínimo serão posicionados no intervalo $[0, 1]$ e o segmento determinado por eles será dividido em 4 partes por meio de 3 valores de referência, neste exemplo intitulados E_1 , E_2 e E_3 . Tomando os elementos como sendo A , B , C e D e a ordenação em que $A > B > C > D$, temos que os elementos B e C podem ocupar qualquer uma das regiões que foram delimitadas no segmento \overline{AB} . Neste caso, os valores das expressões para determinar os valores de referência, garantindo que todos tenham o mesmo comprimento, são:

$$\begin{cases} E_1 = \frac{3}{4}A + \frac{1}{4}D \\ E_2 = \frac{2}{4}A + \frac{2}{4}D \\ E_3 = \frac{1}{4}A + \frac{3}{4}D \end{cases} \quad (3.14)$$

O caso particular da ordenação $A > B > C > D$, com $\sigma = 4$ é exibido na figura 3.2. Podemos notar que há 10 possibilidades para cada ordenação, então, como há $4! = 24$ possíveis ordenações dos elementos, temos que para $\sigma = 4$, o número de estados acessíveis é $M = 240$.

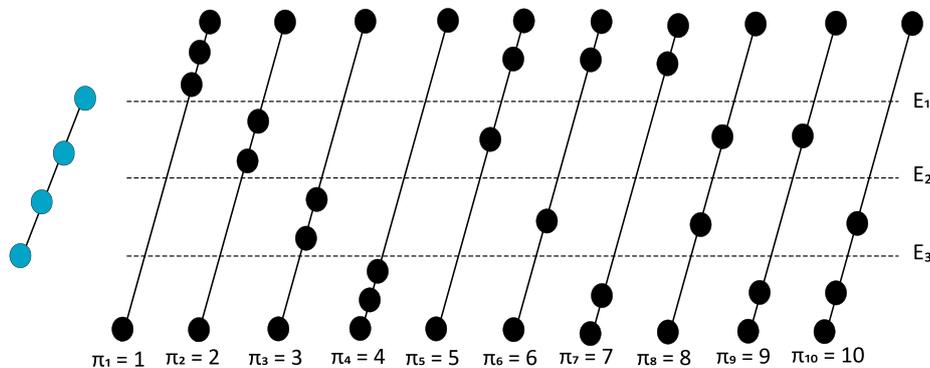


Figura 3.4: As 10 possibilidades de padrões de distância existentes para uma dada ordenação dos elementos no caso em que $\sigma = 4$ em um sistema univariado

Generalizando esse processo no espaço unidimensional, dado um valor qualquer de σ e uma sequência de elementos $x(n)$ com σ elementos, podemos dividir o segmento entre os elementos de máximo e mínimo em σ regiões de mesmo comprimento, utilizando $\sigma - 1$ valores de referência. Então, as expressões serão da forma:

$$\begin{cases} E_1 = \frac{\sigma-1}{\sigma} \cdot \max(x(n)) + \frac{1}{\sigma} \cdot \min(x(n)) \\ E_2 = \frac{\sigma-2}{\sigma} \cdot \max(x(n)) + \frac{2}{\sigma} \cdot \min(x(n)) \\ E_3 = \frac{\sigma-3}{\sigma} \cdot \max(x(n)) + \frac{3}{\sigma} \cdot \min(x(n)) \\ \vdots \\ E_{\sigma-1} = \frac{1}{\sigma} \cdot \max(x(n)) + \frac{\sigma-1}{\sigma} \cdot \min(x(n)) \end{cases} \quad (3.15)$$

É fácil perceber a partir das expressões acima que o maior e o menor número da sequência sempre ocupam as regiões extremas, sendo usados como referências para determinar os valores que irão dividir o segmento formado por eles. Desse modo, eles sempre ocupam a região mais distante da origem e a mais próxima, respectivamente, enquanto os demais pontos podem ocupar qualquer uma das regiões delimitadas por eles. Assim, há sempre $\sigma - 2$ números que podem ocupar qualquer uma das σ regiões delimitadas. Ou seja, há

$$M = \sigma! CR_{\sigma}^{\sigma-2} = \sigma! P_{2\sigma-3}^{\sigma-2, \sigma-1} = \frac{\sigma!(2\sigma-3)!}{(\sigma-2)!(\sigma-1)!} = \frac{\sigma \cdot (2\sigma-3)!}{(\sigma-2)!} \quad (3.16)$$

possibilidades de distribuição, sendo CR o número de combinações completas (MORGADO et al., 2006).

Como o cálculo de M é uma função de σ e considera somente o número de elementos e espaços disponíveis, ele independe do número de componentes do sistema avaliado.

Em seguida, a fim de generalizar esse processo para sistemas multivariados, iremos analisar como o espaço pode ser dividido para os casos de 2 ou 3 dimensões, visto que já observamos como isto acontece no caso unidimensional.

Caso a análise se refira ao espaço bidimensional, ou seja, sistemas com 2 dimensões, precisamos garantir que as σ repartições contidas no plano possuam a mesma área. Nesse caso, dentro do plano, as regiões serão delimitadas por quartos de circunferência centradas na origem, sendo o raio a própria norma dos vetores determinados pelos pontos da sequência.

Então, tomando $\sigma = 3$, sejam A e B os pontos de mínimo e máximo com r e R sendo os raios das circunferências centradas na origem e passando por A e B , respectivamente. Na Figura 3.5 é possível ver uma representação gráfica do plano com os pontos A e B posicionados e suas respectivas circunferências associadas.

O "ponto livre" C irá ocupar uma das 3 regiões determinadas pelos raios α e β , denominadas A_1 , A_2 ou A_3 . Neste caso, é necessário garantir que as 3 regiões tenham a mesma área, para que a probabilidade de C ocupar qualquer uma delas seja igual, ou seja, os eventos devem ser equiprováveis. Para garantir que as regiões A_1 , A_2 e A_3 tenham a mesma área, iremos determinar os raios α e β que assegurem essa condição. Assim, obtemos o seguinte sistema de equações:

$$\frac{\pi(\alpha^2 - r^2)}{4} = \frac{\pi(\beta^2 - \alpha^2)}{4} \quad (3.17)$$

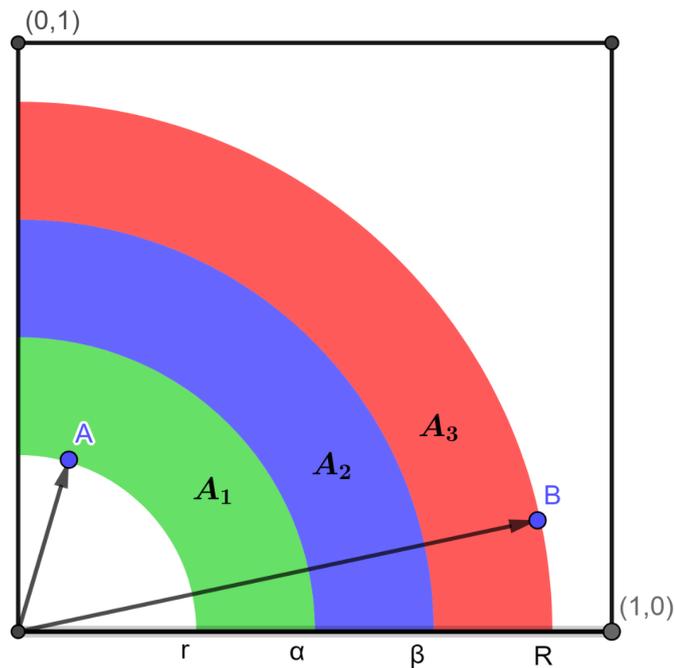


Figura 3.5: Plano dividido em $\sigma = 3$ regiões para avaliação de padrões de distância de sistemas bidimensionais

$$\frac{\pi(\beta^2 - \alpha^2)}{4} = \frac{\pi(R^2 - \beta^2)}{4} \quad (3.18)$$

$$\frac{\pi(\alpha^2 - r^2)}{4} = \frac{\pi(R^2 - \beta^2)}{4} \quad (3.19)$$

Solucinando o sistema composto pelas equações geradas, podemos determinar os valores de α e β em função de r e R , tal que:

$$\alpha = \sqrt{\left(\frac{2r}{3}\right)^2 + \left(\frac{R}{3}\right)^2} \quad (3.20)$$

$$\beta = \sqrt{\left(\frac{2R}{3}\right)^2 + \left(\frac{r}{3}\right)^2} \quad (3.21)$$

Podemos repetir o processo para os demais valores de σ em sistemas bidimensionais, a fim de obter expressões generalizadas, assim como no caso unidimensional. Analogamente, para uma sequência normalizada $x(n)$ e um σ qualquer, seguindo os passos descritos, obteremos as seguintes expressões:

$$\begin{cases} E_1 = \sqrt{\frac{\sigma-1}{\sigma} \cdot \max(x(n))^2 + \frac{1}{\sigma} \cdot \min(x(n))^2} \\ E_2 = \sqrt{\frac{\sigma-2}{\sigma} \cdot \max(x(n))^2 + \frac{2}{\sigma} \cdot \min(x(n))^2} \\ E_3 = \sqrt{\frac{\sigma-3}{\sigma} \cdot \max(x(n))^2 + \frac{3}{\sigma} \cdot \min(x(n))^2} \\ \vdots \\ E_{\sigma-1} = \sqrt{\frac{1}{\sigma} \cdot \max(x(n))^2 + \frac{\sigma-1}{\sigma} \cdot \min(x(n))^2} \end{cases} \quad (3.22)$$

onde $\max(x(n))$ e $\min(x(n))$ representam a maior e menor norma, respectivamente, dentre os vetores de uma sequência de dimensão σ .

Por fim, podemos adaptar os passos adotados para avaliação de sistemas tridimensionais, a fim de encontrar as expressões para segmentar o espaço adequadamente, produzindo regiões com o mesmo volume. Neste caso, a representação tridimensional é exibida na figura 3.6 com $\sigma = 3$, onde os pontos J e K representam, respectivamente, os pontos de mínimo e máximo e o "ponto livre" pode ocupar qualquer uma das 3 regiões determinadas no espaço, denominadas A_1 , A_2 e A_3 .

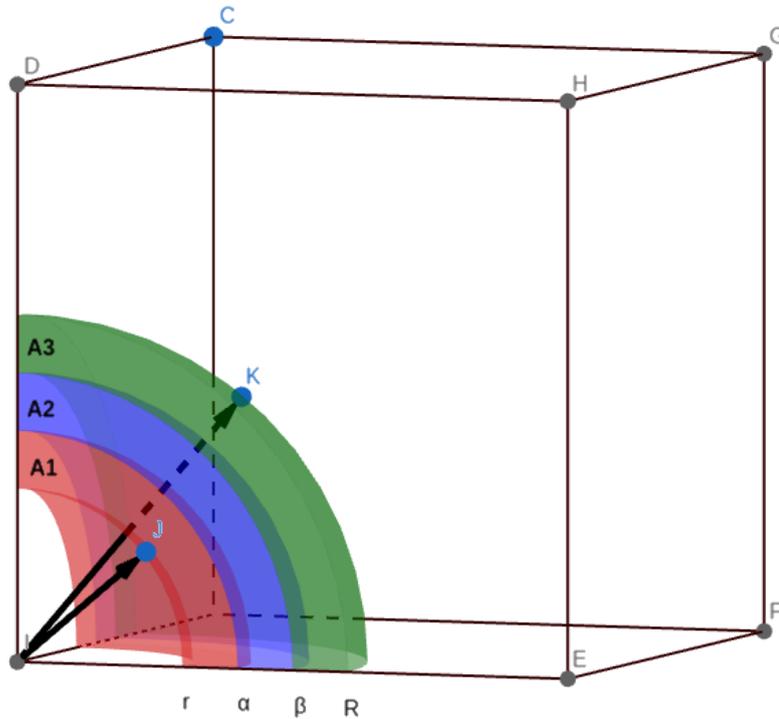


Figura 3.6: Espaço tridimensional dividido em $\sigma = 3$ regiões para avaliação de padrões de distância

Para assegurar que os volumes das regiões A_1 , A_2 e A_3 sejam iguais, temos:

$$\frac{\pi(\alpha^3 - r^3)}{3} = \frac{\pi(\beta^3 - \alpha^3)}{3} \quad (3.23)$$

$$\frac{\pi(\beta^3 - \alpha^3)}{3} = \frac{\pi(R^3 - \beta^3)}{3} \quad (3.24)$$

$$\frac{\pi(\alpha^3 - r^3)}{3} = \frac{\pi(R^3 - \beta^3)}{3} \quad (3.25)$$

Solucinando o sistema composto pelas equações geradas, podemos determinar os valores de α e β em função de r e R , tal que:

$$\alpha = \sqrt[3]{\left(\frac{2r}{3}\right)^3 + \left(\frac{R}{3}\right)^3} \quad (3.26)$$

$$\beta = \sqrt[3]{\left(\frac{2R}{3}\right)^3 + \left(\frac{r}{3}\right)^3}. \quad (3.27)$$

Podemos então seguir a mesma linha para os demais valores de σ em sistemas tridimensionais, a fim de obter expressões generalizadas, assim como nos casos unidimensional e bidimensional. Analogamente, para uma sequência normalizada $x(n)$ e um σ qualquer, seguindo os passos descritos, obteremos as seguintes expressões:

$$\begin{cases} E_1 = \sqrt[3]{\frac{\sigma-1}{\sigma} \cdot \max(x(n))^3 + \frac{1}{\sigma} \cdot \min(x(n))^3} \\ E_2 = \sqrt[3]{\frac{\sigma-2}{\sigma} \cdot \max(x(n))^3 + \frac{2}{\sigma} \cdot \min(x(n))^3} \\ E_3 = \sqrt[3]{\frac{\sigma-3}{\sigma} \cdot \max(x(n))^3 + \frac{3}{\sigma} \cdot \min(x(n))^3} \\ \vdots \\ E_{\sigma-1} = \sqrt[3]{\frac{1}{\sigma} \cdot \max(x(n))^3 + \frac{\sigma-1}{\sigma} \cdot \min(x(n))^3} \end{cases} \quad (3.28)$$

Esse processo pode ser adaptado para sistemas n-dimensionais e as condições serão construídas de modo a satisfazer que todas as regiões delimitadas pelas expressões tenham o mesmo hipervolume (Gipple, 2014). A expressão para cálculo do número M de espaços acessíveis é a mesma, visto que foi construída de forma independente a estrutura do sistema analisado.

Seguindo os passos anteriores, construímos um algoritmo para a avaliação de padrões ordinais e subpadrões de dimensão em sistemas multivariados.

Consequentemente, podemos determinar uma função de densidade de probabilidade multivariada associada que mensure a ocorrência de cada evento s , sendo invariável ao número de dimensões do sistema, dada por:

$$P_{M_v}(\pi_j) = \frac{\#\{s | s \leq N - \sigma + 1; s = j\}}{N - \sigma + 1} \quad (3.29)$$

3.6 Quantificadores de Teoria da Informação Multivariados

Como ambos os quantificadores utilizados, \mathcal{H} e \mathcal{C} são funções da distribuição de probabilidade, a única mudança ocorre na variável que contabiliza o número de espaços acessíveis M , que é também a dimensão da PDF Multivariada.

Dessa forma, a entropia multivariada normalizada é dada por

$$\mathcal{H}_{Mv}[P_{Mv}] = - \sum_{j=1}^M p_j \ln(p_j) / \ln M \quad (3.30)$$

Paralelamente, podemos determinar a expressão da complexidade estatística multivariada \mathcal{C} como sendo da forma

$$\mathcal{C}_{Mv}[P_{Mv}] = Q_J[P_{Mv}, P_e] \cdot \mathcal{H}_{Mv}[P_{Mv}] \quad (3.31)$$

Medimos o desequilíbrio Q_J através do divergente de *Jensen-Shannon*, que quantifica a diferença entre duas (ou mais) distribuições de probabilidade, representado por

$$Q_J[P_{Mv}, P_e] = Q_0 \cdot \mathcal{J}[P_{Mv}, P_e], \quad (3.32)$$

com

$$\mathcal{J}[P_{Mv}, P_e] = \mathcal{S}_{Mv}[(P_{Mv} + P_e)/2] - \mathcal{S}_{Mv}[P_{Mv}]/2 - \mathcal{S}_{Mv}[P_e]/2, \quad (3.33)$$

onde Q_0 é uma constante de normalização ($0 \leq Q_{JS} \leq 1$) igual ao inverso do valor máximo de $\mathcal{J}[P_{Mv}, P_e]$, dado pela expressão

$$Q_0 = -2 \left\{ \left(\frac{M+1}{M} \right) \log(M+1) - 2 \log(2M) + \log(M) \right\}^{-1}. \quad (3.34)$$

Com isso, concluímos a proposição de uma perspectiva multivariada adaptada para a Entropia de Shannon e a Complexidade Estatística. Então, propomos o Plano Causal Multivariado de Complexidade-Entropia (MvCECP) para avaliação e caracterização de sistemas caóticos, estocásticos, periódicos e quaisquer outros conjuntos de séries temporais ($\mathcal{H}_{Mv} \times \mathcal{C}_{Mv}$).



Resultados e Discussões

4.1 Construção da Base de Dados

Para uma caracterização objetiva e perceptível, efetuamos as avaliações dos modelos de sistemas sob as abordagens univariada e multivariada a fim de poder levantar hipóteses sobre o comportamento dos dados em ambos os casos, comparando os resultados.

Inicialmente, foi efetuada a construção de um conjunto com dados de cada um dos sistemas caóticos, estocásticos ou periódicos descritos ao longo da seção 2.2. Cada sistema, seja ele bidimensional ou tridimensional, é gerado a partir de suas séries temporais componentes e cada uma delas foi gerada com 10^6 dados, seguindo sua respectiva modelagem e adotando $\tau = 1$, com auxílio da abordagem de Runge-Kutta-Fehlberg (Press et al., 1992) e condições iniciais dadas por Sprott and Sprott (2003). Um exemplo é exibido na figura 4.2, que ilustra o sistema caótico B_7 .

4.2 Análise Univariada com o CECP

O CECP foi inicialmente proposto por com o intuito de prover um mecanismo adequado para diferenciar sistemas caóticos e processos estocásticos. Apesar de terem uma origem física bastante diferente, as séries temporais decorrentes de sistemas caóticos (CS) compartilham várias propriedades com as séries geradas por processos estocásticos (SP) que as tornam quase indistinguíveis: (1) um espectro de potência de banda larga (PS), (2) uma função de autocorrelação do tipo delta, (3) um comportamento irregular dos sinais medidos, etc. Com isso, essa similaridade tornou possível substituir SP por CS nas mais variadas aplicações práticas (Rosso et al., 2007b).

Posteriormente, o CECP foi utilizado na caracterização de séries temporais decorrentes de mapas caóticos. A localização do sistema no plano (quantificador global) exhibe características específicas típicas associadas à natureza de sua dinâmica (Rosso et al., 2013).

Como o CECP efetua caracterização de séries temporais univariadas, para se analisar um sistema multivariado é necessário efetuar as análises isoladas de cada componente para uma caracterização adequada desse sistema. Na figura 4.1, o sistema B_7 é exibido como exemplo, evidenciando os gráficos de suas séries temporais componentes x , y e z ao longo do tempo.

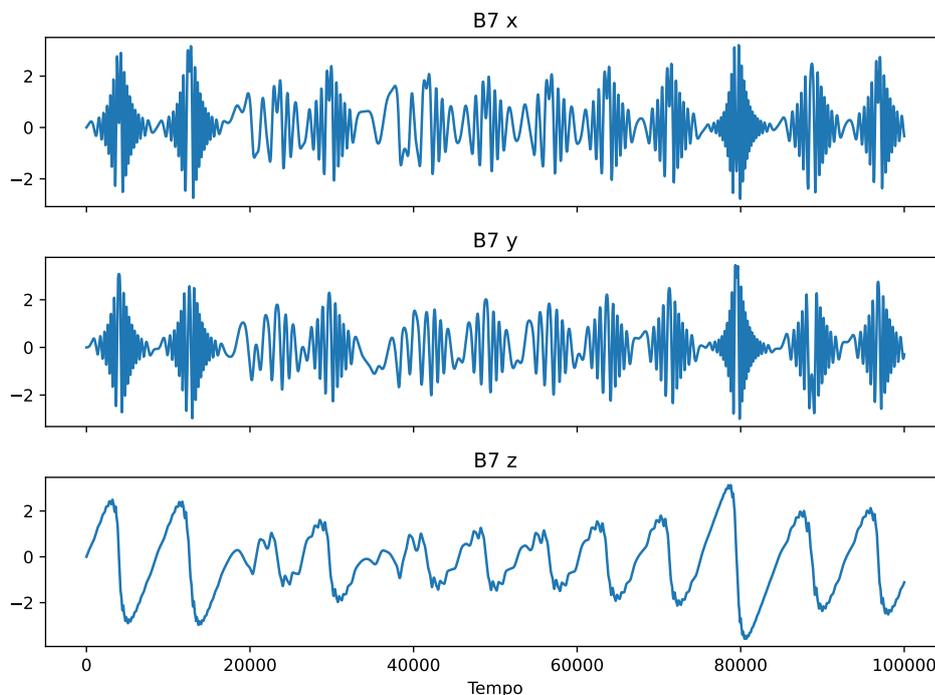


Figura 4.1: Componentes do Sistema Caótico B_7

Para a fase de análises univariadas, dada a dimensão das séries temporais, selecionamos a dimensão de incorporação $D = 6$ para garantir que a condição de validade seja atendida, em que $N \geq D!$. Neste caso, temos $10^6 \geq 720$, o que garante estatísticas confiáveis nos estudos. Na figura 4.2, exibimos os resultados encontrados para a avaliação dos sistemas supracitados, utilizando o CECP. Cada um dos sistemas é representado por uma cor diferente, enquanto que cada componente possui uma forma distinta: para as componentes x adotamos o círculo, para as componentes y utilizamos o quadrado e por fim, para as componentes z foram usados triângulos invertidos.

Na tabela 4.1 são exibidos os valores de entropia e complexidade para cada uma das componentes de todos os sistemas caóticos avaliados.

É fácil perceber que há uma série de sobreposições de informações na imagem, indicando que diversas séries temporais apresentam valores muito próximos de entropia e complexidade e, conseqüentemente, ocupando regiões similares no CECP, o que dificulta a caracterização adequada de cada sistema caótico avaliado.

Há casos de sistemas em que todas as suas componentes ocupam regiões muito próximas no plano, sejam eles tridimensionais, como ocorre com B_7 , Labyrinth e Lorenz, ou bidimensionais, como são os casos de Delayed Logistic e Hénon Map. Com relação ao posicionamento

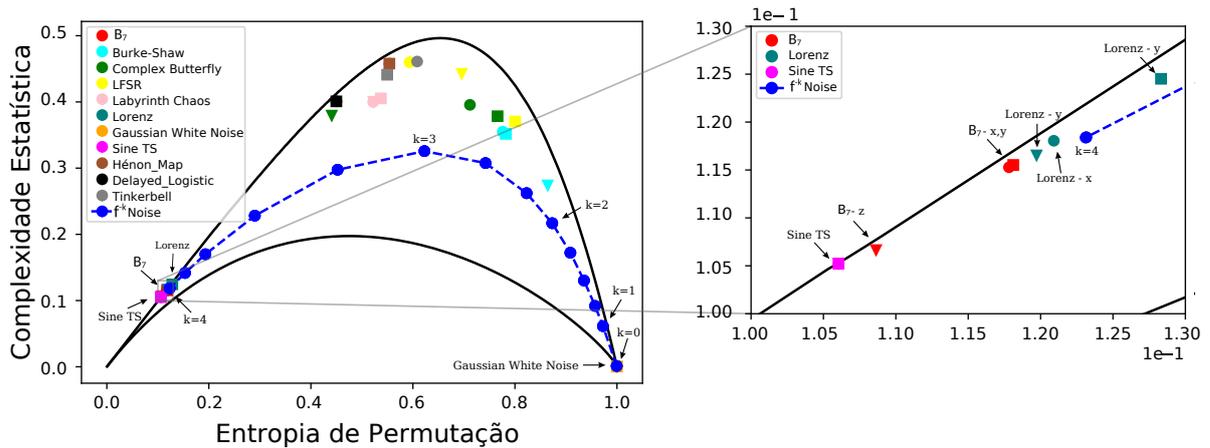


Figura 4.2: Avaliação das séries temporais componentes dos sistemas caóticos, processos estocásticos e dados periódicos no CECP para $D = 6$

no CECP, é possível notar que a maioria das componentes dos mapas caóticos ocupam uma região com elevada complexidade, próximo a curva de complexidade máxima C_{max} , superior a 0.3, tendo como exceção apenas as componentes dos sistemas B_7 e Lorenz, que ocupam uma região em que $\mathcal{H} \sim 0$, além da componente z do sistema Burke que ocupa a região em que a complexidade é de aproximadamente 0.27.

Para ruídos de processo estocástico com espectro de potência $f^{(-k)}$, a localização no CECP exibe um comportamento entrópico decrescente variando de $\mathcal{H} \sim 1$ para $k = 0$ a $\mathcal{H} \sim 0.12$ para $k = 4$. Observe que, para o caso de $k = 0$, ruído não gaussiano não correlacionado, sua localização se aproxima do valor extremo onde o par $(\mathcal{H}, C) = (1, 0)$. Para $k > 0$, os ruídos não gaussianos correlacionados (ruídos coloridos) exibem valores entrópicos mais baixos. Os valores de complexidade estatística associados, que pertencem ao intervalo $0 \leq C \leq 0.39$, situam-se entre os valores de complexidade extrema C_{min} e C_{max} .

Podemos notar também que os pontos que representam o ruído gaussiano e os pontos relativos ao processo estocástico com espectro de potência $f^{(-k)}$, sendo $k = 0$, se sobrepõem na região em que $\mathcal{H} \sim 1$ e $C \sim 0$, indicando um comportamento completamente aleatório.

Por fim, paralelamente, o ruído gaussiano e o sinal periódico ocupam as duas regiões extremas do CECP, onde $\mathcal{H} \sim 1$ e $\mathcal{H} \sim 0$, respectivamente, contrapondo suas dinâmicas. Enquanto o primeiro é completamente aleatório, apresentando efeito de irrelevância (Rosenstein et al., 1993), o segundo tem comportamento previsível, apresentando efeito de redundância (Rosenstein et al., 1993).

4.3 Análise Multivariada com o MvCECP

Como o MvCECP efetua a caracterização de sistemas multivariados, sua análise é feita considerando a correlação entre os dados, englobando o sistema por completo. Na figura 4.3, o sistema

Tabela 4.1: Entropia e Complexidade de cada uma das componentes dos Sistemas Caóticos avaliados no CECP

Sistema Caótico	Var.	Entropia	Complexidade
B_7	x	0.11782	0.11528
	y	0.11811	0.11553
	z	0.10863	0.10664
Burke	x	0.77608	0.35509
	y	0.78208	0.35171
	z	0.86425	0.27395
Complex Butterfly	x	0.71178	0.39561
	y	0.76587	0.37841
	z	0.44079	0.37934
Labyrinth	x	0.52211	0.39991
	y	0.53721	0.40533
	z	0.52246	0.40021
Linear Feedback	x	0.59351	0.45996
	y	0.80028	0.37055
	z	0.69566	0.44251
Lorenz	x	0.12091	0.11805
	y	0.12832	0.12451
	z	0.11972	0.11652
Delayed Logistic	x	0.44986	0.40085
	y	0.44985	0.40084
Henon Map	x	0.55383	0.45789
	y	0.55383	0.45789
Tinkerbell	x	0.60834	0.46103
	y	0.54946	0.44106

B_7 é exibido como exemplo, evidenciando sua disposição no espaço tridimensional.

Para a fase de análises multivariadas, dada a dimensão das séries temporais componentes dos sistemas multivariados, selecionamos a dimensão de incorporação $\sigma = 5$ para garantir que a condição de validade seja atendida, em que $N \geq M$, onde M é o número total de estados acessíveis. A partir da equação 3.17, dado $\sigma = 5$, encontramos $M = 4200$. Neste caso, temos $10^6 \geq 4200$, o que garante estatísticas confiáveis nos estudos. Na figura 4.4, exibimos os resultados encontrados para a avaliação dos sistemas caóticos, estocásticos e periódicos supracitados, utilizando o MvCECP, sendo cada um deles representado por uma cor diferente.

Na tabela 4.2 são exibidos os valores de entropia e complexidade para cada um dos sistemas caóticos avaliados.

Com relação a visualização de dados no MvCECP, podemos afirmar que existe uma separação visível entre os pontos mais bem definida que na perspectiva univariada utilizando o CECP.

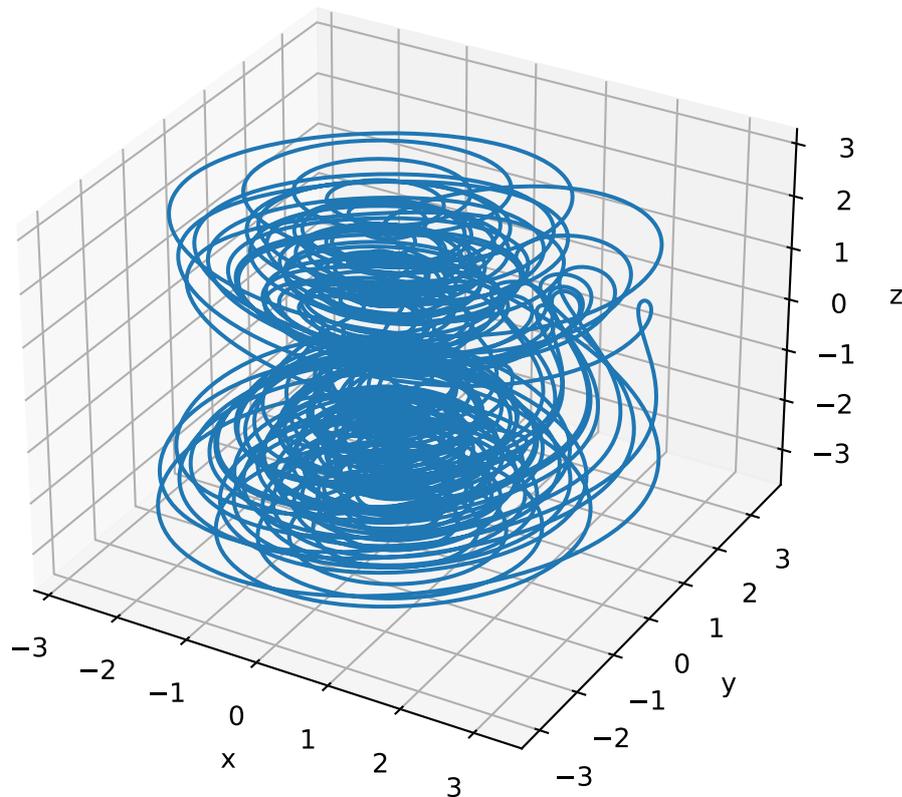


Figura 4.3: Representação Gráfica do Sistema Multivariado B_7 no espaço tridimensional

Tabela 4.2: Entropia e Complexidade dos Sistemas Caóticos avaliados no MvCECP

Sistema Caótico	Entropia	Complexidade
B_7	0.11448	0.11259
Burke	0.88361	0.28759
Complex Butterfly	0.69276	0.47124
Labyrinth	0.59796	0.46431
Linear Feedback	0.68314	0.50199
Lorenz	0.13042	0.12782
Delayed Logistic	0.40529	0.38712
Henon Map	0.63211	0.52068
Tinkerbell	0.50557	0.44849

Assim, utilizando essa abordagem, obtivemos resultados superiores na distinção entre os sistemas caóticos estudados e, conseqüentemente, a partir do conhecimento de propriedades e da dinâmica em cada região do plano, é possível efetuar uma caracterização mais assertiva desses sistemas.

Assim como na abordagem univariada, a maior parte dos sistemas caóticos ocupa uma região com elevada complexidade, superior a 0.3, sendo próxima a curva de complexidade má-

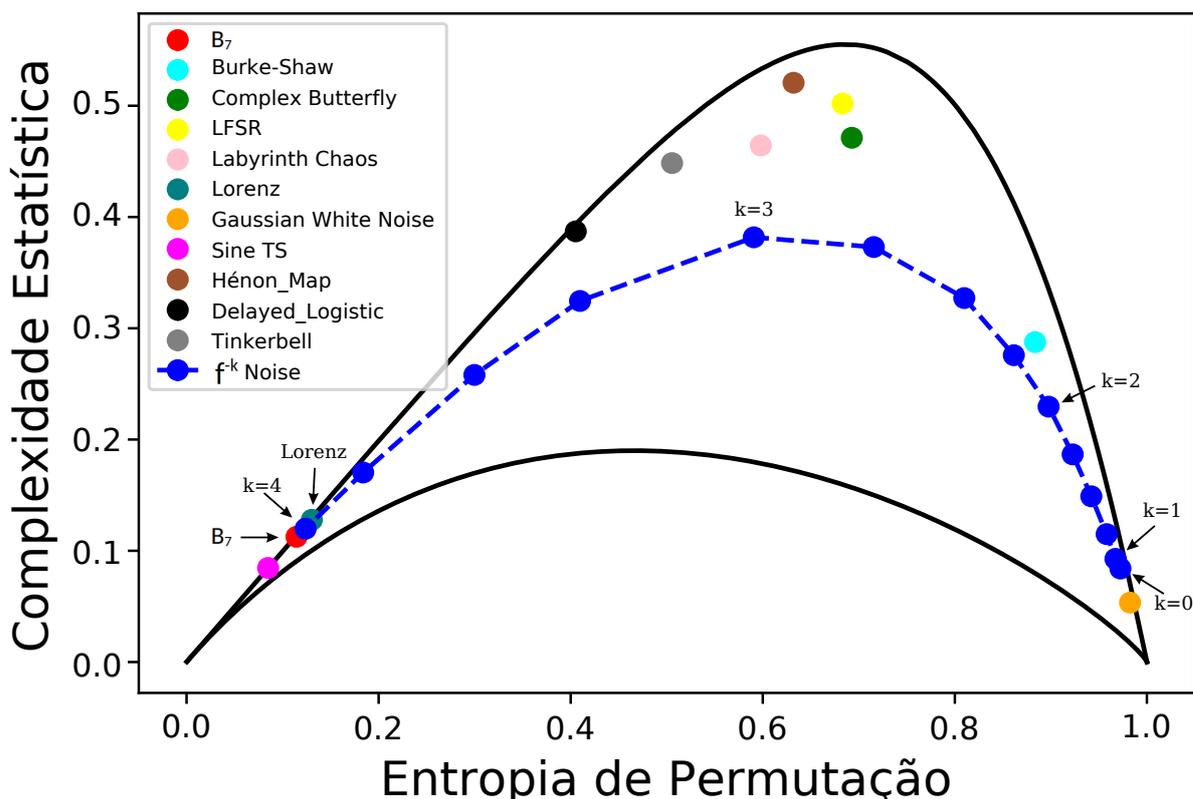


Figura 4.4: Avaliação no dos sistemas caóticos, processos estocásticos e dados periódicos no MvCECP para $\sigma = 5$

xima C_{max} , que neste caso tem limite superior ao caso univariado, chegando a alcançar valores próximos a 0.6. Isso ocorre porque os valores da curva C_{max} dependem do número de estados acessíveis pela distribuição de probabilidade. Como $M > D!$, os valores máximos de complexidade para cada par de entropia são superiores no caso multivariado.

As exceções, assim como no caso univariado, são os sistemas B_7 e Lorenz que ocupam uma região em $\mathcal{H} \sim 0$, além de Burke, que ocupa uma região em que $\mathcal{C} \sim 0.27$, assim como sua componente z na abordagem univariada, o que revela uma maior influência dessa componente sobre as demais na dinâmica do sistema multivariado.

Nos casos de sistemas em que suas componentes ocupavam regiões muito próximas, sejam eles tridimensionais, como ocorre com B_7 , Labyrinth e Lorenz, ou bidimensionais, como são os casos de Delayed Logistic e Hénon Map, é possível notar que a representação multivariada ocupa uma região similar no MvCECP.

Para ruídos de processo estocástico com espectro de potência $f^{(-k)}$, como as 3 componentes de cada ruído ocupam regiões extremamente próximas, os sistemas multivariados apresentam resultados muito semelhantes ao caso univariado. A localização desses ruídos no MvCECP também exibe um comportamento entrópico decrescente variando de $\mathcal{H} \sim 1$ para $k = 0$ a $\mathcal{H} \sim 0.12$ para $k = 4$. Os valores de complexidade estatística associados, que pertencem ao intervalo $0 \leq C \leq 0.39$, situam-se entre os valores de complexidade extrema C_{min} e C_{max} .

Por fim, também de modo similar ao caso univariado, o ruído gaussiano e o sinal periódico ocupam as duas regiões extremas do CECP, onde $\mathcal{H} \sim 1$ e $\mathcal{H} \sim 0$, respectivamente, contrapondo suas dinâmicas. Isso acontece porque as 3 componentes de cada um deles ocupam regiões extremamente próximas, então os sistemas multivariados apresentam resultados muito semelhantes a abordagem univariada.

Podemos notar ainda que o ruído gaussiano ocupa a região mais próxima ao local onde $\mathcal{H} \sim 1$ e $\mathcal{C} \sim 0$, indicando ter o comportamento com dinâmica mais próxima da completa aleatoriedade dentre todos os sistemas analisados.

Com isso, podemos concluir que o MvCECP cumpre bem seu objetivo de conseguir distinguir entre os comportamentos caótico e estocástico, além do periódico, ofertando ganhos consideráveis na distinção e caracterização da dinâmica de sistemas caóticos multivariados.

5

Considerações Finais

5.1 Conclusões

Neste trabalho, investigamos a aplicação do Plano Causal de Complexidade-Entropia (CECP) na tarefa de distinguir e caracterizar sistemas multivariados de diversas naturezas, visto que sua avaliação é feita somente sobre séries temporais univariadas, enquanto a grande maioria dos sistemas existentes é multivariada. Dessa forma, as quantificações realizadas com o plano $\mathcal{H} \times \mathcal{C}$ univariado, são feitas percorrendo com janelas deslizantes cada uma das séries temporais componentes do sistema, o que implica que a correlação existente entre elas não é avaliada.

Além disso, nessa abordagem, a construção das Funções de Densidade de Probabilidade (PDF) é baseada na criação de partições com o número total de estados acessíveis. Cada símbolo é determinado a partir dos padrões ordinais apresentados por valores relativos vizinhos na série temporal avaliada por uma janela deslizante de dimensão D . Entretanto, por só considerar os padrões ordinais intrínsecos as séries temporais, essa abordagem acarreta na perda de alguns detalhes de informações de amplitude da série original.

Diante dessas duas potenciais melhorias a serem realizadas na avaliação feita pelo CECP, propusemos o Plano Causal Multivariado de Complexidade-Entropia (MvCECP). Nessa abordagem, efetuamos uma adaptação na construção da PDF para que as séries temporais sejam avaliadas conjuntamente, de modo a garantir que a correlação existente entre elas seja considerada nos cálculos de Entropia e Complexidade. Como o R^n não é um corpo ordenável para $n > 1$, utilizamos as normas dos pontos formados pela concatenação das componentes de cada sistema como métrica para estudo dos padrões ordinais. Assim, é possível ordenar os pontos com relação a uma referência estabelecida, que no nosso caso é a origem do sistema, o que só é possível porque todas as séries temporais componentes são normalizadas no início do processo.

Além disso, desenvolvemos uma técnica para divisão do espaço n-dimensional em σ regiões com mesmo hipervolume, onde σ é a dimensão de incorporação adotada. Assim, é possível

determinar as diferenças de amplitude existentes entre os elementos avaliados, mais uma vez utilizando as informações de norma dos pontos formados pelas séries temporais componentes dos sistemas em estudo.

A fim de comparar as abordagens e avaliar a eficácia do MvCECP, utilizamos 3 conjuntos de dados distintos: um composto por sistemas caóticos, contando com modelos bidimensionais e tridimensionais; um conjunto de sistemas estocásticos, contando com dois tipos distintos de ruídos; e um último conjunto composto por dados de um sistema periódico produzido a partir da função seno. Tanto o conjunto de dados estocásticos quanto o de dados periódicos foram modelados para serem tridimensionais, de modo a assegurar uma avaliação adequada na abordagem multidimensional.

O CECP foi inicialmente proposto por com o intuito de prover um mecanismo adequado para diferenciar sistemas caóticos e processos estocásticos e, posteriormente, foi utilizado na caracterização de séries temporais decorrentes de mapas caóticos. Ao realizar os experimentos com ambas as abordagens, univariada e multivariada, utilizando os conjuntos de dados mencionados, constatamos que o MvCECP, assim como o CECP é eficaz na distinção entre os comportamentos caótico, estocástico e periódico.

Entretanto, o MvCECP demonstrou ser mais eficaz que o CECP na distinção entre sistemas caóticos, sendo portanto, mais adequado para efetuar uma caracterização de sistemas multidimensionais com dinâmica caótica.

5.2 Trabalhos Futuros

Com o desenvolvimento do MvCECP e a comprovação de sua eficácia em caracterizar a dinâmica de sistemas caóticos, além de ser capaz de distinguir sistemas caóticos e estocásticos, pretendemos efetuar um estudo de caracterização de mapas caóticos multivariados, distinguindo-os entre suas categorias: mapas não-inversíveis, mapas dissipativos e mapas conservativos. Além disso, a partir da posição apresentada pelos conjuntos de dados no plano, poderemos inferir propriedades comuns a sistemas que integrem a mesma categoria.

Outra possibilidade é avaliar o impacto que a amostragem dos dados de um determinado sistema pode exercer sobre sua caracterização no plano. Nesse estudo, é possível investigar quais são os intervalos de amostragem ótima de sistemas multivariados com diversas aplicações práticas. A amostragem ótima é encontrada quando um determinado subconjunto de elementos de um dado sistema para $\tau > 1$, atinge a maior complexidade possível para uma amostra. Esse estudo tem um impacto prático em diversos campos, especialmente quando aplicado na coleta, tráfego e armazenamento de dados em uma rede de sensores sem fio (RSSF).

Além disso, também é importante investigar o impacto que a presença de ruídos adicionados a um conjunto de dados de um determinado sistema pode exercer sobre sua caracterização no MvCECP, sendo importante averiguar até onde os conjuntos suportam preservar sua invari-

abilidade ao acréscimo de ruídos, propriedade essencial para detecção de eventos anômalos em sistemas de monitoramento de dados.

Por fim, o estudo também pode ser estendido para a análise de diversos conjuntos de dados multidimensionais. Em especial, desejamos investigar a efetividade do método na caracterização ou clusterização de imagens coloridas, utilizando seus canais de cor como referências para abordagem adotada com séries temporais. Todavia, o leque de possibilidades a serem exploradas é imenso, passando também por estudos sobre som, buscando distinguir entre estilos musicais, mercado de ações, dentre muitos outros campos de aplicação.

Referências bibliográficas

- Mosabber Uddin Ahmed and Danilo P Mandic. Multivariate multiscale entropy: A tool for complexity analysis of multichannel data. *Physical Review E*, 84(6):061918, 2011.
- Christoph Bandt and Bernd Pompe. Permutation entropy: a natural complexity measure for time series. *Physical review letters*, 88(17):174102, 2002.
- João B Borges, Heitor S Ramos, Raquel AF Mini, Osvaldo A Rosso, Alejandro C Frery, and Antonio AF Loureiro. Learning and distinguishing time series dynamics via ordinal patterns transition graphs. *Applied Mathematics and Computation*, 362:124554, 2019.
- Eduarda TC Chagas, Alejandro C Frery, Juliana Gambini, Magdalena M Lucini, Heitor S Ramos, and Andrea A Rey. Statistical properties of the entropy from ordinal patterns. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(11):113118, 2022.
- Konstantinos E Chlouverakis and JC Sprott. A comparison of correlation and lyapunov dimensions. *Physica D: Nonlinear Phenomena*, 200(1-2):156–164, 2005.
- Kevin M Cuomo, Alan V Oppenheim, and Steven H Strogatz. Synchronization of lorenz-based chaotic circuits with applications to communications. *IEEE Transactions on circuits and systems II: Analog and digital signal processing*, 40(10):626–633, 1993.
- Per-Erik Danielsson. Euclidean distance mapping. *Computer Graphics and image processing*, 14(3):227–248, 1980.
- Luciana De Micco, Juana Graciela Fernández, Hilda A Larrondo, Angelo Plastino, and Osvaldo A Rosso. Sampling period, statistical complexity, and chaotic attractors. *Physica A: Statistical Mechanics and its Applications*, 391(8):2564–2575, 2012.
- Jean-Louis Deneubourg and Simon Goss. Collective patterns and decision-making. *Ethology Ecology & Evolution*, 1(4):295–311, 1989.
- Leandro dos Santos Coelho and Viviana Cocco Mariani. Firefly algorithm approach based on chaotic tinkerbell map applied to multivariable pid controller tuning. *Computers & Mathematics with Applications*, 64(8):2371–2382, 2012.

- Ahmed S Elwakil, S Ozoguz, and Michael Peter Kennedy. Creation of a complex butterfly attractor using a novel lorenz-type system. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 49(4):527–530, 2002.
- David P Feldman, Carl S McTague, and James P Crutchfield. The organization of intrinsic computation: Complexity-entropy diagrams and the diversity of natural information processing. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 18(4):043106, 2008.
- Bent Fuglede and Flemming Topsøe. Jensen-shannon divergence and hilbert space embedding. In *International symposium on Information theory, 2004. ISIT 2004. Proceedings.*, page 31. IEEE, 2004.
- Jake Gipple. The volume of n-balls. *Rose-Hulman Undergraduate Mathematics Journal*, 15(1):14, 2014.
- Hermann Haken. Analogy between higher instabilities in fluids and lasers. *Physics Letters A*, 53(1):77–78, 1975.
- Shaobo He, Kehui Sun, and Huihai Wang. Multivariate permutation entropy and its application for complexity analysis of chaotic systems. *Physica A: Statistical Mechanics and its Applications*, 461:812–823, 2016.
- Neyram Hemati. Strange attractors in brushless dc motors. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 41(1):40–45, 1994.
- Robert C Hilborn et al. *Chaos and nonlinear dynamics: an introduction for scientists and engineers*. Oxford University Press on Demand, 2000.
- Chieh Su Hsu. *Cell-to-cell mapping: a method of global analysis for nonlinear systems*, volume 64. Springer Science & Business Media, 2013.
- F Kaffashi, MS Scher, SM Ludington-Hoe, and KA Loparo. An analysis of the kangaroo care intervention using neonatal eeg complexity: a preliminary study. *Clinical neurophysiology*, 124(2):238–246, 2013.
- Stuart A Kauffman et al. *The origins of order: Self-organization and selection in evolution*. Oxford University Press, USA, 1993.
- Karsten Keller and Mathieu Sinn. Ordinal analysis of time series. *Physica A: Statistical Mechanics and its Applications*, 356(1):114–120, 2005.
- PW Lamberti, MT Martin, A Plastino, and OA Rosso. Intensive entropic non-triviality measure. *Physica A: Statistical Mechanics and its Applications*, 334(1-2):119–131, 2004.

- Jacques Laskar. A numerical experiment on the chaotic behaviour of the solar system. *Nature*, 338(6212):237–238, 1989.
- Christophe Letellier, P Dutertre, J Reizner, and G Gouesbet. Evolution of a multimodal map induced by an equivariant vector field. *Journal of Physics A: Mathematical and General*, 29(17):5359, 1996.
- Christophe Letellier, Tsvetelin D Tsankov, Greg Byrne, and Robert Gilmore. Large-scale structural reorganization of strange attractors. *Physical Review E*, 72(2):026212, 2005.
- Jing Li, Jiaqing Yan, Xianzeng Liu, and Gaoxiang Ouyang. Using permutation entropy to measure the changes in eeg signals during absence seizures. *Entropy*, 16(6):3049–3061, 2014.
- Peng Li, Cheng-Yu Liu, Li-Ping Li, Li-Zhen Ji, Shou-Yuan Yu, and Chang-Chun Liu. Multiscale multivariate fuzzy entropy analysis. 2013.
- Ricardo Lopez-Ruiz, Héctor L Mancini, and Xavier Calbet. A statistical measure of complexity. *Physics letters A*, 209(5-6):321–326, 1995.
- Edward N Lorenz. Deterministic nonperiodic flow. *Journal of atmospheric sciences*, 20(2):130–141, 1963.
- René Lozi. Un attracteur étrange (?) du type attracteur de hénou. *Le Journal de Physique Colloques*, 39(C5):C5–9, 1978.
- Fang Ma, Qingju Fan, and Guang Ling. Complexity-entropy causality plane analysis of air pollution series. *Fluctuation and Noise Letters*, 21(02):2250011, 2022.
- MT Martin, A Plastino, and OA Rosso. Generalized statistical complexity measures: Geometrical and analytical properties. *Physica A: Statistical Mechanics and its Applications*, 369(2):439–462, 2006.
- James Massey. Shift-register synthesis and bch decoding. *IEEE transactions on Information Theory*, 15(1):122–127, 1969.
- Augusto César MORGADO, João Bosco Pitombeira de CARVALHO, Paulo Cezar Pinto CARVALHO, and Pedro FERNANDEZ. Análise combinatória e probabilidade: com as soluções dos exercícios. *Coleção do Professor de Matemática*, ed, 9, 2006.
- Sayan Mukherjee, Sanjay Kumar Palit, Santo Banerjee, MRK Ariffin, Lamberto Rondoni, and DK Bhattacharya. Can complexity decrease in congestive heart failure? *Physica A: Statistical Mechanics and its Applications*, 439:93–102, 2015.

- Givanildo L Nascimento, Cristopher GS Freitas, Osvaldo A Rosso, and Andre LL Aquino. Data sampling algorithm based on complexity-entropy plane for smart sensing applications. *IEEE Sensors Journal*, 21(22):25831–25842, 2021.
- SGOPAL Patro and Kishore Kumar Sahu. Normalization: A preprocessing stage. *arXiv preprint arXiv:1503.06462*, 2015.
- Steven M Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, 88(6):2297–2301, 1991.
- AR Plastino and A Plastino. Symmetries of the fokker-planck equation and the fisher-frieden arrow of time. *Physical Review E*, 54(4):4423, 1996.
- William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. Numerical recipes in c. 2. *Cambridge University*, 1992.
- Steen Rasmussen, Carsten Knudsen, Rasmus Feldberg, and Morten Hindsholm. The coreworld: Emergence and evolution of cooperative structures in a computational chemistry. *Physica D: Nonlinear Phenomena*, 42(1-3):111–134, 1990.
- Haroldo V Ribeiro, Luciano Zunino, Renio S Mendes, and Ervin K Lenzi. Complexity–entropy causality plane: A useful approach for distinguishing songs. *Physica A: Statistical Mechanics and its Applications*, 391(7):2421–2428, 2012.
- Michael T Rosenstein, James J Collins, and Carlo J De Luca. A practical method for calculating largest lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena*, 65(1-2):117–134, 1993.
- OA Rosso, L Zunino, DARIO GABRIEL Pérez, Alejandra Figliola, Hilda A Larrondo, M Garavaglia, María Teresa Martín, and A Plastino. Extracting features of gaussian self-similar stochastic processes via the bandt-pompe approach. *Physical Review E*, 76(6):061114, 2007a.
- Osvaldo A Rosso and Cristina Masoller. Detecting and quantifying stochastic and coherence resonances via information-theory complexity measurements. *Physical Review E*, 79(4):040106, 2009.
- Osvaldo A Rosso, HA Larrondo, María Teresa Martin, A Plastino, and Miguel A Fuentes. Distinguishing noise from chaos. *Physical review letters*, 99(15):154102, 2007b.
- Osvaldo A Rosso, Luciana De Micco, Hilda A Larrondo, María T Martín, and Angel Plastino. Generalized statistical complexity measure. *International Journal of Bifurcation and Chaos*, 20(03):775–785, 2010.

- Oswaldo A Rosso, Felipe Olivares, Luciano Zunino, Luciana De Micco, André LL Aquino, Angelo Plastino, and Hilda A Larrondo. Characterization of chaotic maps using the permutation bandt-pompe probability distribution. *The European Physical Journal B*, 86(4): 1–13, 2013.
- Jacob Savir and William H McAnney. A multiple seed linear feedback shift register. In *Proceedings. International Test Conference 1990*, pages 657–659. IEEE, 1990.
- Claude Elwood Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27: 623–656, 1948.
- Robert Shaw. Strange attractors, chaotic behavior, and information flow. *Zeitschrift für Naturforschung A*, 36(1):80–112, 1981.
- Luiz Eduardo Virgilio Silva, Brenno Caetano Troca Cabella, Ubiraci Pereira da Costa Neves, and Luiz Otavio Murta Junior. Multiscale entropy-based methods for heart rate variability complexity analysis. *Physica A: Statistical Mechanics and its Applications*, 422:143–152, 2015.
- Miguel C Soriano, Luciano Zunino, Laurent Larger, Ingo Fischer, and Claudio R Mirasso. Distinguishing fingerprints of hyperchaotic and stochastic dynamics in optical chaos from a delayed opto-electronic oscillator. *Optics letters*, 36(12):2212–2214, 2011.
- Julien Clinton Sprott and Julien C Sprott. *Chaos and time-series analysis*, volume 69. Oxford university press Oxford, 2003.
- René Thomas. Deterministic chaos seen in terms of feedback circuits: Analysis, synthesis, "labyrinth chaos". *International Journal of Bifurcation and Chaos*, 9(10): 1889–1905, 1999.
- Valentina A Unakafova, Anton M Unakafov, and Karsten Keller. An approach to comparing kolmogorov-sinai and permutation entropy. *The European Physical Journal Special Topics*, 222(2):353–361, 2013.
- Guo-Cheng Wu and Dumitru Baleanu. Discrete chaos in fractional delayed logistic maps. *Nonlinear Dynamics*, 80:1697–1703, 2015.
- Yi Yin and Pengjian Shang. Weighted multiscale permutation entropy of financial time series. *Nonlinear Dynamics*, 78(4):2921–2939, 2014.
- Massimiliano Zanin, Luciano Zunino, Oswaldo A Rosso, and David Papo. Permutation entropy and its main biomedical and econophysics applications: A review. *Entropy*, 14(8): 1553–1577, 2012.

- Yali Zhang and Pengjian Shang. Permutation entropy analysis of financial time series based on hill's diversity number. *Communications in Nonlinear Science and Numerical Simulation*, 53:288–298, 2017.
- Luciano Zunino, Miguel C Soriano, Ingo Fischer, Osvaldo A Rosso, and Claudio R Mirasso. Permutation-information-theory approach to unveil delay dynamics from time-series analysis. *Physical Review E*, 82(4):046212, 2010a.
- Luciano Zunino, Massimiliano Zanin, Benjamin M Tabak, Darío G Pérez, and Osvaldo A Rosso. Complexity-entropy causality plane: A useful approach to quantify the stock market inefficiency. *Physica A: Statistical Mechanics and its Applications*, 389(9):1891–1901, 2010b.
- Luciano Zunino, Benjamin M Tabak, Francesco Serinaldi, Massimiliano Zanin, Darío G Pérez, and Osvaldo A Rosso. Commodity predictability analysis with a permutation information theory approach. *Physica A: Statistical Mechanics and its Applications*, 390(5):876–890, 2011.
- Luciano Zunino, Aurelio Fernández Bariviera, M Belén Guercio, Lisana B Martinez, and Osvaldo A Rosso. On the efficiency of sovereign bond markets. *Physica A: Statistical Mechanics and its Applications*, 391(18):4342–4349, 2012.