UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

JOÃO VITOR GONÇALVES DA SILVA
Orientador: Vander Luis de Souza Freitas

# INVESTIGANDO MÉTRICAS E ALGORITMOS PARA COMPARAÇÃO DE SÉRIES TEMPORAIS

Ouro Preto, MG
2023

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

JOÃO VITOR GONÇALVES DA SILVA

# INVESTIGANDO MÉTRICAS E ALGORITMOS PARA COMPARAÇÃO DE SÉRIES TEMPORAIS

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

**Orientador:** Vander Luis de Souza Freitas

Ouro Preto, MG
2023

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
INSTITUTO DE CIENCIAS EXATAS E BIOLOGICAS
DEPARTAMENTO DE COMPUTACAO

# FOLHA DE APROVAÇÃO

**João Vitor Gonçalves da Silva**

**Investigando métricas e algoritmos para comparação de séries temporais**

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Ciência da Computação

Aprovada em 9 de Outubro de 2024.

Membros da banca

Vander Luis de Souza Freitas (Orientador) - Doutor - Universidade Federal de Ouro Preto
Guilherme Augusto Lopes Silva (Examinador) - Mestre - Universidade Federal de Ouro Preto
Aurelienne Aparecida Souza Jorge (Examinadora) - Mestre - Instituto Nacional de Pesquisas Espaciais

Vander Luis de Souza Freitas, Orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 9/10/2024.





R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35402-163
Telefone: 3135591692 - www.ufop.br

*Dedico esse trabalho aos meus pais, Roseni e Joel, e minhas irmãs Josiane e Rosiane, pelos ensinamentos, pelo apoio incondicional e por me darem força de continuar ao longo desta jornada.*

# Agradecimentos

Agradeço aos meus pais e minhas irmãs, pelo suporte que me forneceram para manter minha vida acadêmica. As minhas amigas Júlia, Myllene e Sabrina por me acompanharem por toda a jornada e por todo apoio fornecido. Agradeço ao professor Vander Freitas e ao CSILab pela orientação durante a monografia, à Universidade Federal de Ouro Preto e ao Departamento de Computação.

“O simples bater de asas de uma borboleta no Brasil pode ocasionar um tornado no Texas”

Edward Lorenz

# Resumo

Séries temporais são medidas de dados feitas ao longo de um período de tempo. A comparação entre séries é uma prática útil quando deseja-se identificar diferenças ou similaridades em alguns aspectos. Uma aplicação derivada dessas comparações é a geração de redes funcionais, caracterizada pela criação de conexões entre pontos de dados similares, os quais podem ser séries temporais. Estas redes são comumente geradas a partir de dados meteorológicos, sendo úteis na detecção de teleconexões entre eventos, como chuvas e mudanças abruptas de temperatura da superfície terrestre. Dado essa necessidade, saber exatamente quais métricas ou algoritmos utilizar é de suma importância, já que existem vários tipos de séries temporais e nem todo método gera resultados ideais, no contexto da aplicação de interesse. Essa monografia propõe gerar séries sintéticas com diversas propriedades como séries temporais periódicas, caóticas, com ruídos e combinações entre elas a fim de avaliar quais métodos melhor capturam as diferenças entre os pares de séries. Os algoritmos e métricas analisados foram o *Dynamic Time Warping*, Coeficiente de Correlação de Pearson e Informação Mútua, que são métricas mais utilizadas na literatura. Os resultados são similares ao se comparar séries derivadas de combinações entre diferentes tipos de dinâmica (caótica e periódica), mas DTW, em específico, foi a única abordagem a detectar similaridades entre séries caóticas, e mateve uma relação linear com o aumento de ruído para comparação de séries e suas versões ruidosas.

**Palavras-chave**: Séries temporais. Comparação de séries temporais. Redes funcionais.

# Abstract

Time series are measurements of data taken over a period of time. Comparing time series is a useful practice when seeking to identify differences or similarities in certain aspects. A derived application of these comparisons is the generation of functional networks, characterized by creating connections between similar data points, which can be time series. These networks are commonly generated from meteorological data and are useful in detecting teleconnections between events, such as rainfall and abrupt changes in the surface temperature of the Earth. Given this need, knowing exactly which metrics or algorithms to use is of utmost importance, as there are various types of time series, and not every method produces ideal results in the context of the application of interest. This study proposes to generate synthetic time series with diverse properties, such as periodic time series, chaotic time series, with noise, and combinations thereof, in order to evaluate which methods better capture the differences between pairs of time series. The algorithms and metrics analyzed were Dynamic Time Warping, Pearson Correlation Coefficient, and Mutual Information. The results are similar when comparing series derived from combinations of different types of dynamics (chaotic and periodic), but DTW, specifically, was the only approach to detect similarities between chaotic series. Additionally, DTW maintained a linear relationship with the increase in noise when comparing series and their noisy versions.

**Keywords**: Time series. Time series comparison. Functional networks.

# Lista de Ilustrações

# Lista de Abreviaturas e Siglas

| | |
|---|---|
| ML | Mapa Logístico |
| IM | Informação Mútua |
| DTW | *Dynamic Time Warping* |
| CCP | Coeficiente de Correlação de Pearson |

# Lista de Símbolos

| | |
|---|---|
| $x_n$ | Valor da variável na iteração $n$ do Mapa Logístico |
| $r$ | Parâmetro do Mapa Logístico |
| $\xi(t)$ | Fonte de ruído |
| $\delta(t)$ | Intensidade do ruído |
| $P_{(X,Y)}$ | Função de massa de probabilidade conjunta |
| $P_X$ | Distribuição marginal da série temporal $X$ |

# Sumário

# 1 Introdução

Séries temporais são medições de valores de um fenômeno ao longo de um período. Sendo utilizadas em diversas áreas, como área financeira e previsão do tempo, essas mensurações têm o propósito de avaliar informações essenciais sobre as situações em questão, tais como a identificação de tendências e anomalias (Brockwell; Davis, 2002).

Na área da medicina, é notável que algumas doenças possuem tendência de ocorrer com maior frequência a depender da época do ano. Caso catalogado esses dados, é possível compreender os padrões de doença ao longo dos anos e aplicar ações em épocas específicas para evitar mortes nas épocas de maior ocorrência (Antunes; Cardoso, 2015). Na área financeira, o mercado de ações se mostra em constante mudança, na qual séries temporais se mostram úteis em catalogar essas mudanças e detectar padrões, sendo possível prever próximos valores, auxiliando na tomada de decisão de investimentos (Arroyo; Espínola; Maté, 2011).

Em alguns outros aspectos científicos, além da analise de dados, é preciso verificar similaridades entre os dados para agrupá-los. Essa verificação é realizada for funções de cálculo de distância, que verificam o quanto duas séries são relacionadas (Wang *et al.*, 2013).

Um desafio é saber qual métrica ou função de distância utilizar. Em Lhermitte *et al.* (2011) são utilizados três conjuntos de dados sobre o ecossistema em formato de séries temporais para determinar quais medidas de similaridades são mais eficazes em cada caso, para deteccção de mudanças no meio ambiente. O artigo também aponta a importância da escolha do método correto para a situação que deseja-se obter resultados, já que cada método pode ser eficaz em casos especificos. Em Ferreira *et al.* (2021), várias técnicas são utilizadas para a mesma base de dados com o intuito de gerar redes funcionais. No artigo, o *Dynamic Time Warping* se mostrou eficaz na geração de redes com conexões de longo alcance entre fenômenos atmosféricos, enquanto outros métodos se mostram eficazes na detecção de similaridades entre séries temporais de fenômenos mais próximos.

## 1.1 Justificativa e motivação

Com a análise de séries temporais, tem-se a necessidade de comparação entre esses dados para verificar similaridades e diferenças entre os conjuntos. Entretanto, há vários tipos de séries e não se sabe ao certo qual método escolher para comparar quaisquer tipos de séries temporais. Como por exemplo, o Coeficiente de Correlação de Pearson leva em conta apenas as relações lineares entre as séries, negligenciando relações mais complexas. Então, existe a necessidade de catalogar os métodos e verificar em quais casos eles são mais propícios na comparação.

Uma motivação prática é a possibilidade de criar redes funcionais a partir dos conjuntos

de séries temporais (Ferreira *et al.*, 2021; Jorge *et al.*, 2020; Donges *et al.*, 2009; Diniz, 2022), já que um dos problemas atuais é justamente a escolha da métrica ou algoritmo para comparação de séries. Espera-se definir quais deles melhor capturam as diferenças entre séries de diferentes tipos, como periódicas, ruidosas, caóticas e combinações das anteriores.

## 1.2 Objetivo geral e objetivos específicos

O objetivo geral desta monografia é avaliar quais métricas e algoritmos melhor capturam semelhanças entre séries temporais de diferentes tipos. São utilizados *Dynamic Time Warping*, Informação Mútua e Coeficiente de Correlação de Pearson como métricas e algoritmos para o estudo da comparação das séries, que são métricas mais populares na literatura. As séries comparadas serão séries ruidosas, periódicas, caóticas e combinações das anteriores, como estudos de caso. Os objetivos específicos são:

- Gerar séries temporais sintéticas a partir de funções periódicas, ruídos aditivos e do Mapa Logístico;
- Comparar conjuntos de séries com diferentes características, a partir do *Dynamic Time Warping*, Informação Mútua e Coeficiente de Correlação de Pearson;
- Comparar séries temporais obtidas de *datasets* do Kaggle para ter análises dos algorítmos e métricas para dados reais;
- Analisar graficamente e quantitativamente, dentre as três técnicas de comparação de séries, quais melhor capturam diferenças entre elas, nos diferentes contextos.

## 1.3 Organização do Trabalho

A monografia está organizada da seguinte forma. No Capítulo 2 está contido todo o embasamento teórico necessário para a compreensão do trabalho. O Capítulo 3 descreve os passos metodológicos para atingir os objetivos da pesquisa, seguido do Capítulo 4, com os resultados obtidos e o Capítulo 5, com as considerações finais e trabalhos futuros.

### 1.3.1 Estrutura da Monografia

**Capítulo 1:** Introdução.

**Capítulo 2:** Revisão Bibliográfica.

**Capítulo 3:** Desenvolvimento.

**Capítulo 4:** Resultados e Discussões.

**Capítulo 5:** Conclusão e Trabalhos Futuros.

# 2 Revisão Bibliográfica

Este capítulo contém uma discussão da fundamentação teórica e dos trabalhos relacionados, para entendimento da temática do projeto.

## 2.1 Fundamentação Teórica

### 2.1.1 Geração de Séries Temporais

Esta seção introduz os métodos que são utilizados para geração de séries temporais na Monografia. O Mapa Logístico foi popularizado por May (1976), considerado por ele um sistema simples, mas com comportamento complexo. A partir de um ponto inicial, é possível gerar uma sequência de pontos, cada um dependendo do ponto anterior e de um parâmetro definido a priori. Esse Mapa apresenta regimes periódicos de diferentes períodos e caóticos, a depender do parâmetro escolhido.

#### 2.1.1.1 Mapa Logístico

Como mostrado em May (1976), modelos matemáticos simples podem ser utilizados para estudar e entender modelos matemáticos complexos. Um exemplo disso é o Mapa Logístico que gera pontos a partir de uma condição inicial e um único parâmetro, sendo possível obter trajetórias periódicas e caóticas a depender da escolha do parâmetro. Mesmo sendo uma função simples, serve para estudos de população e sua taxa de crescimento, resultando em uma analise detalhada já que leva em consideração as condições de restrição do ambiente, levando a uma previsão de população com uma maior precisão. Sua formulação é a seguinte:

$$x_{n+1} = rx_n(1 - x_n) + \xi(n), \tag{2.1}$$

onde $x_n$ é um valor entre 0 e 1, que representa o valor da variável na iteração $n$ e $r$ é o único parâmetro do modelo, que pode representar tanto taxa de crescimento quanto de decaimento. Para a realização dos experimentos, também é adicionada uma fonte de ruído $\xi(n) \in [-\delta, \delta]$, com distribuição uniforme e intensidade $\delta$. Quando $\xi = 0$, a Equação (2.1) se resume ao Mapa Logístico tradicional, apresentado por May (1976).

Para uma representação gráfica, a Figura 2.1 apresenta o diagrama de bifurcação do Mapa Logístico, sem ruídos, que evidencia o comportamento da função de acordo com a variação de $r$. O eixo das abcissas representa o parâmetro $r$ e nas ordenadas estão os valores que $x$ assume após o modelo chegar em seu regime final. Isto é, independentemente da condição inicial $x_0$, após um número alto de iterações do modelo, ele se estabiliza e apresenta órbitas de diferentes períodos:

sendo período 1, por exemplo, quando $r = 2.9$, o mesmo valor se repetindo constantemente ao longo das iterações; período 2, por exemplo, quando $r = 3.2$, onde dois valores se alternam continuamente em pontos consecutivos; e em regimes caóticos, onde os valores nunca se repetem em nenhum intervalo da função, indefinidamente.

É possível observar na Figura 2.2 que quando o valor é $r = 3.567$, o Mapa Logístico gera uma série temporal periódica, mas para $r = 3.956$, o comportamento é caótico.

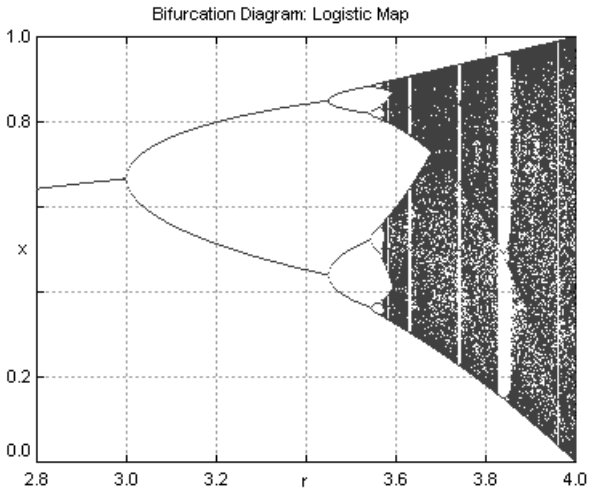



Figura 2.1 – Diagrama de Bifurcação do Mapa Logístico gerado pelo programa Chaos for Java. Fonte: (Davies, 2003)

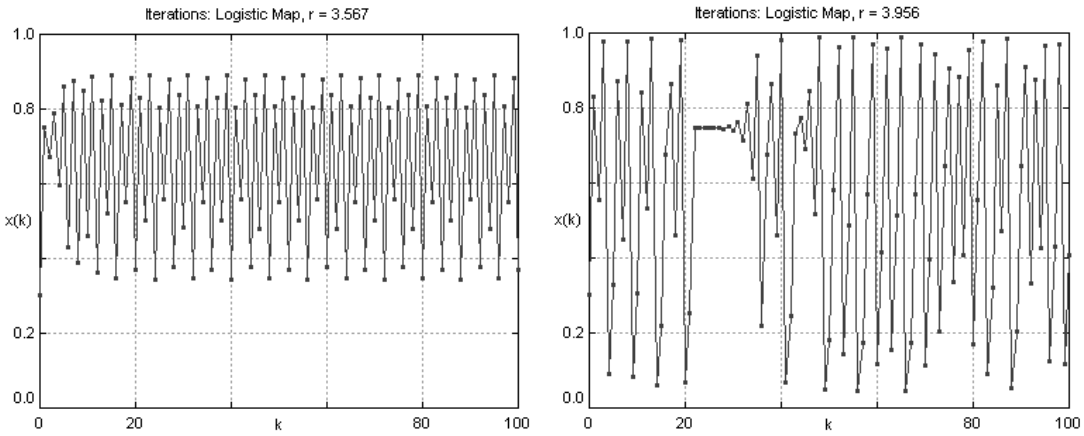



Figura 2.2 – Séries temporais geradas pelo Mapa Logístico com $r = 3.567$, à esquerda, e $r = 3.956$ à direita, pelo programa Chaos for Java.

#### 2.1.1.2 Séries geradas via uma função senoidal

Esta seção apresenta outra forma de geração de séries sintéticas, a partir de uma função periódica com adição de ruído. Trata-se da função seno com adição de uma fonte de ruído com distribuição uniforme:

$$f(t) = \frac{sen(t) + 5}{10} + \xi(t), \tag{2.2}$$

sendo $t$ o ponto no tempo e $\xi(t)$ a fonte de ruído, o qual é gerado com uma intensidade $\delta$. Cada série gerada será diferente, aumentando a discrepância à medida que $\delta$ aumenta. Essa função é utilizada em Silva *et al.* (2021) e na Figura 2.3 é possível ter um exemplo do comportamento esperado da função.

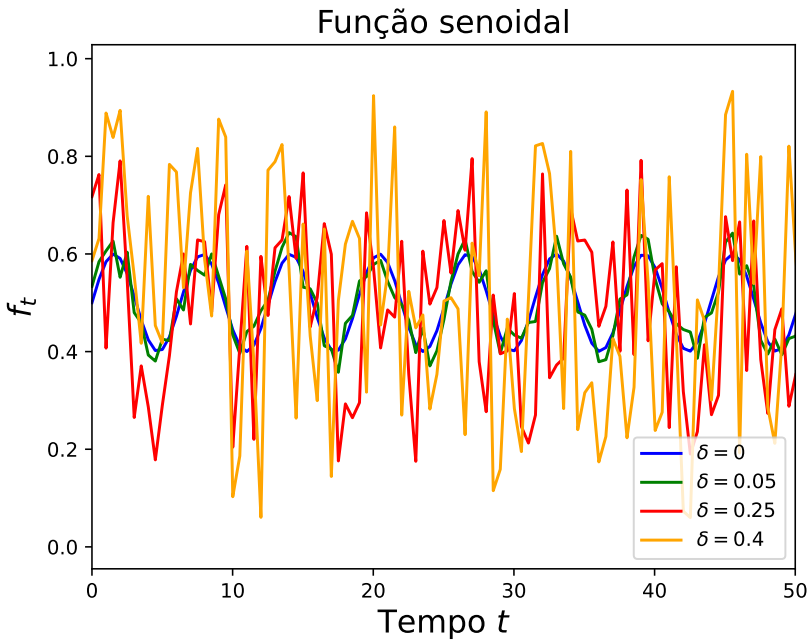



Figura 2.3 – Comportamento da função senoildal da Equação (2.2) com ruídos de diferentes intensidades $\delta$.

### 2.1.2 Métricas e Algoritmos de Comparação entre Séries Temporais

Nessa seção, serão apresentados alguns algoritmos e métricas que são populares na comparação de séries temporais.

#### 2.1.2.1 Dynamic Time Warping

O *Dynamic Time Warping* (DTW) é um algoritmo capaz de medir a similaridade entre duas sequências em um espaço de tempo, mesmo que elas possuam tamanhos diferentes. Foi desenvolvido por Sakoe e Chiba (1978) e é utilizado para reconhecer padrões e similaridades, calculando a distância média entre as séries temporais. Como entrada, o DTW recebe duas sequências de tamanhos variados e constrói a matriz de custo mínimo entre elas. O DTW consegue criar um alinhamento de alta precisão entre séries de diferentes tamanhos, apesar de haver a necessidade de definir um parâmetro para poder interpretar seu resultado com maior eficácia.

Inicialmente, define-se a função de distância utilizada para a comparação entre os pontos

adjacentes das séries. Normalmente, utiliza-se as Equações 2.3 ou 2.4, mostradas a baixo:

$$\delta(i,j) = |s_i - t_j|, \tag{2.3}$$

$$\delta(i,j) = (s_i - t_j)^2, \tag{2.4}$$

onde $s_i$ é o $i$-ésimo valor observado na série $s$ e $t_j$ o $j$-ésimo da série $t$. Ao definir a função, gera-se uma matriz $M \times N$, tal que $M$ é o número de elementos na série $s$ e $N$ é o número de elementos na série $t$. Após aplicar a função de distância entre as séries, realiza-se a busca do menor caminho ótimo dentro da matriz, realizando a soma dos custos de movimento horizontalmente, verticalmente, ou na diagonal. Então, é feito o calculo do canto superior direito até ao canto inferior esquerdo, sempre escolhendo o menor custo, junto da soma desses valores. Ao fim da execução, tem-se o valor do custo mínimo de alinhar as duas séries temporais. Valores baixos de custo mínimo significam que as séries são muito semelhantes ou possuem padrão de comportamento parecido. Como é possível ver na Figura 2.4, mesmo séries com tamanhos diferentes podem ser comparadas utilizando este método. O algoritmo que é utilizado nesta monografia encontra-se implementado e disponibilizado no GitHub [1].

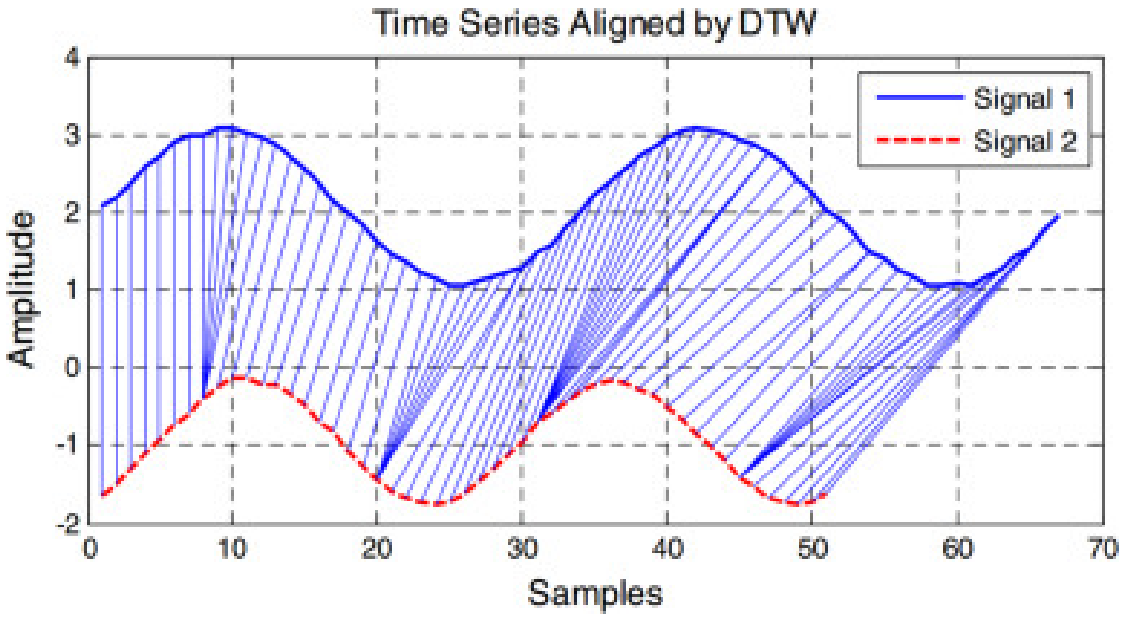



Figura 2.4 – Duas séries temporais alinhadas pelo algorítmo DTW. Fonte: (Lu *et al.*, 2016)

#### 2.1.2.2 Informação Mútua

A Informação Mútua (IM) quantifica a similaridade de pares de variáveis aleatórias. Essa medida faz uso da probabilidade conjunta e da probabilidade marginal das variáveis (Hutter, 2001). Assim como no DTW, a IM não possui uma escala fixa, dificultando a interpretabilidade de seus resultados, não indicando direção da dependência (correlação inversa) nem o tipo (linear ou não-linear). O conceito é definido pela seguinte formula:

[1] <https://github.com/DynamicTimeWarping/dtw-python>

$$I(X;Y) = \sum\sum P_{(X,Y)}(x,y)\log(\frac{P_{(X,Y)}(x,y)}{P_X(x)P_Y(y)}), \tag{2.5}$$

onde $P_{(X,Y)}$ é a função massa de probabilidade conjunta das séries temporais $X$ e $Y$, e $P_X$ e $P_Y$ são as distribuições marginais dessas mesmas séries, as quais ditam individualmente as probabilidades de ocorrência de $X$ e $Y$ respectivamente. Nesse caso, quanto maior o valor da IM, mais as séries temporais são relacionadas entre si.

A implementação da IM da biblioteca Scikit-learn é utilizada nesta monografia, a partir do código disponível no GitHub[2].

#### 2.1.2.3 Coeficiente de Correlação de Pearson

O Coeficiente de Correlação de Pearson (CCP) quantifica a relação linear entre dois vetores $X$ e $Y$, que podem ser, inclusive, séries temporais. Comumente utilizado em análise de dados, o coeficiente determina principalmente o quanto duas variáveis aleatórias estão ligadas entre si pelo seu grau de correlação:

$$r = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}} \tag{2.6}$$

sendo que $r$ pode assumir valores reais entre $-1$ e $1$, e $\overline{x}$ e $\overline{y}$ são as médias das sequências $X$ e $Y$. O valor de $r$ define a similaridade entre as séries: quanto mais perto do valor de $1$, representa uma similaridade maior, quanto mais perto de $-1$, representa uma similaridade inversa, e $0$ representa nenhuma similaridade (Schober; Boer; Schwarte, 2018). O CCP apenas mede a correlação linear entre as séries, não conseguindo indicar possíveis relações não-lineares, além de seu intervalo de valores possíveis pode não ser suficiente para representar com precisão relações complexas.

A implementação que é utilizada neste trabalho vem da biblioteca SciPy[3].

### 2.1.3 Grafos e Redes Funcionais

Grafos são estruturas compostas por vértices (ou nós) que estão conectados por arestas (Rahman *et al.*, 2017). Um sistema pode possuir vários vértices com multiplas arestas. Essas conexões podem ser direcionadas, indicando um fluxo em uma direção específica, ou não-direcionadas, quando não há distinção de direção entre os vértices conectados.

No contexto de redes funcionais, os grafos são utilizados para representar interações entre elementos de um sistema. Quando se aplica comparação de séries temporais, pode-se construir uma rede funcional a partir dessas comparações. Nesse processo, as séries temporais são representadas como vértices, enquanto as conexões (arestas) indicam a similaridade ou relação funcional entre elas. Isso permite identificar padrões de conexão e formar comunidades de séries

[2] <https://github.com/scikit-learn/scikit-learn>
[3] <https://github.com/scipy/scipy>

similares. As redes são uteis em áreas como análise de dados, onde a identificação de padrões e a formação de comunidades podem revelar informações valiosas sobre o comportamento do sistema analisado.

## 2.2 Trabalhos Relacionados

Há uma demanda para armazenar e agrupar dados em intervalos de tempo específicos, para previsão de tendências e detecção de padrões. Essa coleção de dados constitui uma série temporal (Brockwell; Davis, 2002). A análise desses padrões desempenha um papel crucial em diversas áreas da ciência, sendo as previsões um exemplo relevante. Em Donate *et al.* (2013), um estudo é apresentado, no qual uma rede neural é treinada utilizando três técnicas de design automático de redes neurais voltados para previsões de séries temporais. O autor utiliza cinco séries como conjunto de dados para avaliar os métodos, sendo elas: *Passengers*, que contém dados números de passageiros de uma linha área internacional medidos mensalmente; *Temperature*, que contém a temperatura média de uma região dentro de um espaço de tempo; *Dow-Jones*, que contém os lucros mensais da empresa; *Quebec*, que representa o número de nascimentos diariamente da cidade; *Mackey-Glass*, que é uma equação diferencial que gera séries temporais caóticas. O objetivo do trabalho era criar uma rede neural que recebesse qualquer série temporal e fosse capaz de fazer previsões.

Dada a necessidade de análise de séries temporais, ao longo do tempo, foram desenvolvidas várias técnicas com o intuito de aprimorar a comparação entre elas, como exemplo a TSAPI (Silva, 2022). O autor utiliza algoritmos baseados em redes complexas e também em algoritmos tradicionais, visando comparar séries temporais. A principal vantagem da API reside na sua eficiência no processamento de múltiplas séries temporais, uma vez que se trata de uma aplicação distribuída, executada em várias máquinas de forma paralela. Além disso, a API oferece uma variedade de opções para comparações.

Na internet, existem alguns serviços que realizam a tarefa de analise de séries temporais, tais como a Azure Data Explorer[4]. Baseada na computação de núvem da Azure, essa ferramenta auxilia no processamento de grandes quantidades de dados usando a KQL (Linguagem de Consulta Kusto), que possui suporte nativo para análise de séries temporais. Em resumo, a plataforma permite criação de séries temporais, análise de regressão e detecção de sazonalidade para auxiliar no entendimento dos dados. Como mostrado na Figura 2.5, o software é capaz de realizar a comparação de múltiplas séries temporais, além de oferecer uma vizualiação desses dados. No quesito de comparação de dados, a plataforma oferece apenas o Coeficiente de Correlação de Pearson, sendo essa uma limitação.

[4] <https://learn.microsoft.com/pt-br/azure/data-explorer/kusto/query/time-series-analysis>

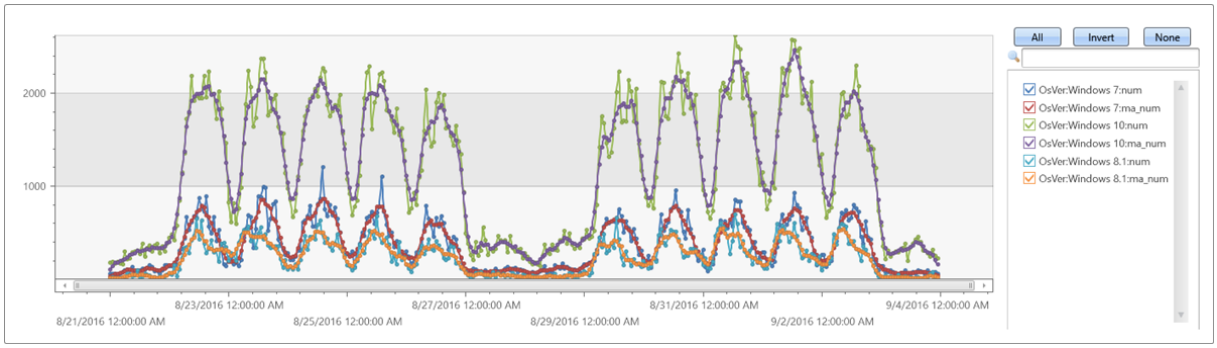



Figura 2.5 – Azure Data Explorer com múltiplas séries temporais. Fonte: <https://learn.microsoft.com/pt-br/azure/data-explorer/kusto/query/time-series-analysis>

Uma das áreas em que a comparação de séries temporais se mostra útil é na geração de redes funcionais. Ferreira *et al.* (2021) gera redes funcionais a partir de uma base de dados de séries temporais climáticas, com o intuito de mapear pontos de interação climática no planeta Terra. Esse estudo utiliza diversas funções, como o Coeficiente de Correlação de Pearson e a Informação Mútua, para identificar similaridades entre as bases de dados e criar as redes funcionais. Essas redes permitem destacar inter-relações entre os pontos das séries temporais que são relevantes para a análise do clima global ou regional, contribuindo em pesquisas relacionadas às mudanças climáticas.

Nesse mesmo contexto, Donges *et al.* (2009) demonstram duas abordagens na criação de redes complexas a partir da comparação de séries temporais de dados relacionados à temperatura terrestre a fim de comparar modelos lineares e não-lineares. No artigo, os autores fazem uso tanto do Coeficiente de Correlação de Pearson quanto da Informação Mútua. É possível observar que as duas medidas encontraram resultados similares quanto aos dados de temperaturas regionais. Entretanto, ao utilizar a comparação em dados de escala global, a Informação Mútua encontrou mais pares similares de séries em comparação à CCP, sugerindo que este consiga detectar melhor as relações não-lineares entre os nós.

# 3 Desenvolvimento

Esta monografia busca avaliar em quais casos as métricas e algoritmos propostos para comparação de séries temporais são mais efetivos. Para tanto, serão avaliadas comparações entre os tipos de séries a seguir:

- Periódicas, geradas pela Equação (2.2) e pelo Mapa Logístico, da Seção 2.1.1.1;
- Caóticas, também geradas pelo Mapa Logístico;
- Combinações entre séries caóticas e periódicas, para capturar diferentes nuances.
- Periódicas com adição de ruído, a exemplo do trabalho de Silva *et al.* (2021);
- Séries temporais reais de *datasets* do Kaggle e suas versões ruidosas.

Em se tratando das métricas e algoritmos para comparação de séries, tenciona-se utilizar Coeficiente de Correlação de Pearson (CCP), Informação Mútua (IM) e *Dynamic Time Warping* (DTW). No quesito de implementação, é utilizado a biblioteca *SciPy* em Python para Coeficiente de Correlação de Pearson; Para a Informação Mútua, é utilizado a biblioteca *Scikit-Learn* em Python; O algoritmo do *Dynamic Time Warping* se encontra no GitHub[1] como repositório. Em todos os casos, serão utilizadas duas séries temporais como entrada, seguindo as limitações dos algorítmos e métricas utilizadas.

Para gerar as séries temporais, os dois métodos utilizados são o Mapa Logístico, da Seção 2.1.1.1, e a Função senoidal, da Seção 2.1.1.2. Para o Mapa Logístico, o valor inicial é fixado em $x_0 = 0.3$ para todos os casos. São realizadas $1000$ iterações da função, com o descarte das $500$ primeiras que são considerados pontos transientes, já que o sistema só se estabiliza depois de certo tempo, como mostra o exemplo da Figura 3.1.

[1] <https://github.com/DynamicTimeWarping/dtw-python>

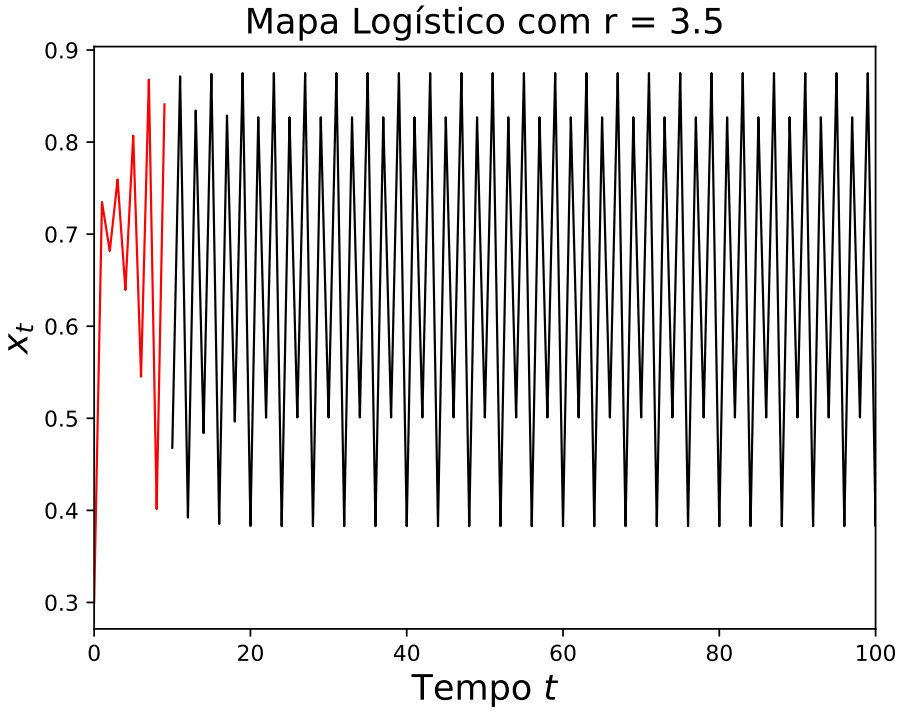



Figura 3.1 – Pontos transientes destacados em vermelho, onde é possível observar que a série temporal ainda não se estabilizou.

Para a Equação (2.2), utiliza-se 500 iterações da função para fins comparativos, com $t \in [0, \ 50]$ e incrementos de $\Delta t = 0.1$. A fonte de ruído $\xi(t)$ possui como único parâmetro a intensidade $\delta \in \{0, \ 1\}$. Para as séries dos *datasets*, o ruído $\xi(t)$ possui intensidade $\delta \in \{0, \ 20\}$, já que seus valores originais são maiores. São gerados 100 valores de ruídos variando entre a intensidade $\delta$ e são feitas 100 iterações de cada método para cada valor de intensidade de ruído, a fim de obter o desvio padrão de cada método, apresentado pelo sombreado nas figuras.

As séries sintéticas comparadas nos experimentos possuem propriedades de séries reais em alguns contextos. Dentro do contexto de clima, espera-se que os valores tenham baixa previsibilidade, o que pode ser ilustrado com séries caóticas. Ao mesmo tempo, encontra-se certa periodicidade no comportamento de algumas variáveis climáticas e é essa a motivação de utilizarmos também séries periódicas e mesmo combinações de séries periódicas e caóticas.

## 3.1 Experimentos

Esta seção detalha os 10 experimentos realizados. Todos foram feitos em uma máquina Acer Nitro 5, com sistema operacional Windows 11 Home, 24gb RAM e processador Intel Core i5 11400H.

### 3.1.1 Experimento 1 - Função senoidal com ruídos

No experimento 1 realiza-se a avaliação das métricas e algoritmos previamente citados entre séries períodicas geradas pela Equação (2.2) com a adição de ruídos. Com o objetivo de verificar a eficácia das métricas para detecção de ruídos, a comparação se dá por uma série temporal períodica sem ruídos, ou seja, $\delta = 0$, com outra série onde a perturbação será aplicada. A Figura 3.2 mostra a representação da série de forma gráfica.

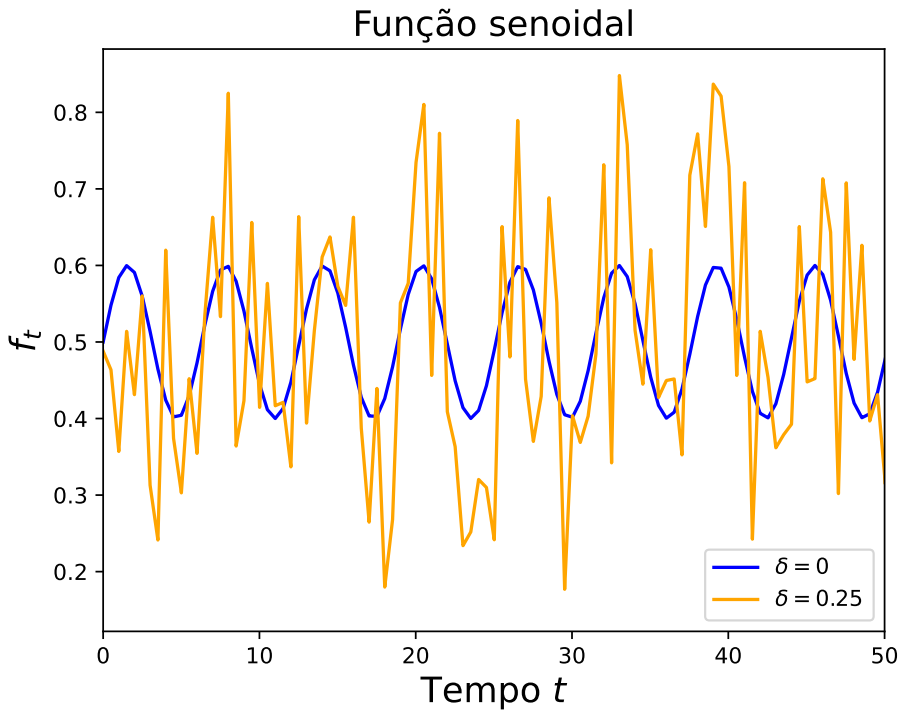



Figura 3.2 – Série temporal gerada pela função senoidal e sua versão ruidosa com $\delta = 0.25$.

### 3.1.2 Experimento 2 - Mapa Logístico com $r = 3.5$ (período 4) e ruído

No segundo experimento também é realizada a avaliação das métricas e algoritmos em séries períodicas com adição de ruídos, mas geradas pela Equação (2.1), demonstrado na Figura 3.3. Tal como o experimento anterior, é gerado uma série períodica sem adição de ruídos com o parâmetro do Mapa Logístico definido em $r = 3.5$ e de valor inicial $x_0 = 0.3$, utilizando 500 iterações, para comparar com outra de mesmo paramêtro, mas com ruído de intensidade $\delta$. Sendo assim, temos uma série de período 4, ou seja, existem 4 valores $x_1, x_2, x_3$ e $x_4$ que se repetem durante a série.

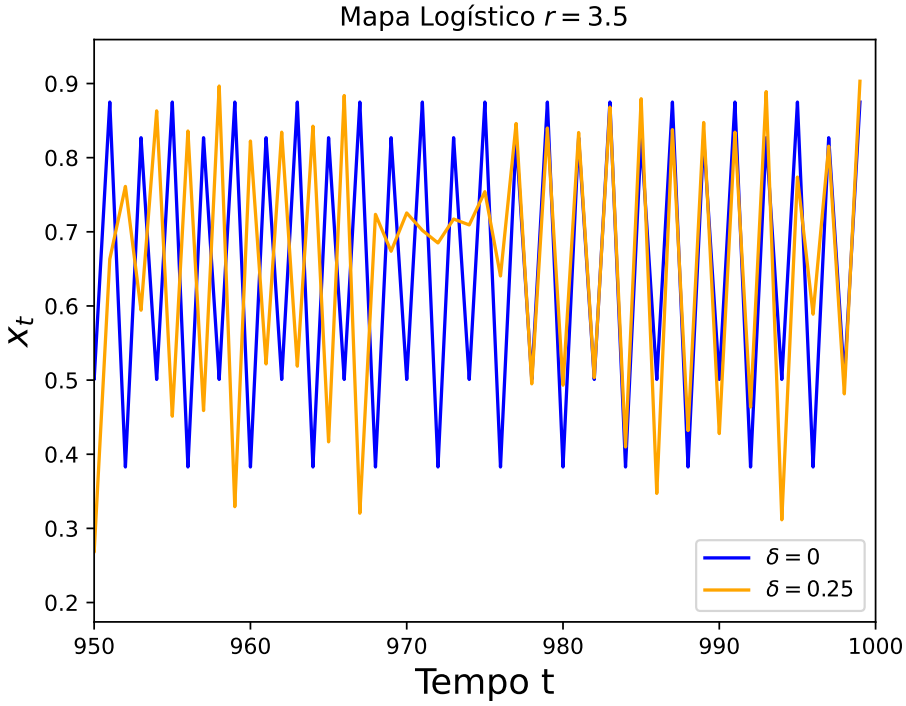



Figura 3.3 – Série temporal gerada pelo Mapa Logístico com $r = 3.5$ e sua versão ruidosa com intensidade $\delta = 0.25$.

### 3.1.3 Experimento 3 - Mapa Logístico com $r = 3.95$ (caótico) e ruído

Um outro experimento realizado é com séries temporais caóticas geradas pelo Mapa Logístico com adição de ruídos. Primeiro gera-se uma série temporal com o parâmetro $r = 3.95$ e de valor inicial $x_0 = 0.3$ para obter uma série caótica, e em seguida gera-se a mesma série com ruídos $\xi(t)$, mostrado na Figura 3.4.

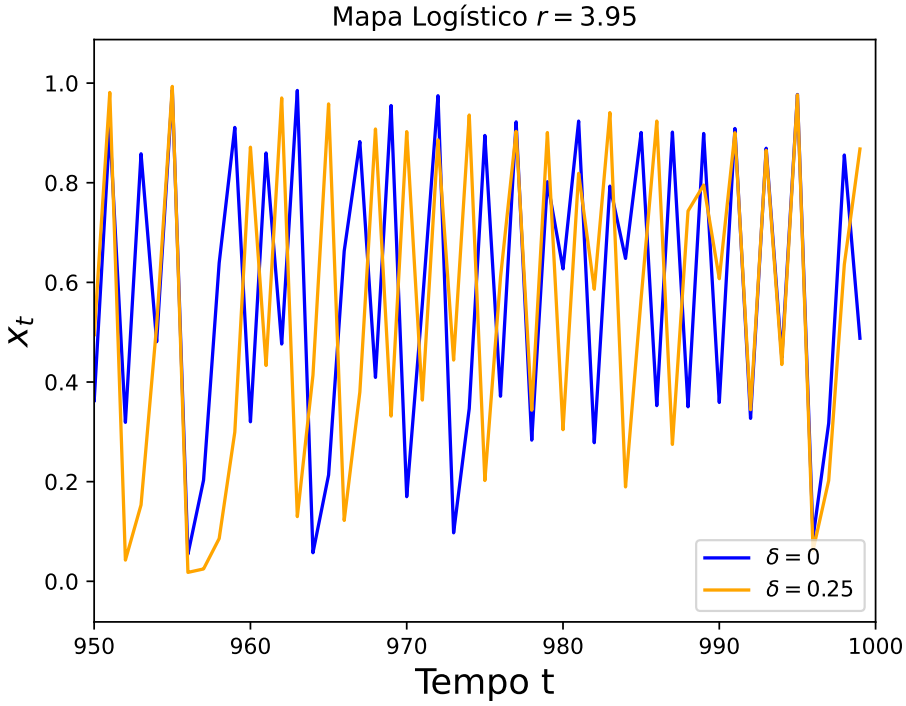



Figura 3.4 – Série temporal gerada pelo Mapa Logístico com $r = 3.95$ e sua versão ruidosa com intensidade $\delta = 0.25$.

### 3.1.4 Experimento 4 - Mapa Logístico com $r = 3.8$ (caótico) e $r = 3.87$ (caótico)

Nessa parte, são geradas duas séries temporais caóticas diferentes pelo Mapa Logístico com parâmetros próximos. A primeira série é gerada com o parâmetro $r = 3.8$ e valor inicial $x_0 = 0.3$, e a segunda série possui o parâmetro $r = 3.87$ e valor inicial $x_0 = 0.3$, como é possível ver na Figura 3.5.

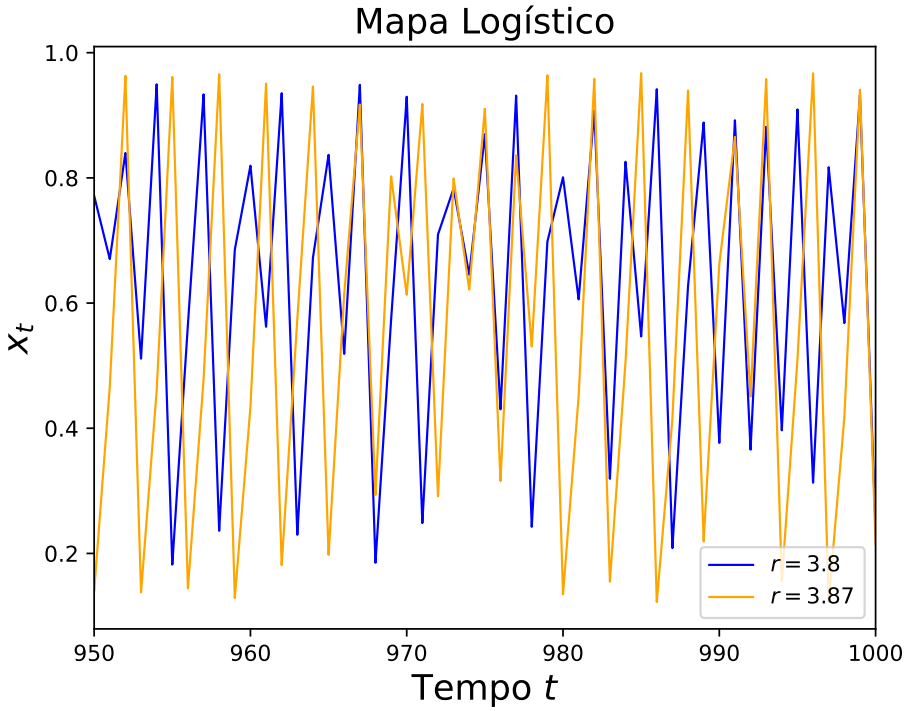



Figura 3.5 – Séries temporais gerada pelo Mapa Logístico com $r = 3.8$ em azul, e $r = 3.87$ em laranja.

### 3.1.5 Experimento 5 - Mapa Logístico com $r = 3.678$ (caótico) e $r = 4$ (caótico)

Aqui, são geradas duas séries temporais caóticas diferentes pelo Mapa Logístico com parâmetros distantes. A primeira série é gerada com o parâmetro $r = 3.678$ e valor inicial $x_0 = 0.3$, e a segunda série possui o parâmetro $r = 4$ e valor inicial $x_0 = 0.3$, como mostra a Figura 3.6.

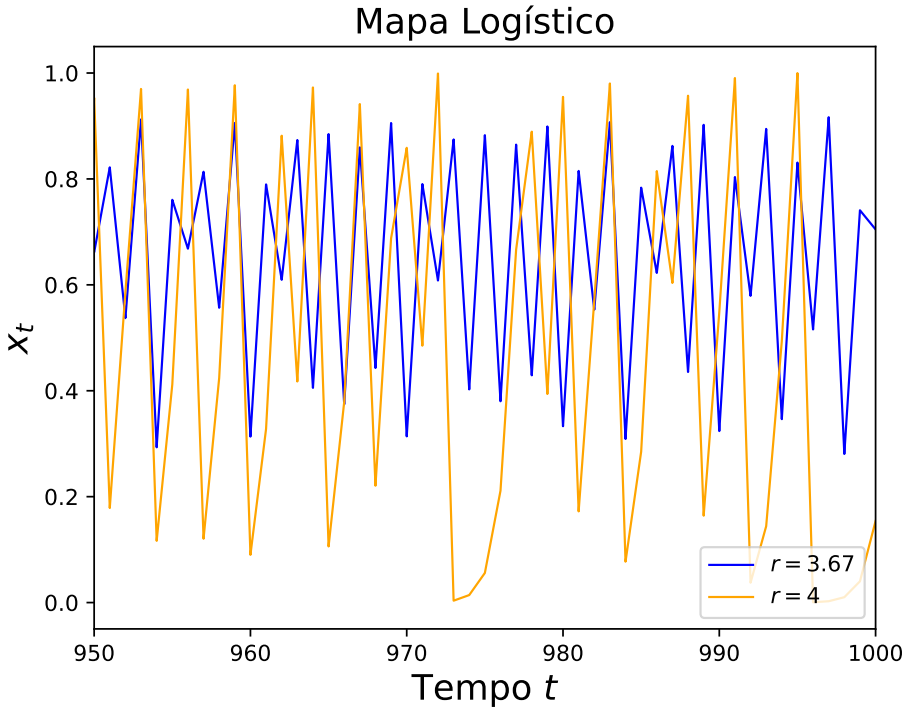



Figura 3.6 – Séries temporais geradas pelo Mapa Logístico com $r = 3.67$ em azul, e $r = 4$ em laranja.

### 3.1.6 Experimento 6 - Mapa Logístico com $r = 3.891$ (caótico) e $r = 3.2$ (período 4)

Nesse experimento, é realizada a comparação de uma série temporal caótica de parâmetro $r = 3.891$ e valor inicial $x_0 = 0.3$ com uma série temporal períodica gerada pelo Mapa Logístico de período 4, com parametro $r = 3.2$ e de valor inicial $x_0 = 0.3$, mostrados na Figura 3.7.

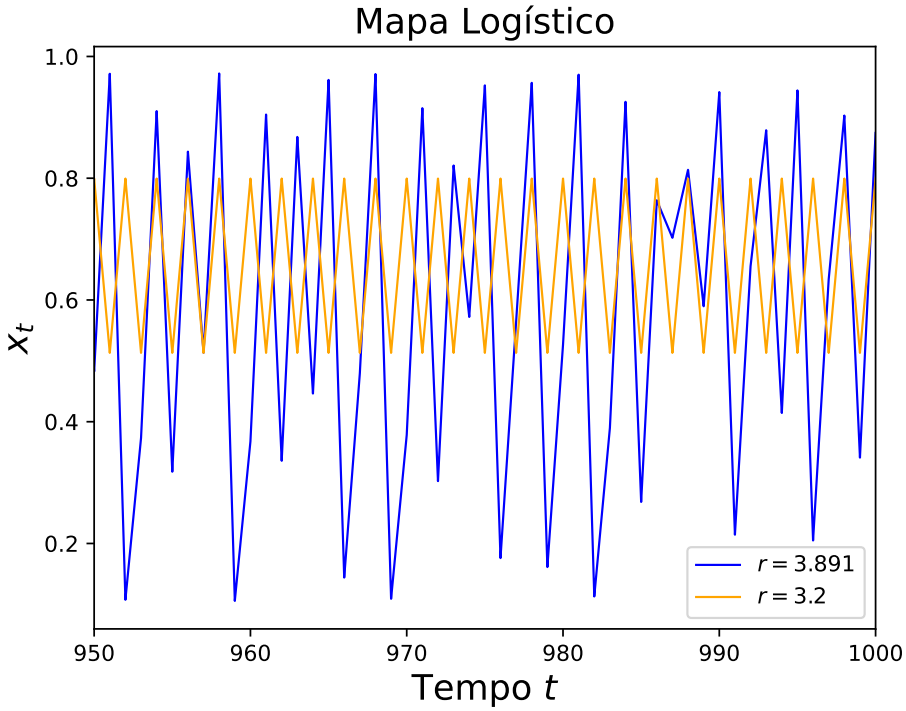



Figura 3.7 – Séries temporais geradas pelo Mapa Logístico, sendo $r = 3.891$ em azul e $r = 3.2$ em laranja.

### 3.1.7 Experimento 7 - Combinação de séries do Mapa Logístico com $r = 3.88$ (caótico) e $r = 3.5$ (período 4) comparada com o Mapa Logístico com $r = 3.9592$ (caótico)

Nesse experimento, realiza-se a soma de uma série caótica de parâmetro $r = 3.88$ e valor inicial $x_0 = 0.3$ com outra períodica gerada pelo Mapa Logístico de parâmetro $r = 3.5$ e de valor inicial $x_0 = 0.3$, sendo a série períodica de período 4, com o propósito de comparar com uma outra série caótica de parâmetro $r = 3.9592$ e valor inicial $x_0 = 0.3$. Todas as séries são mostradas na Figura 3.8.

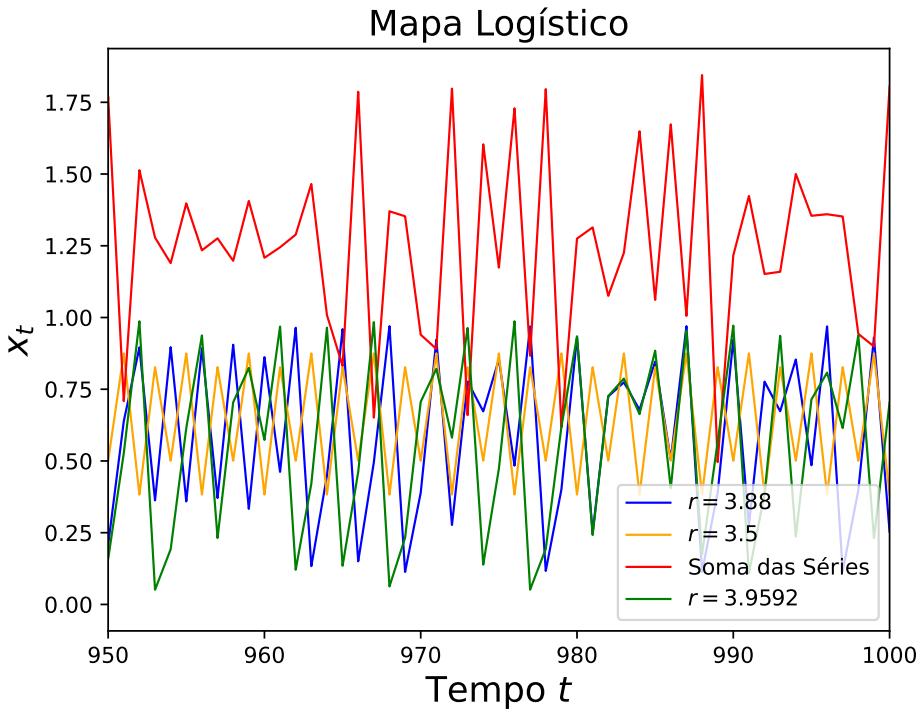



Figura 3.8 – Séries temporais geradas pelo Mapa Logístico, com $r = 3.88$ em azul, $r = 3.5$ em laranja, a soma dessas duas séries em vermelho e a série com $r = 3.9592$ em verde.

### 3.1.8 Experimento 8 - Combinação de séries do Mapa Logístico com $r = 3.9$ (caótico) e senoidal comparada com o Mapa Logístico com $r = 3.891$ (caótico)

Nesse experimento também é realizada a combinação entre diferentes séries temporais. A soma se dá por uma série caótica de parâmetro $r = 3.9$ e valor inicial $x_0 = 0.3$, com a série periódica gerada pela Equação (2.2). A série resultante dessa soma é comparada com uma outra série caótica de parâmetro $r = 3.891$ e valor inicial $x_0 = 0.3$. O gráfico das séries são mostrados na Figura 3.9.

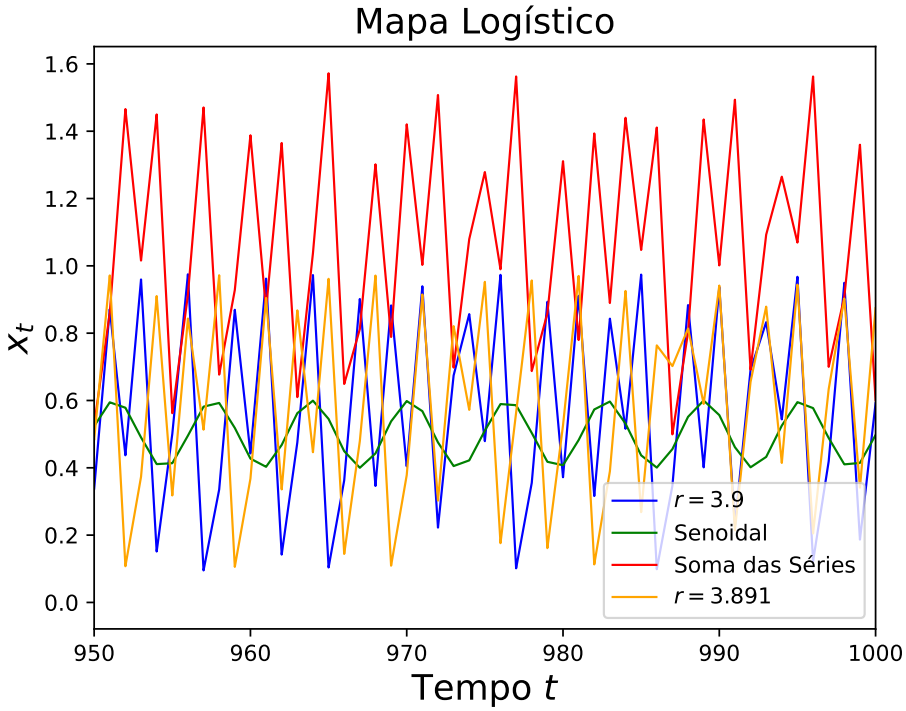



Figura 3.9 – Série temporal gerada pelo Mapa Logístico, com $r = 3.9$ em azul, e a série temporal gerada pela Equação (2.2) em verde. A soma dessas duas séries destacada em vermelho e a série temporal do Mapa Logístico com $r = 3.891$ em laranja.

### 3.1.9 Experimento 9 - Combinação de caótica + periódica com caótica + periódica, com mesmas séries caóticas e séries periódicas diferentes

Para esse experimento, compara-se duas séries temporais resultantes de uma soma. A primeira série é gerada pela soma de uma série caótica de parâmetro $r = 3.734$ e valor inicial $x_0 = 0.3$ com uma série períodica gerada pelo Mapa Logístico de período 4, com o parâmetro $r = 3.5$ e de valor inicial $x_0 = 0.3$. A segunda série é gerada pela soma da mesma série caótica mencionada acima com a série períodica gerada pela Equação (2.2). As séries desse experimentos são demonstradas nas Figuras 3.10 e 3.11.

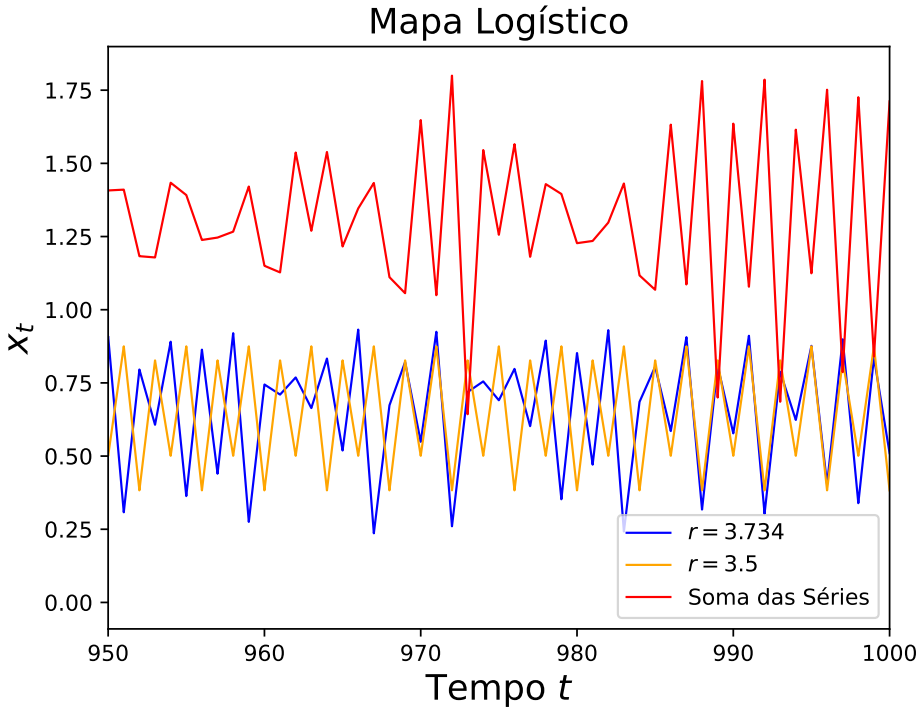



Figura 3.10 – Séries temporais geradas pelo Mapa Logístico com $r = 3.734$ em azul, $r = 3.5$ em laranja e a soma dessas duas séries em vermelho.

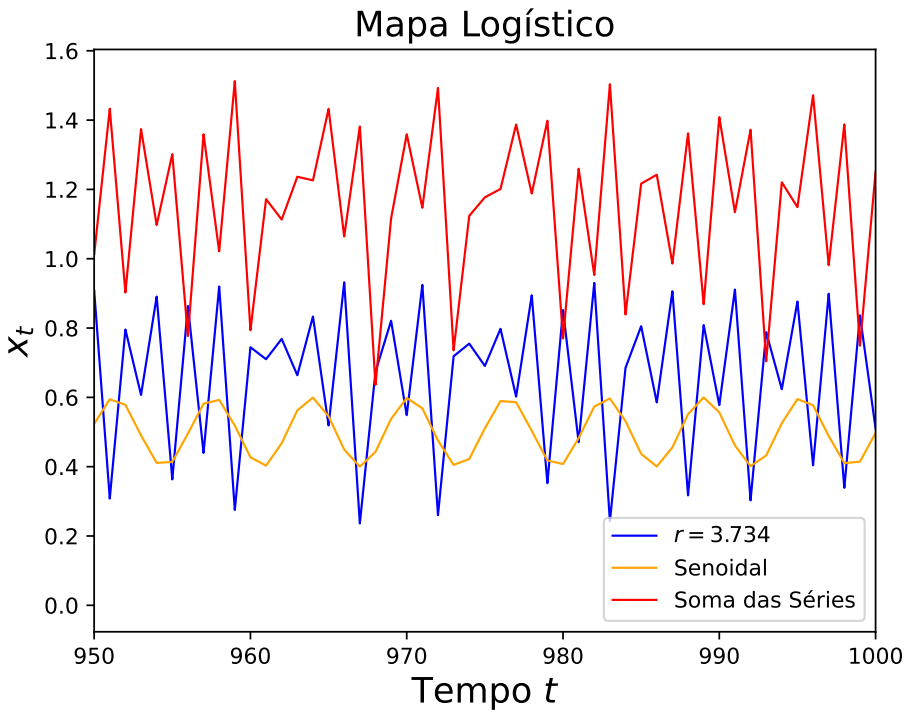



Figura 3.11 – Séries temporais geradas para o experimento. Em azul há a série gerada pelo Mapa Logístico com $r = 3.734$. Em laranja a série é gerada pela função senoidal. E em vermelho está a soma dessas duas séries.

### 3.1.10 Experimento 10 - Dados reais com ruído aditivo

Para o último experimento, são utilizados *datasets* do Kaggle, onde contém 4 tipos de séries temporais provenientes de dados reais. Essas séries são as seguinte: Eletricidade, que mostra a produção elétrica diária de uma cidade, mostrado na Figura 3.12; Temperaturas, que mostra as temperaturas mínimas diárias de Melbourne, mostrado na Figura 3.13; Cerveja, que mostra a venda mensal de cerveja na Austrália, representado na Figura 3.14; e Shampoo, que representa a venda mensal de shampoos em um período de três anos, mostrado na Figura 3.15. Todos esses dados se encontram disponíveis no Kaggle[2]. Os experimentos se dão pela comparação das próprias séries temporais com suas versões ruidosas, a partir de uma fonte de ruído $\xi(t)$ com intensidade $\delta$, como nos outros experimentos com ruídos.

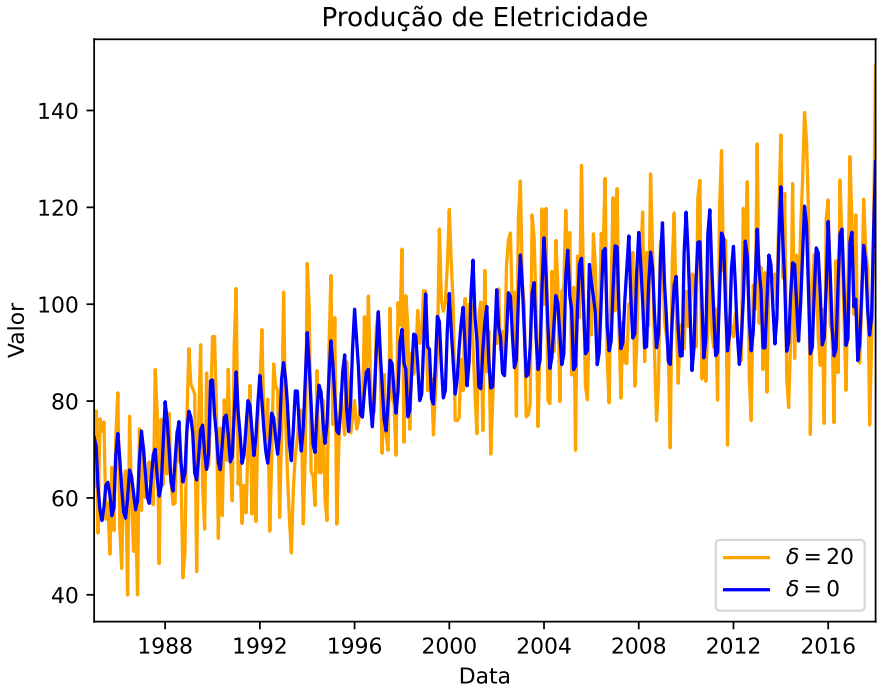



Figura 3.12 – Série temporal obtida do *dataset* do *Kaggle* que representa a produção de energia de uma elétrica e sua versão com ruídos de itensidade $\delta$.

[2] <https://www.kaggle.com/datasets/shenba/time-series-datasets/>

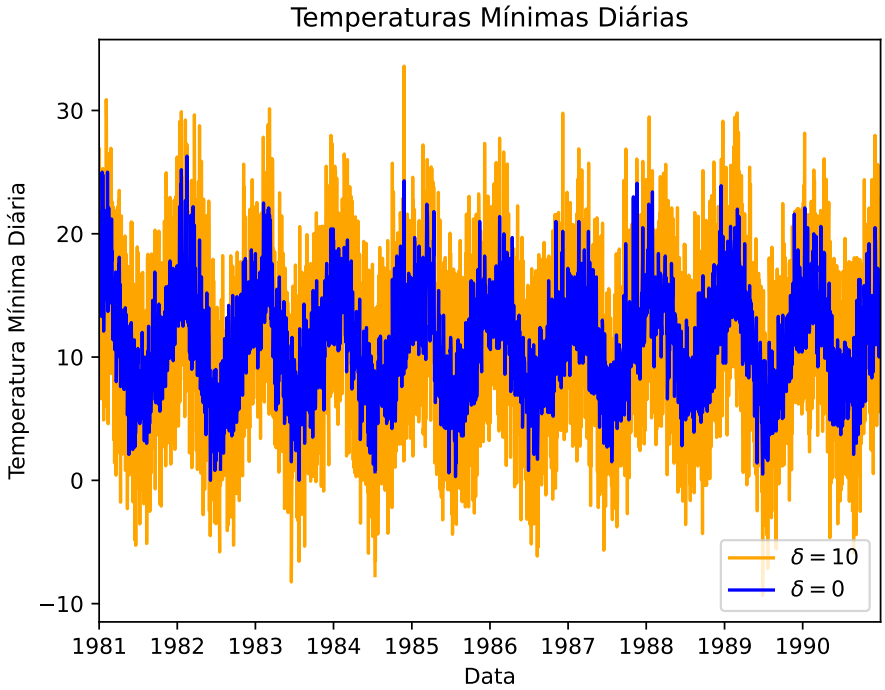



Figura 3.13 – Série temporal obtida do *dataset* do *Kaggle* que representa as temperaturas mínimas diárias ao longo dos anos e sua versão com ruídos de itensidade $\delta$.

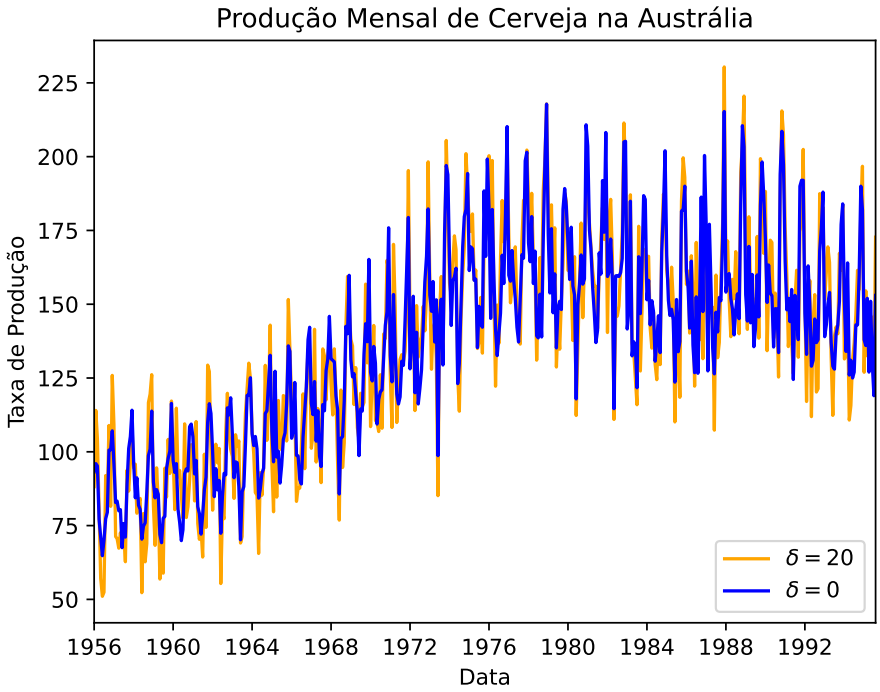



Figura 3.14 – Série temporal obtida do *dataset* do *Kaggle* que representa produção mensal de cerveja na Australia e sua versão com ruídos de itensidade $\delta$.

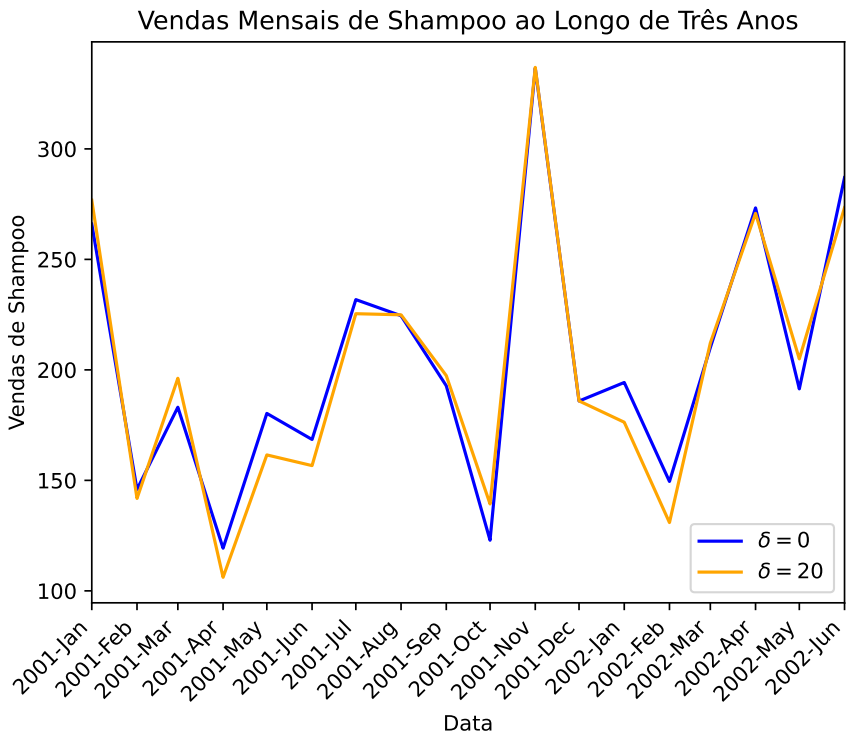
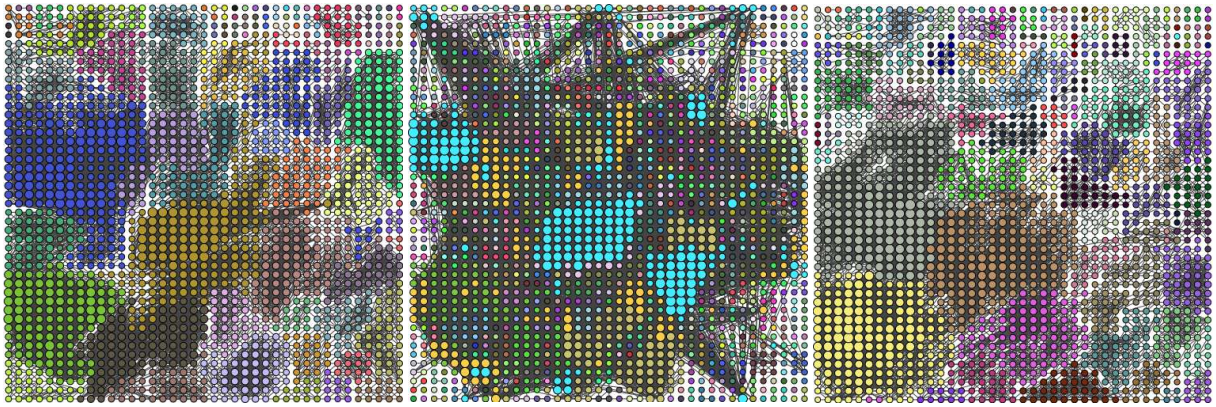



Figura 3.15 – Série temporal obtida do *dataset* do *Kaggle* que representa a venda mensal de shampoos ao longo de três anos e sua versão com ruídos de itensidade $\delta$.

# 4 Resultados

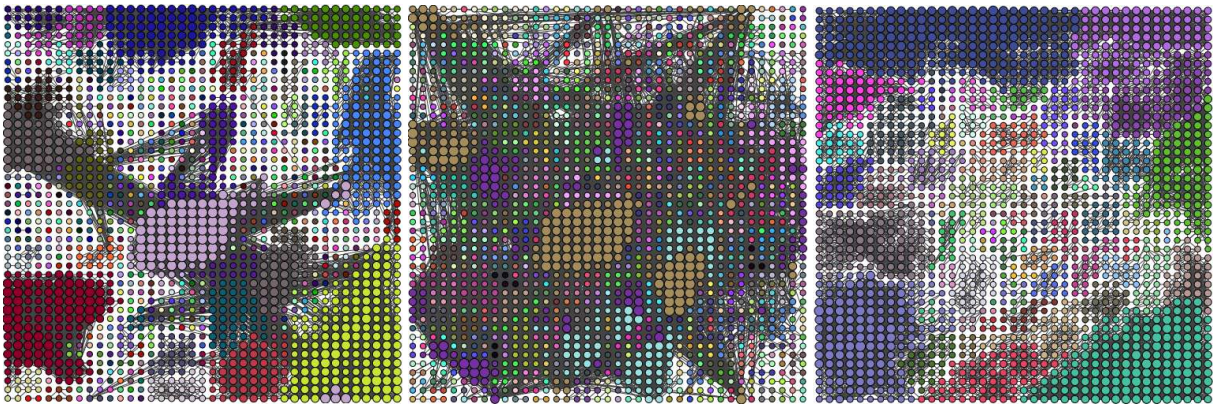
Antes de relatar os resultados obtidos nos experimentos, trago uma contextualização da aplicação das métricas e algoritmos de comparação de séries temporais na geração de redes funcionais, a fim de guiar a interpretação do leitor.

De acordo com o trabalho de Diniz (2022), onde o autor gerou redes funcionais a partir de séries temporais, as redes funcionais geradas por IM e CCP são relativamente parecidas, tendo concentrações de grupos de mesma cor nos grafos gerados. Já para as redes geradas por DTW, é possível notar conexões de longas distâncias, como é possível observar na Figura 4.1. No caso onde as redes funcionais foram geradas a partir de uma limiarização local, via a extração do backbone, aquelas geradas por CCP e DTW também demonstram conexões de longa distância, mas com CCP ainda preservando grupos concentrados, enquanto o IM mantém sua concentração de comunidades sem conexões de longo alcance, como mostrado na Figura 4.2. Esse comportamento das redes funcionais também é notado no trabalho de Valério (2024), onde as comunidades demonstradas por MI e CCP possuem uma concentração de grupos locais grandes e poucas conexões de longa distância e as geradas por DTW mostram conexões de longo alcance. As redes geradas com a utilização de *backbones* conseguem captar algumas conexões de longo alcance utilizando o CCP.



Rede CCP Rede DTW Rede IM

Figura 4.1 – Redes funcionais geradas por comparação de séries temporais com distinção de comunidades por cores sem *backbones*. Fonte: (Diniz, 2022)



Rede CCP Rede DTW Rede IM

Figura 4.2 – Redes funcionais geradas por comparação de séries temporais com distinção de comunidades por cores com *backbones*. Fonte: (Diniz, 2022)

Em Ferreira *et al.* (2021), as redes funcionais geradas apontam que o DTW foi o que mais se destacou em captar padrões em teleconexões que as outras métricas e algorítmos não foram capazes de detectar, indicando que a métrica é promissora em captar padrões não-lineares. No contexto de dinâmica climática, isso é importante pois fenômenos podem influenciar a temperatura mesmo a longas distâncias.

Os resultados obtidos na presente monografia apontam que as séries temporais sintéticas com adição de ruído seguem um mesmo padrão. As nuances detectadas pelos três métodos são demonstradas nas Figuras 4.3, 4.4 e 4.5. Como é possível notar, a métrica de Coeficiente de Correlação de Pearson, nos três casos, diminui a sensibilidade a ruídos acima de $\delta \approx 0.5$, não gerando novos valores distantes e apenas indicando que as séries comparadas são diferentes. Para a Informação Mútua, isso ocorre ainda mais cedo, quando $\delta \approx 0.2$. O *Dynamic Time Warping* consegue detectar bem as alterações de acordo com o aumento da intensidade do ruído, tendo uma relação linear com a intensidade do ruído. Se tratando do tempo de execução, o algoritmo do DTW foi o mais custoso, devido a necessidade do cálculo da matriz de custo mínimo, seguido da IM e, por último, CCP.

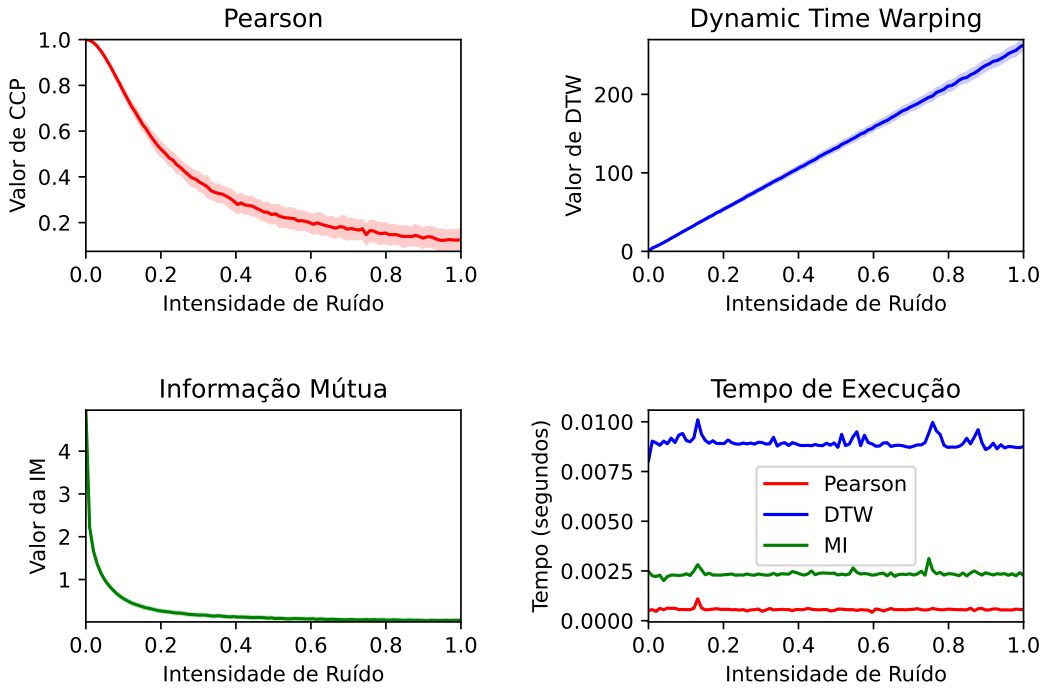



Figura 4.3 – Resultados do Experimento 1 - Função senoidal com ruídos.

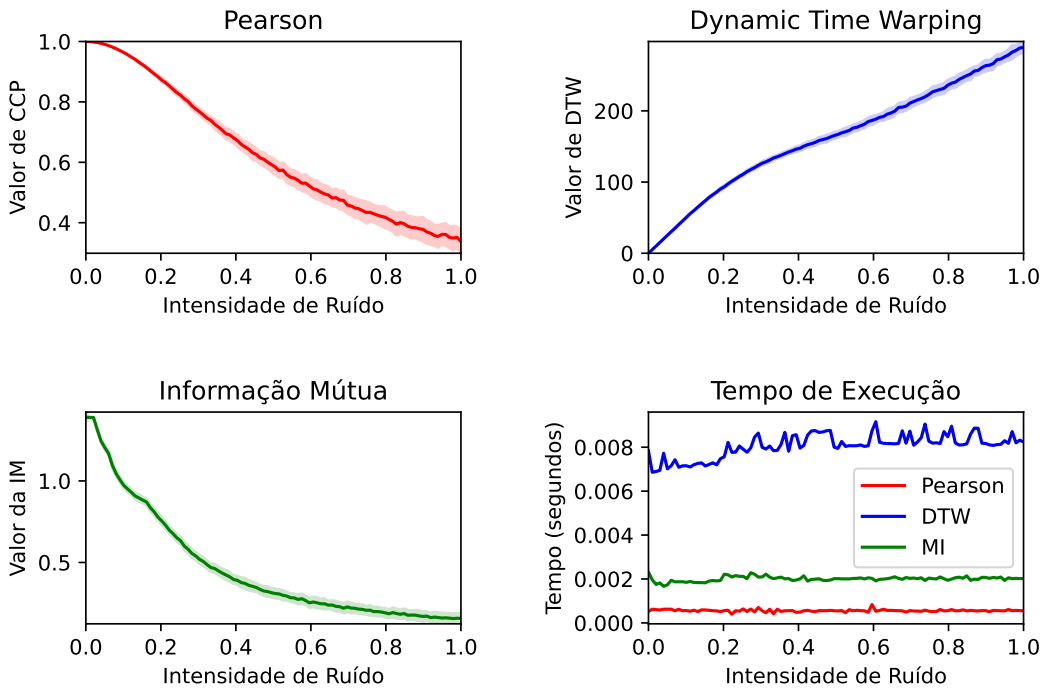



Figura 4.4 – Resultados do Experimento 2 - Mapa Logístico com r = 3.5 (período 4) e ruído.

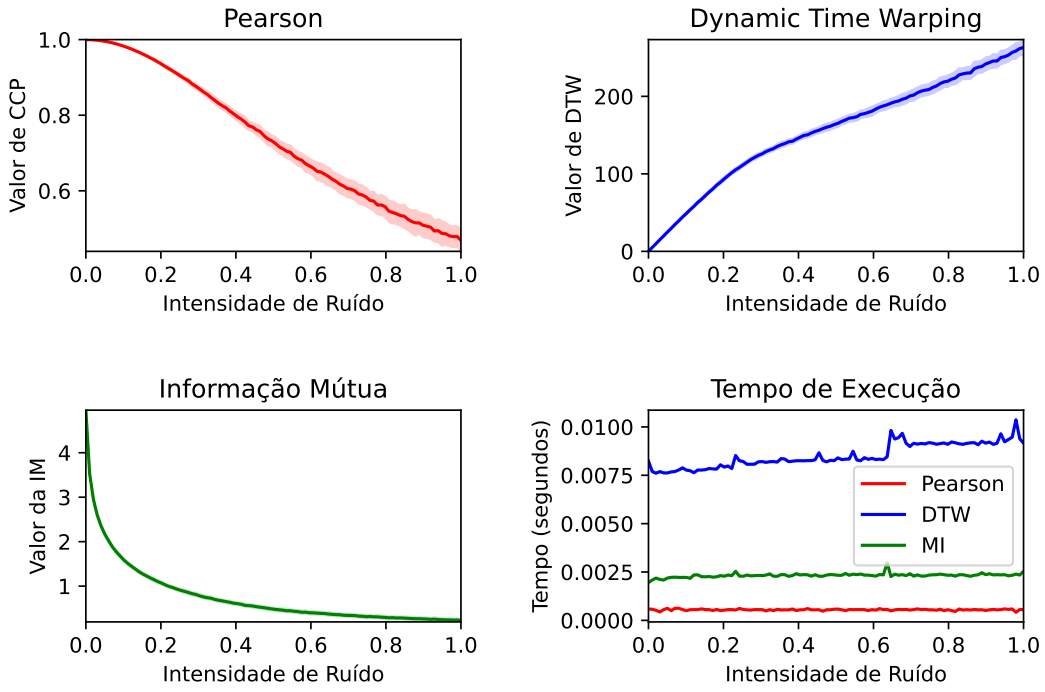



Figura 4.5 – Resultados do Experimento 3 - Mapa Logístico com r = 3.95 (caótico) e ruído.

Os Experimentos voltados para séries sintéticas caóticas (comparação entre séries caóticas e a fusão entre caóticas com periódicas) são mostrados nas Figuras 4.6, 4.9 e 4.8. Para esses casos, apenas o *Dynamic Time Warping* aponta alguma semelhança entre essas séries, tomando os valores anteriores de DTW como referência. Tanto o Informação Mútua quanto o Coeficiente de Correlação de Pearson não detectaram semelhanças nesses Experimentos, resultando em valores muito baixos, indicando que as séries são diferentes e não possuem correlação. Em questão de tempo de execução, o DTW se mostra com o maior custo.

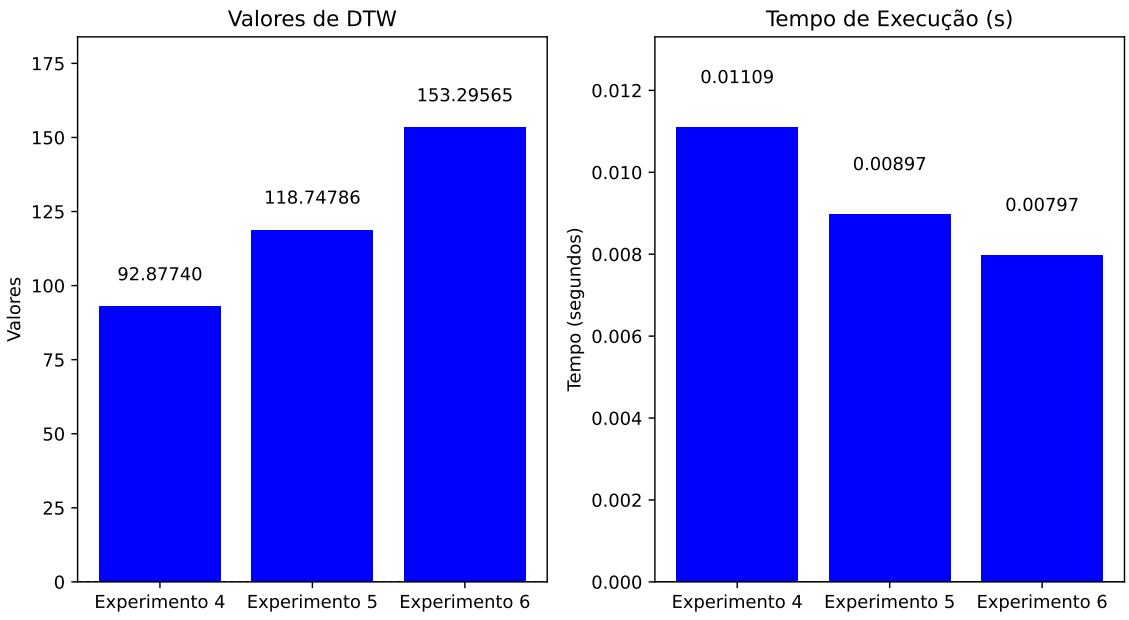



Figura 4.6 – Resultados do algorítmo *Dynamic Time Warping* para os Experimentos 4, 5 e 6

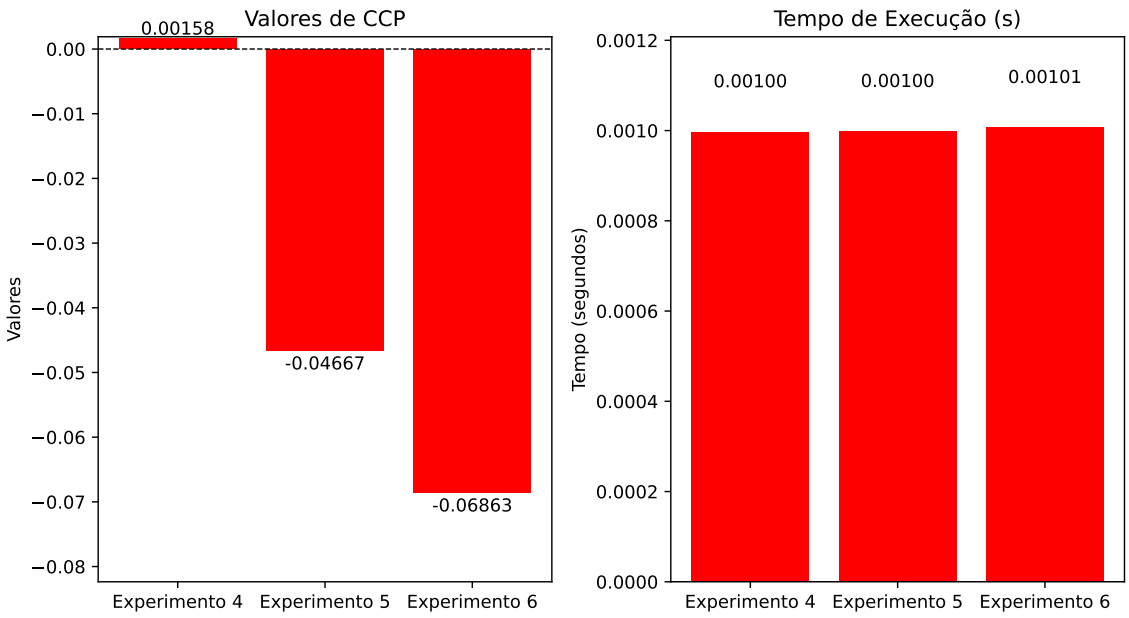



Figura 4.7 – Resultados do algorítmo Coeficiente de Correlação de Pearson para os Experimentos 4, 5 e 6

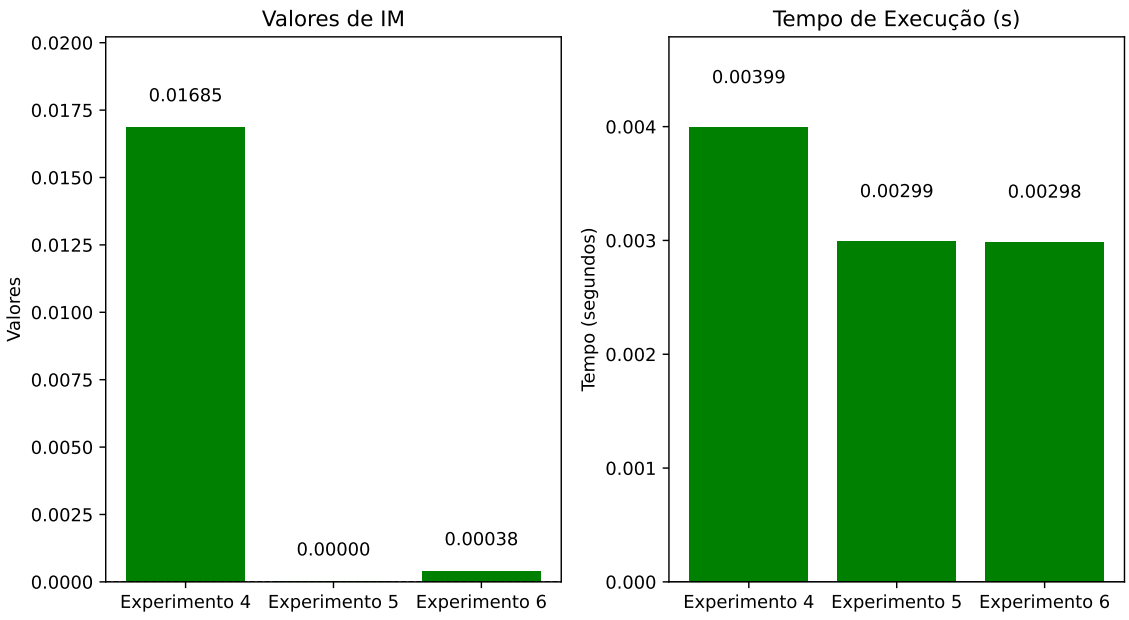



Figura 4.8 – Resultados do algorítmo Informação Mútua para os Experimentos 4, 5 e 6

Já em relação aos experimentos de comparação das somas de séries sintéticas, é possível observar que todas as métricas conseguiram detectar o mesmo nível de similaridade, como mostra as Figuras 4.9, 4.10 e 4.11. Levando em conta os valores do Experimento anterior, pode-se afirmar que todas, com exceção do Experimento 9, não são séries similares. Os valores obtidos de DTW para os Experimentos 7 e 8 indicam que as séries comparadas estão distantes entre si, assim como os valores de IM e CCP indicam que as séries são diferentes e não possuem correlação entre si.

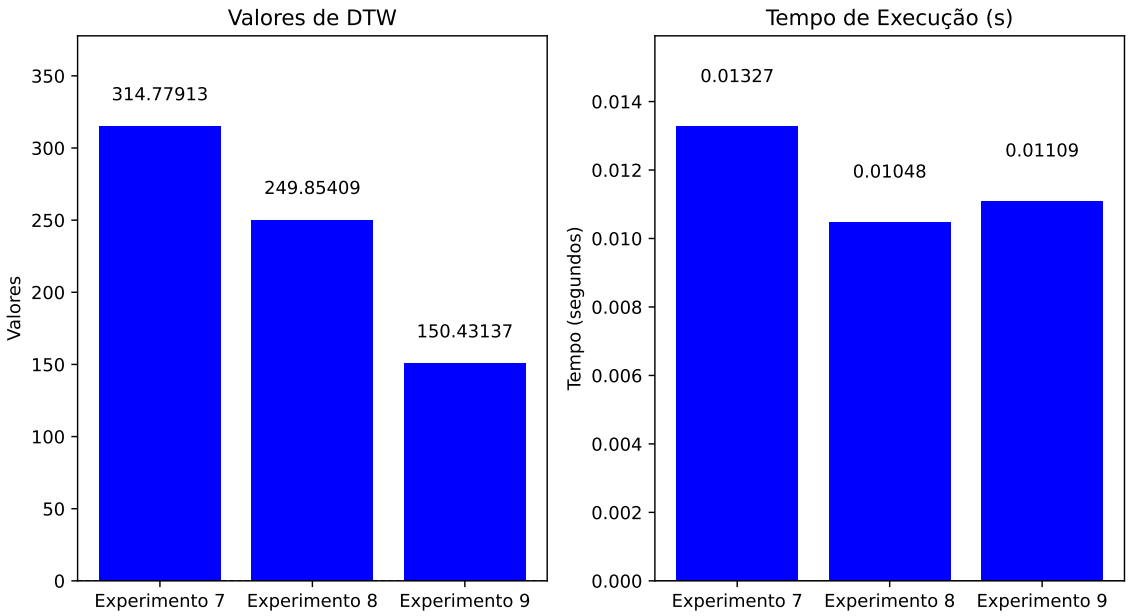



Figura 4.9 – Resultados do algorítmo *Dynamic Time Warping* para os Experimentos 7, 8 e 9.

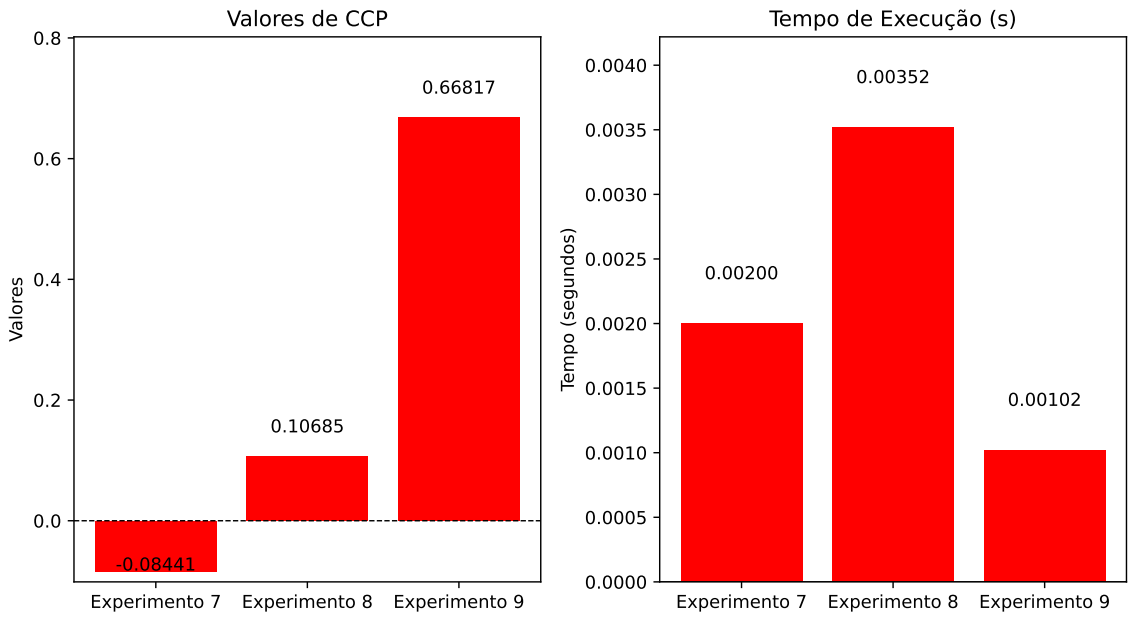



Figura 4.10 – Resultados da métrica Coeficiente de Correlação de Pearson para os Experimentos 7, 8 e 9.

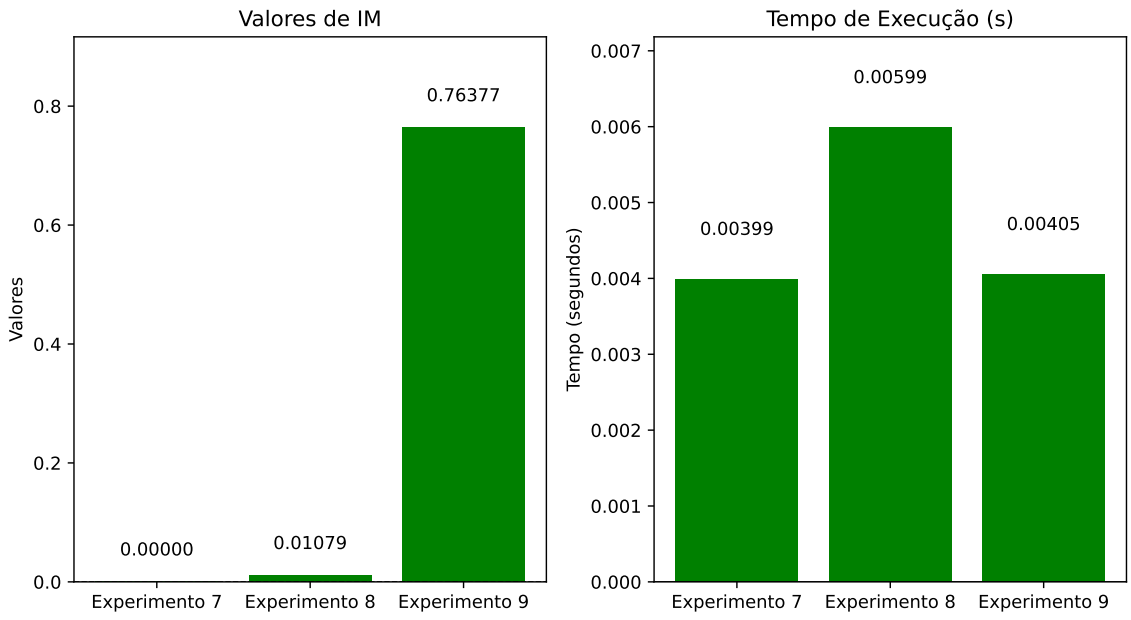



Figura 4.11 – Resultados do algorítmo Informação Mútua para os Experimentos 7, 8 e 9.

Para uma visualização ampla, a Figura 4.12 mostra todos os valores obtidos nos experimentos com as séries sintéticas. Para os experimentos de incremento de ruído, foram selecionados os valores obtidos ao compara a série com intensidade de ruído igual a zero ($\sigma = 0$), com valor médio de ruído ($\sigma = 0.5$) e com o máximo de ruído ($\sigma = 1$).

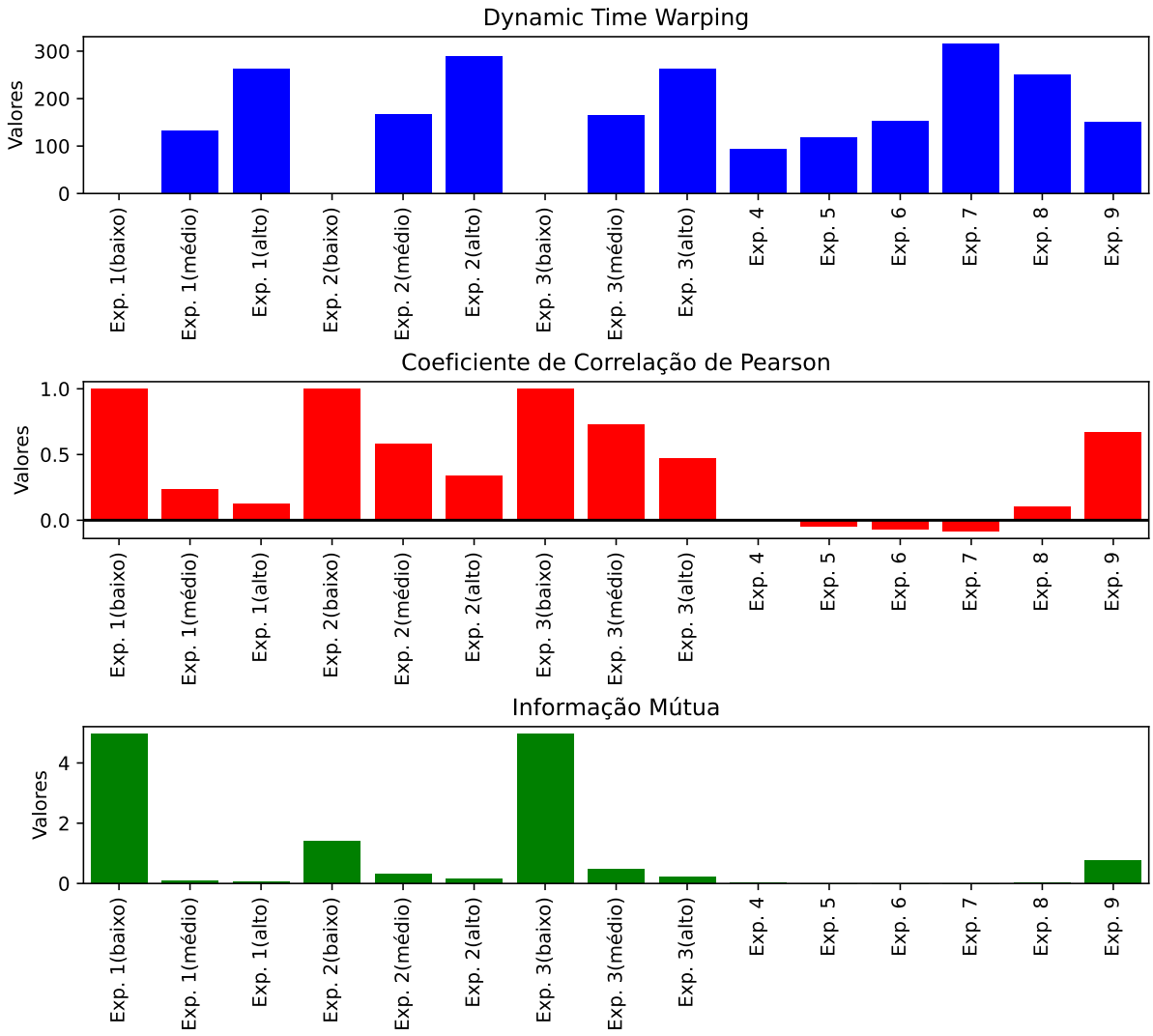



Figura 4.12 – Resultados gerais dos Experimentos de 1 a 9 das três métricas.

Para os experimentos feitos utilizando séries temporais reais dos *datasets* do *Kaggle* e suas versões ruídosas, o Coeficiente de Correlação de Pearson detecta semelhanças mesmo com ruídos altos, com exceção do Experimento 10.2, onde a métrica detecta com precisão os aumentos de intensidade de ruído. A Informação Mútua, assim como nos Experimentos anteriores de ruído, tende a diminuir a sensibilidade ao aumento de intensidade de ruído quando $\delta \approx 5$, não gerando novos valores de IM. A relação entre a intensidade do ruído e DTW continua quase linear, como nos experimentos anteriores, e com o maior custo entre eles. Há uma exceção no Experimento 10.4, onde é possível ver na Figura 4.16 que o tempo da Informação Mútua é maior do que os outros dois métodos, já que a série em questão é pequena em tamanho. Também é possível observar que há um grande desvio padrão obtido pelas métricas e algorítmos nesse Experimento.

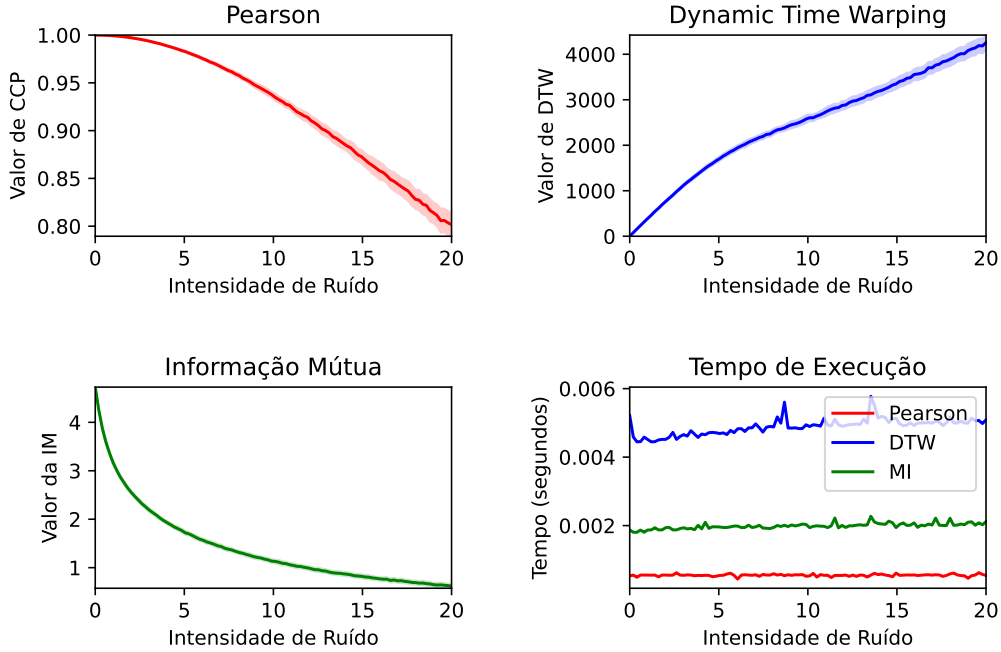



Figura 4.13 – Resultados do Experimento 10.1 - Produção de energia de uma elétrica com adição de ruídos.

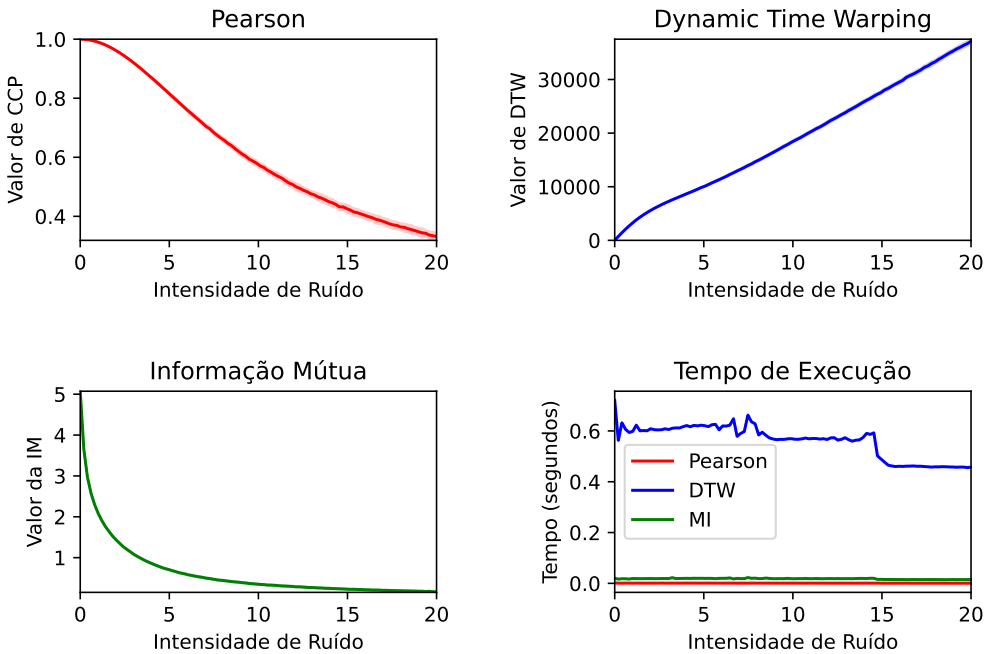



Figura 4.14 – Resultados do Experimento 10.2 - Temperaturas mínimas diárias ao longo dos anos com adição de ruídos.

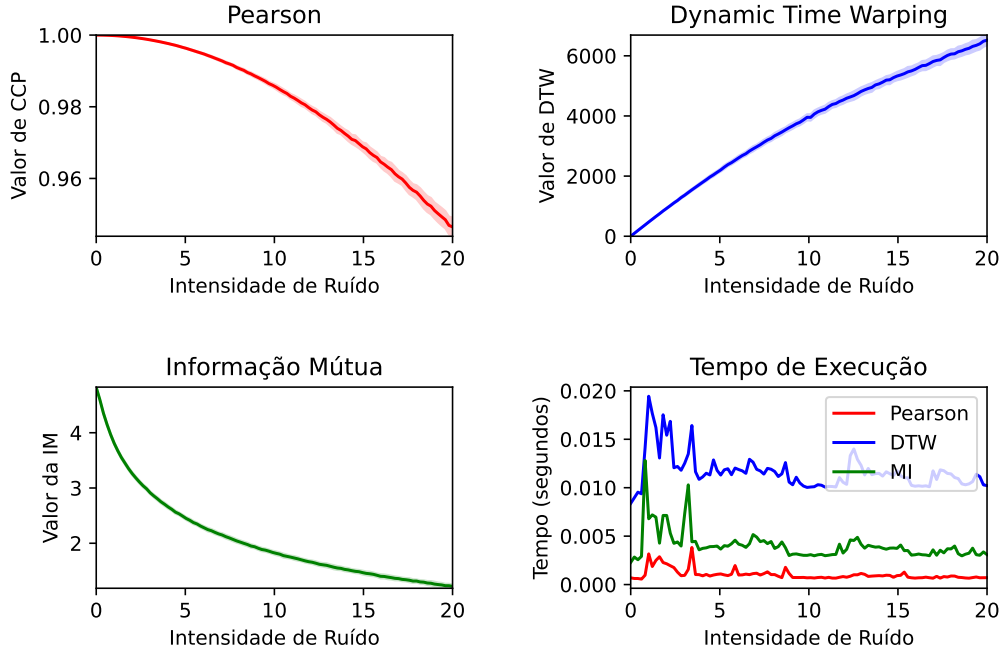



Figura 4.15 – Resultados do Experimento 10.3 - Produção mensal de cerveja na Australia com adição de ruídos.

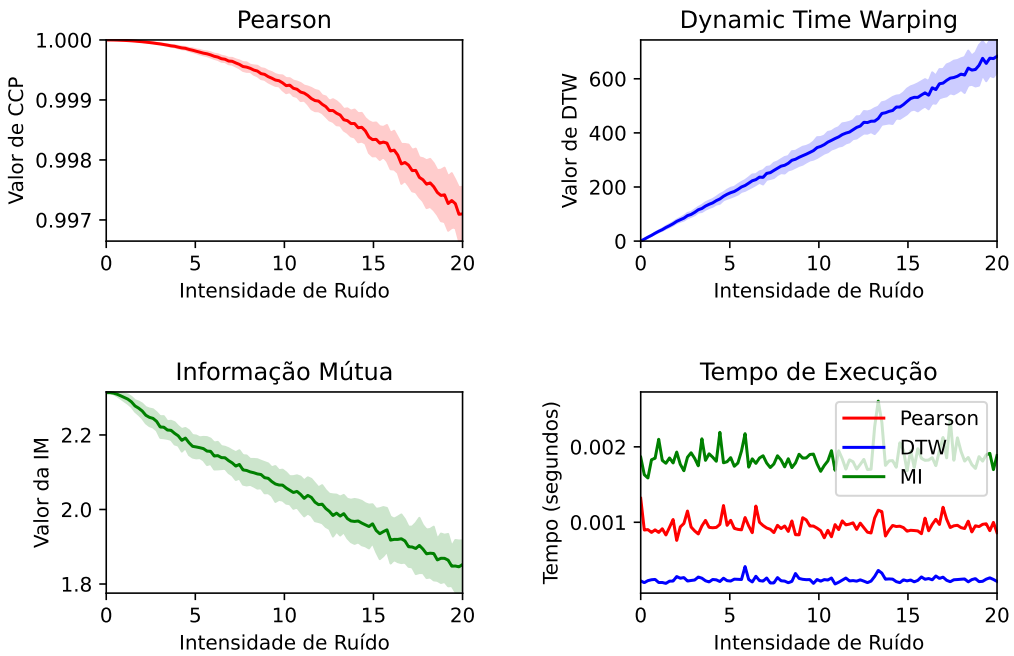



Figura 4.16 – Resultados do Experimento 10.4 - Venda mensal de shampoos ao longo de três anos com adição de ruídos.

A Figura 4.17 mostra os valores obtidos nos experimentos realizados com as séries do *dataset* do Kaggle. Como esses experimentos também envolviam aumento de ruído, no gráfico

demonstra os valores com ruído baixo ($\sigma = 0$), valor de ruído médio ($\sigma = 10$) e o valor máximo de ruído ($\sigma = 20$).

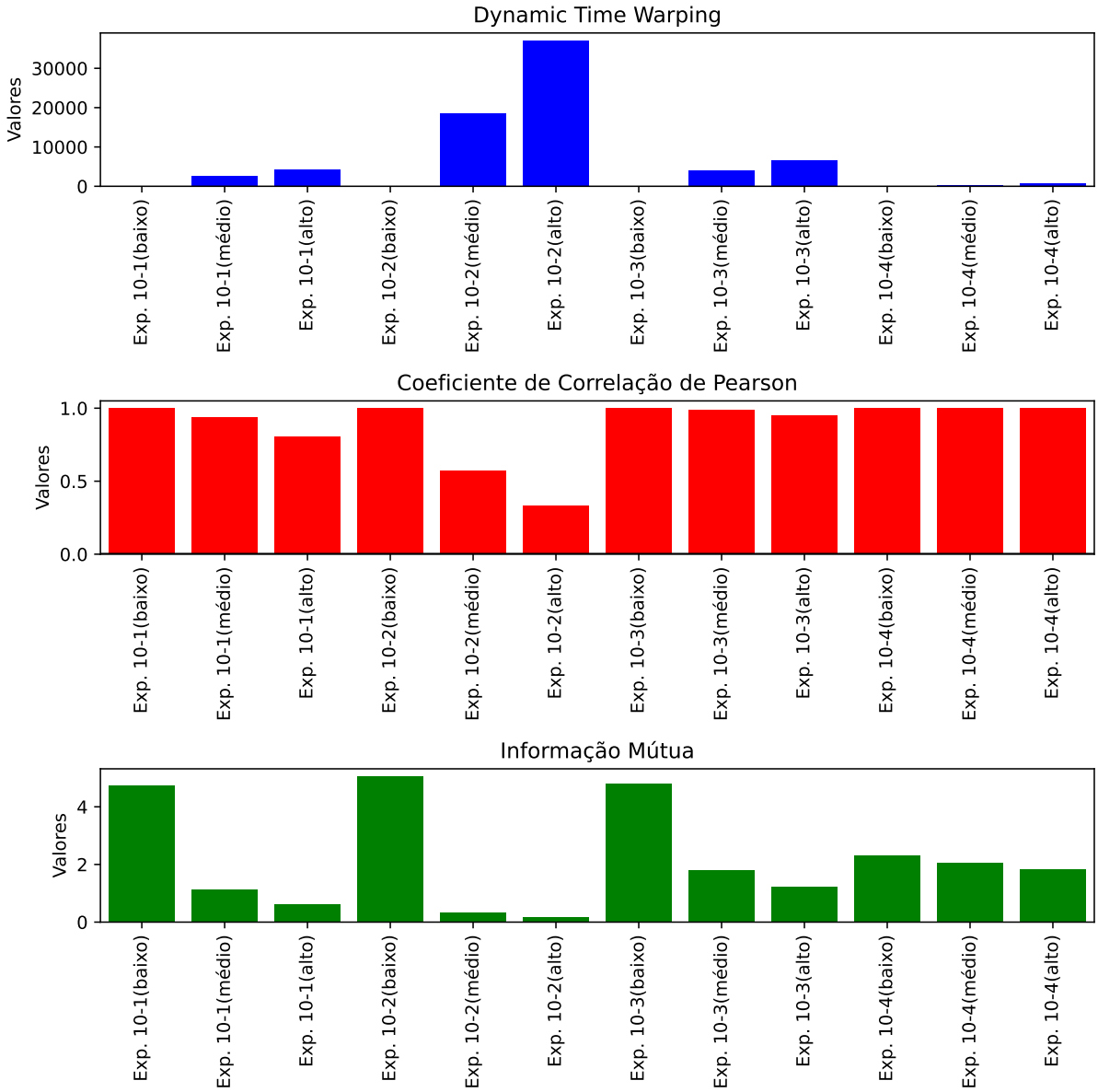



Figura 4.17 – Resultados gerais dos Experimentos de 10-1 ao 10-4 das três métricas.

De modo geral, na questão das séries sintéticas com adição de ruído, tanto a Informação Mútua quanto o Coeficiente de Correlação de Pearson detectam as alterações nas séries até certo ponto. Após isso, as métricas perdem a sensibilidade ao aumento de intensidade do ruído, o que não é ideal em casos onde precisa-se de uma precisão melhor, pois elas deixaram de distinguir as séries com suas versões ruidosas. Enquanto isso, o *Dynamic Time Warping* se manteve linear ao aumento da intensidade do ruído, captando bem as diferenças entre as séries. Quando as comparações de séries caóticas, o DTW captou semelhanças entre as séries comparadas que os outros métodos não conseguiram, apenas indicando que as séries não possuiam relações. Já

para as combinações de séries, as três métricas demonstraram um comportamento parecido nos experimentos, resultando em valores que indicavam as mesmas conclusões nos experimentos. Nos experimentos das séries de dados reais, CCP não conseguiu identificar as alterações com o aumento do ruído. Já a IM manteve o padrão dos outros experimentos com ruído, deixando de ser sensível à alterações a medida que a intensidade do ruído se aproxima de um certo valor. O DTW também se manteve de forma padrão nos experimentos, mantendo uma relação linear com o aumento do ruído.

# 5 Conclusão

Esta monografia teve como objetivo catalogar o comportamento das métricas e algorítmos voltados para comparação de séries temporais de diferentes tipos e seus respectivos tempos de processamento. Foram utilizadas séries sintéticas geradas tanto pelo Mapa Logístico quanto por uma função senoidal e séries de dados reais, obtidas de *datasets* do Kaggle.

O *Dynamic Time Warping* se mostrou como um bom algorítmo para detecção de similaridades a longa distância, sendo o mais indicado para buscar nuances não-lineares. Ele manteve uma relação linear com o aumento de intensidade do ruído, não perdendo sua sensibilidade a esse aumento, diferente das outras métricas. Também foi o melhor para comparar séries caóticas, identificando similaridades que as outras métricas não conseguiram detectar. Entretando, uma desvantagem significativa do DTW é o tempo elevado de execução ao lidar com séries temporais grandes, devido ao cálculo da matriz de custo mínimo. Além disso, os valores resultantes do DTW podem ser de difícil interpretação, exigindo a definição de um parâmetro de similaridade de referência para facilitar a análise dos resultados.

Enquanto isso, o Coeficiente de Correlação de Pearson, para as séries sintéticas, detectou as nuances com certa precisão, mas perde a sensibilidade quando o valor do ruído chega a certo ponto. Nas séries do Kaggle, ele não detectou diferenças grandes mesmo com intensidades de ruídos altas para os casos onde as séries apresentavam valores elevados, apesar de ainda ter a sensibilidade a alterações. Para a Informação Mútua, em todos os casos de ruído, após certo valor de ruído, ele não detecta com muita precisão as alterações na intensidade, não gerando novos valores, não refletindo as mudanças nas séries.

Levando em conta todo o estudo, é possível afirmar que cada métrica e algorítmo possuem seus pontos fortes e suas limitações. O DTW demonstrou que mantém uma relação linear com o aumento de ruído, evidenciando que ele captura bem às mudanças, apesar do seu alto tempo de processamento. Já o CCP se mostrou útil em casos onde as séries possuem valores numa certa faixas e ruídos distoantes desses valores, e teve baixo tempo de processamento.

## 5.1 Trabalhos Futuros

É necessário definir uma escala melhor para o DTW. Uma abordagem eficaz seria comparar a série temporal com uma versão embaralhada de si mesma. Espera-se que os valores obtido sejam uma espécie de paramêtro para uma análise mais precisa.

Uma outra continuação possível para esse trabalho seria a aplicação das séries temporais geradas para a criação de redes funcionais sintéticas para observar com maior precisão os comportamentos das métricas e fazer uma comparação direta dos resultados obtidos neste trabalho

com as redes criadas.

# Referências


ANTUNES, J. L. F.; CARDOSO, M. R. A. Uso da análise de séries temporais em estudos epidemiológicos. **Epidemiologia e Serviços de Saúde**, SciELO Brasil, v. 24, p. 565–576, 2015.

ARROYO, J.; ESPÍNOLA, R.; MATÉ, C. Different approaches to forecast interval time series: a comparison in finance. **Computational Economics**, Springer, v. 37, p. 169–191, 2011.

BROCKWELL, P. J.; DAVIS, R. A. **Introduction to time series and forecasting**. [S.l.]: Springer, 2002.

DAVIES, B. **Exploring Chaos: Theory And Experiment (Studies in Nonlinearity)**. [S.l.]: Westview Press, 2003. ISBN 978-0813341279.

DINIZ, I. d. S.

**Geração de redes funcionais a partir de séries temporais de um radar meteorológico**, Ouro Preto, 2022.

DONATE, J. P. *et al.* Time series forecasting by evolving artificial neural networks with genetic algorithms, differential evolution and estimation of distribution algorithm. **Neural Computing and Applications**, Springer, v. 22, p. 11–20, 2013.

DONGES, J. F. *et al.* Complex networks in climate dynamics: Comparing linear and nonlinear network construction methods. **The European Physical Journal Special Topics**, Springer, v. 174, n. 1, p. 157–179, 2009.

FERREIRA, L. N. *et al.* The effect of time series distance functions on functional climate networks. **The European Physical Journal Special Topics**, Springer, v. 230, p. 2973–2998, 2021.

HUTTER, M. Distribution of mutual information. **Advances in neural information processing systems**, v. 14, 2001.

JORGE, A. A. *et al.* Geographical complex networks applied to describe meteorological data. **Proceedings XXI GEOINFO**, v. 21, 2020.

LHERMITTE, S. *et al.* A comparison of time series similarity measures for classification and change detection of ecosystem dynamics. **Remote sensing of environment**, Elsevier, v. 115, n. 12, p. 3129–3152, 2011.

LU, B. *et al.* Constrained selective dynamic time warping of trajectories in three dimensional batch data. **Chemometrics and Intelligent Laboratory Systems**, Elsevier, v. 159, p. 138–150, 2016.

MAY, R. M. Simple mathematical models with very complicated dynamics. **Nature**, Springer, v. 261, p. 459–467, 1976.

RAHMAN, M. S. *et al.* **Basic graph theory**. [S.l.]: Springer, 2017. v. 9.

SAKOE, H.; CHIBA, S. Dynamic programming algorithm optimization for spoken word recognition. **IEEE transactions on acoustics, speech, and signal processing**, IEEE, v. 26, n. 1, p. 43–49, 1978.

SCHOBER, P.; BOER, C.; SCHWARTE, L. A. Correlation coefficients: appropriate use and interpretation. **Anesthesia & analgesia**, Wolters Kluwer, v. 126, n. 5, p. 1763–1768, 2018.

SILVA, V. de P. **TSAPI: Uma Aplicação Distribuída Para Comparação De Séries Temporais**. 55 p. Especializção — Ciência da Computação, Universidade Federal de Ouro Preto, Ouro Preto, 2022.

SILVA, V. de P. *et al.* Comparando algoritmos para conversão de séries temporais em grafos. In: **Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 9, n. 1**. [S.l.: s.n.], 2021.

VALéRIO, D. M.

**Correlações entre redes funcionais de dados de chuva com propriedades do solo.**, Ouro Preto, 2024.

WANG, X. *et al.* Experimental comparison of representation methods and distance measures for time series data. **Data Mining and Knowledge Discovery**, Springer, v. 26, p. 275–309, 2013.