UNIVERSIDADE FEDERAL DE MINAS GERAIS

INSTITUTO DE CIÊNCIAS EXATAS

DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

GUILHERME MENDES DE OLIVEIRA

CARACTERIZAÇÃO E MODELAGEM POR APRENDIZADO DE MÁQUINA EM SISTEMAS CAÓTICOS

Belo Horizonte

2023

GUILHERME MENDES DE OLIVEIRA

CARACTERIZAÇÃO E MODELAGEM POR APRENDIZADO DE MÁQUINA EM SISTEMAS CAÓTICOS

Monografia apresentada ao Curso de Sistemas de Informação do Departamento de Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do grau de bacharelado em Sistemas de Informação.

Orientador: Prof. Dr. Wagner Meira Junior.

Belo Horizonte

2023

RESUMO

Este estudo visa realizar o estudo de sistemas dinâmicos com ênfase na dinâmica populacional, orientado a partir da atividade de caracterização de uma base de dados reais como originada de um sistema dinâmico.

Serão discutidas as principais propriedades apontadas na literatura que poderão ser usadas como pistas para validar a hipótese de caracterização, assim como os principais modelos matemáticos que tratam da dinâmica populacional que serão utilizados também como forma de identificar possíveis equações geradoras dos dados.

Além disso, serão utilizadas técnicas complementares para a atividade de caracterização, dessa forma, além da discussão do referencial teórico haverá a discussão dos resultados da aplicação prática no conjunto de dados reais.

Ademais, espera-se apresentar conclusões fundamentadas na revisão teórica e na análise empírica dos resultados, contribuindo para o avanço do conhecimento de caracterização de dinâmica caótica em conjuntos de dados reais.

Palavras-chave: Modelagem de dados, caracterização de sistemas dinâmicos, análise de séries temporais.

ABSTRACT

This study aims to conduct research on dynamic systems with an emphasis on population dynamics, based on the characterization of a real database originating from a dynamic system.

The main properties highlighted in the literature, which can serve as clues to validate the characterization hypothesis, will be discussed. Additionally, the main mathematical models addressing population dynamics will be used to identify potential equations that generate the data.

Furthermore, complementary techniques will be employed for the characterization activity. In addition to the theoretical framework, the discussion will also include the practical application results on the real dataset.

Moreover, the study aims to present conclusions based on the theoretical review and empirical analysis of the results, contributing to the advancement of knowledge in the characterization of chaotic dynamics in real datasets.

Keywords: Data modeling, dynamic systems characterization, time series analysis.

LISTA DE FIGURAS

FIGURA 1 Exemplo de plotagem do Atrator de Lorenz…………………………………….…9

FIGURA 2 Exemplo de Atrator de ciclo no modelo Lotka-Volterra…………………………13

FIGURA 3 Evidência do fenômeno caótico no Mapa Logístico obtido com pequena alteração nas condições iniciais (X) e o mesmo valor de r (3.99)..13

FIGURA 4 Exemplo de Bifurcação no Mapa Logístico obtido com a variação do parâmetro r para o mesmo valor de X……………………………………………………………………...14

FIGURA 5 Histogramas dos atributos da base de dados……………………………………..23

FIGURA 6 Linha da matriz de correlação entre os atributos da base de dados……………….24

FIGURA 7 Plotagem do resultado do ajuste à curva - Mapa Logístico……………………….25

FIGURA 8 Plotagem do resultado do ajuste à curva - Modelo de Gompertz………………….26

FIGURA 9 Plotagem do resultado do ajuste à curva - Modelo de Bertalanffy-Richards……..27

FIGURA 10 Plotagem do valor da autocorrelação em função do 'lag'...................................28

FIGURA 11 Plotagem da série temporal para avaliação de tendência……………………….28

FIGURA 12 Teste de estacionariedade - Método de Janelas Móveis (Média) Iteração 1…….29

FIGURA 13 Teste de estacionariedade -Método de Janelas Móveis (Variância) Iteração 1…..29

FIGURA 14 Aplicação do Teorema de Takens……………………………………………….30

SUMÁRIO

RESUMO 3
ABSTRACT 4
LISTA DE FIGURAS 5
SUMÁRIO 6
1 Introdução 9
2 Sistemas Dinâmicos 11
2.1 Equilíbrio 12
2.2 Estabilidade 13
2.3 Atratores 14
2.4 Caos 15
3 Modelos Matemáticos de Dinâmica Populacional 16
3.2 Dinâmica Populacional 16
3.1 Modelo Malthusiano 17
3.1 Modelos de Verhulst 17
3.2 Modelo de Gompertz 18
3.3 Modelo de Bertalanffy-Richards 19
4 Formas de Caracterização 20
4.1: Autocorrelação 21
4.2: Estacionariedade 21
4.3: Teorema de Takens 22
4.4 Expoente de Lyapunov 22
5 Análise Exploratória e Experimental 24
5.1 Análise Exploratória 24
5.2 Caracterização da Base de Dados 26
5.3 Análise Experimental 29
5.3.1: Autocorrelação 29
5.3.2: Estacionariedade 30
5.3.3: Teorema de Takens 32
5.3.4: Expoente de Lyapunov 33
6 Conclusão 34
7 Referências Bibliográficas 35

1 Introdução

Os estudos dos sistemas dinâmicos são fundamentais para compreender o comportamento complexo e a evolução de fenômenos da natureza ao longo do tempo e são essenciais para a modelagem e predição em diversas áreas do conhecimento, como biologia, ecologia, economia, demografia, dentre outras. A escolha do tema é fundamentada na ideia de ser capaz de estabelecer a caracterização de uma base de dados que pertence a uma área de estudo dos sistemas dinâmicos, no caso a dinâmica populacional.

Sendo possível, posteriormente, aplicar técnicas de predição mais alinhadas aos sistemas dinâmicos, essa atividade inteira será dividida em duas etapas, para este trabalho de Monografia I será restrita à caracterização e na Monografia II a parte de aplicação das técnicas de aprendizado de máquina para predições.

Alguns modelos matemáticos são muito utilizados na temática da dinâmica populacional com o intuito de realizar o entendimento e a predição dessas populações, são eles o modelo Malthusiano, o modelo de Verhulst, o modelo de Gompertz e o Modelo de Bertalanffy-Richards. Cada um desses modelos apresenta características específicas, mas contribuíram e contribuem para a compreensão do crescimento e evolução da dinâmica das populações.

A série histórica do tamanho da população do Brasil será considerada a variável de interesse principal com o objetivo de tornar a atividade um pouco menos complexa, do que seria considerando uma base multidimensional. Embora existam outros fatores, discutidos nas teorias demográficas, que impactam diretamente a taxa de crescimento e, consequentemente, o tamanho da população.

Posteriormente, poderão ser consideradas outras variáveis relevantes para uma análise mais completa na aplicação das técnicas de aprendizado de máquina para predição na Monografia II.

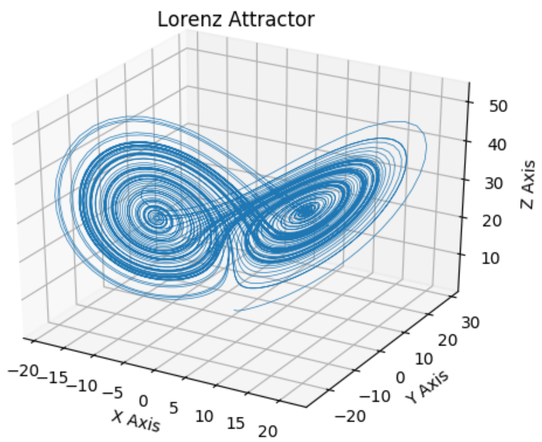
Para caracterizar os dados como um sistema dinâmico além da abordagem de aproximação aos modelos matemáticos mencionados, serão utilizadas técnicas específicas para investigar o potencial caráter dinâmico e revelar padrões e estruturas do conjunto de dados sendo elas a autocorrelação, o Teorema de Takens e o espaço de fases.

Este trabalho está organizado em cinco seções além desta introdutória. Na seção 2, serão apresentados os fundamentos teóricos dos sistemas dinâmicos, destacando os principais conceitos. Na seção 3, serão discutidos os modelos matemáticos mencionados, juntamente com suas premissas. A seção 4 apresentará as técnicas pesquisadas para a caracterização, na Seção 5 abordará a tratativa da caracterização com os testes e a utilização dos modelos matemáticos no conjunto de dados juntamente com um breve exploratório do mesmo. Por fim, na seção 6, serão apresentadas as conclusões do trabalho.

2 Sistemas Dinâmicos

Ao longo da história vários cientistas trabalharam para tentar estender os avanços obtidos por Poincaré que desenvolveu uma estrutura teórica para analisar e compreender os sistemas dinâmicos, mas foi Lorenz quem conseguiu estabelecer uma perspectiva prática do caso mostrando como pequenas perturbações podem ter grandes efeitos nesses sistemas. [1,2]

FIGURA 1 Exemplo de plotagem do Atrator de Lorenz



Fonte: Matplotblib, 2002. Disponível em: Matplot.org

Há de se estabelecer que nos sistemas dinâmicos existem duas abordagens, os mapas iterativos e as equações diferenciais. [1]

Equação iterativa: Diz respeito a evolução do sistema considerando o tempo de forma discreta e expressam o instante *n+1* como uma relação recorrência com o instante *n*.

$$x_{n+1} = f(x_n)$$

(2.1)

Equação diferencial: a evolução do sistema considerando o tempo de forma contínua envolve derivadas, representando as taxas de mudança das variáveis dependentes x em relação às variáveis independentes *t*.

$$\frac{dx}{dt} = g(x, t)$$

(2.2)

Conforme os estudos de Strogatz (1994) e Alligood et. al. (1996) existem alguns conceitos no que diz respeito a observação, análise e compreensão de sistemas dinâmicos e seu comportamento ao longo do tempo.

2.1 Equilíbrio

O estado de equilíbrio em sistemas dinâmicos diz respeito a manutenção do estado, isto é, quando há um equilíbrio entre os fatores que induzem crescimento ou redução levando a variável a ter um valor constante ao longo do tempo. [1,2]

Se imaginarmos uma espécie na natureza, o estado de equilíbrio da mesma ocorre quando a taxa de natalidade e a mortalidade se igualam fazendo com que o tamanho da população se mantenha constante ao longo do tempo. Matematicamente, o equilíbrio ocorre quando a derivada da equação diferencial que descreve o sistema em relação ao tempo é igual a zero para os valores das variáveis do sistema, conforme exemplo abaixo da equação logística. [2]

$$\frac{dP}{dt} = r \cdot P \cdot \left(1 - \frac{P}{K}\right)$$

(2.3)

Supondo que r seja igual a 0,1 e K=100 podemos calcular o equilíbrio da seguinte forma

$$\frac{dP}{dt} = 0,1 \cdot P \cdot \left(1 - \frac{P}{100}\right) = 0$$

(2.4)

Quando a população é zero ao longo do tempo ela nem cresce nem diminui, estabiliza em zero e quando ela atinge a capacidade de suporte do ambiente, o crescimento se iguala a zero e ao longo do tempo a população se estabiliza em zero.

2.2 Estabilidade

A estabilidade é um conceito que avalia a dinâmica do sistema em relação a perturbações que atuam em suas variáveis no que diz respeito ao estado de equilíbrio. [1,2]

Se imaginarmos a mesma espécie anterior em uma condição inicial de equilíbrio e que por algum fator, como por exemplo, a ocorrência de predação influenciando na sua taxa de mortalidade a estabilidade pode ser analisada.

Caso essa mortalidade propicie, por exemplo, um aumento na taxa de natalidade e que essa relação de predação volte a ser constante no futuro, a perturbação não foi suficiente para que o sistema se afastasse do equilíbrio, portanto há estabilidade nesse sistema hipotético.

Vale ressaltar que o modelo logístico apresentado anteriormente tem a premissa que a taxa de crescimento *r* é constante, portanto para esse novo cenário com a presença de um predador tem-se o modelo Lotka-Volterra. que considera a interação entre essas duas espécies.

$$\frac{dR}{dt} = aR - bRP$$
$$\frac{dP}{dt} = cRP - dP$$

(2.5)

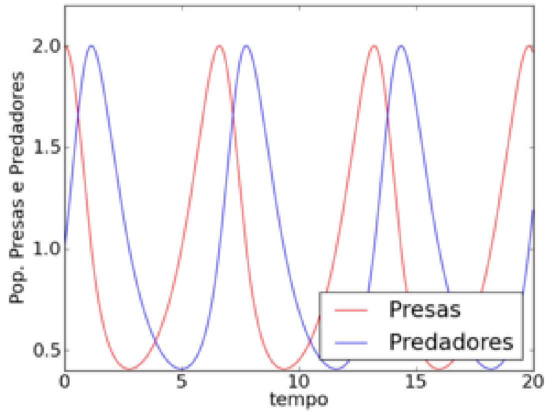
Tendo *R* como o valor da população de presas , P o valor da população de predadores. Os parâmetros *a, b, c e d* são valores das taxas de crescimento e mortalidade das populações

2.3 Atratores

Atratores são valores para os quais as variáveis do sistema tendem a assumir quando o sistema é perturbado, o estado resultante pode apontar comportamentos de característica recorrente ou de estabilidade. [1,2]

No caso do modelo de Lotka-Volterra pode-se observar, conforme a Figura 2, a presença de um atrator que gera um comportamento cíclico na evolução das populações de presas e predadores, nesse caso as populações para esse atrator convergem para determinados valores ao longo do tempo.

FIGURA 2 Exemplo de Atrator de ciclo no modelo Lotka-Volterra.

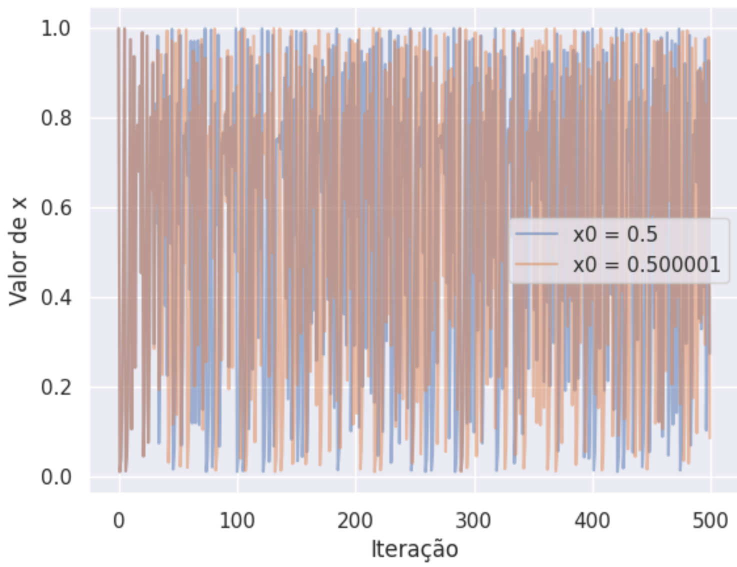


Fonte: Wikipedia,2013. Disponível em: Wikipedia.org

2.4 Caos

O caos, conforme Figura 3, é um fenômeno inerente a alguns sistemas dinâmicos que além da forte sensibilidade às condições iniciais apresentam trajetórias complexas em ciclos determinados que não se repetem tornando o sistema complexo e imprevisível. [1,2]

FIGURA 3 Evidência do fenômeno caótico no Mapa Logístico obtido com pequena alteração nas condições iniciais (*X*) e o mesmo valor de *r* (3.99).



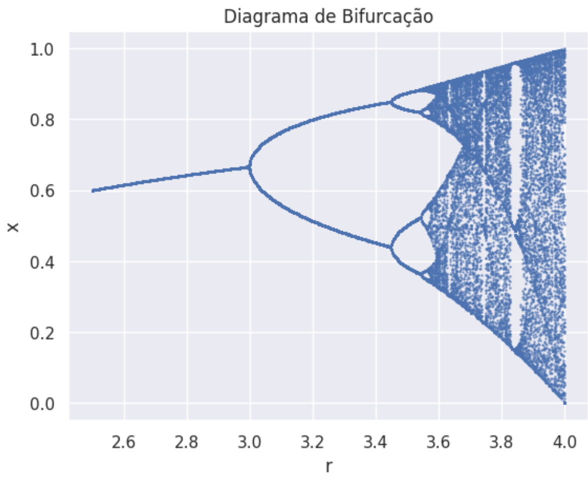
Fonte: Elaborado pelo autor.

No sistema dinâmico representado pelos mapas logísticos esse comportamento caótico é manifestado quando o parâmetro r atinge um valor crítico, produzindo as bifurcações.

A bifurcação é uma característica presente em alguns sistemas dinâmicos, na qual ocorre uma mudança no comportamento do sistema à medida que os parâmetros das equações são variados. Essa variação dos parâmetros é uma ferramenta analítica que nos permite estudar os diferentes regimes de comportamento que um sistema pode apresentar dado um mesmo valor inicial.

No exemplo da Figura 4, utiliza-se o mapa logístico, onde dado o mesmo valor inicial, variando o parâmetro *r*, desconsiderando as primeiras 100 iterações para que o sistema alcançasse equilíbrio, pode-se observar os valores possíveis que o sistema assume para cada valor de *r*.

FIGURA 4 Exemplo de Bifurcação no Mapa Logístico obtido com a variação do parâmetro *r* para o mesmo valor de *X*.



Fonte: Elaborado pelo autor.

3 Modelos Matemáticos de Dinâmica Populacional

3.2 Dinâmica Populacional

Segundo Lee (1987) os estudos da dinâmica populacional são reflexo da interação entre diversos fatores demográficos como as taxas de nascimento e mortalidade, o envelhecimento da população, além de fatores culturais, socioeconômicos e tecnológicos.

Dessa forma os estudos demográficos em relação aos fatores que impactam no crescimento populacional foram fundamentais para que os modelos matemáticos evoluíssem e fossem importantes na análise do comportamento da população ao longo da história.

Em seus estudos Lee (2003) apresenta como ao longo do tempo a composição e os fatores estruturantes da população foram mudando, tal fenômeno chamado de transição demográfica reflete também como os modelos matemáticos foram evoluindo para acompanhar os comportamentos observados pelos estudos demográficos.

3.1 Modelo Malthusiano

Um dos primeiros pesquisadores na história a apresentar modelos matemáticos que envolvesse a dinâmica populacional foi Thomas Malthus, para ele a população humana tende a crescer exponencialmente, já os recursos disponíveis tendem a crescer linearmente sendo assim o crescimento populacional eventualmente ultrapassaria a capacidade de produção de alimentos.

Dessa forma, Malthus considerava que a taxa de crescimento populacional era uma constante que dependeria apenas da simples diferença observada entre natalidade e mortalidade. [5, 6]

$$\frac{dP(t)}{dt} = r \cdot P(t) \tag{3.1}$$

$$r = T_{\text{natalidade}} - T_{\text{mortalidade}} \tag{3.2}$$

3.1 Modelos de Verhulst

O modelo de Verhulst, ou modelo logístico, é uma extensão do modelo malthusiano proposto por Pierre-François Verhulst. Verhulst estendeu a ideia da limitação dos recursos no ambiente conforme dito por Malthus e introduziu um fator limitante no crescimento por parte do ambiente adicional na equação diferencial.

No modelo de Verhulst, a taxa de crescimento populacional agora depende da interação entre a população e a capacidade máxima do ambiente, partindo da premissa que o tamanho da população interfere na competição por recursos e também no comportamento reprodutivo. [2,6]

A equação diferencial do modelo logístico é dada por:

$$\frac{dP}{dt} = r \cdot P\left(1 - \frac{P}{K}\right) \tag{3.3}$$

- P é o tamanho da população;
- r é a taxa de crescimento populacional;
- K é a quantidade máxima de indivíduos que o ambiente pode sustentar, conhecido também como a capacidade de suporte;
- $\left(1 - \frac{P}{K}\right)$ é o fator de crescimento

Portanto o fator de crescimento implica que o crescimento populacional diminui à medida que a população se aproxima da capacidade de suporte resultando em um equilíbrio populacional em torno de *K*.

3.2 Modelo de Gompertz

O modelo de Gompertz também pode ser considerado uma extensão do modelo Malthusiano, proposto por Benjamin Gompertz. O modelo de Gompertz também implementa a ideia de um limite máximo para o número de indivíduos suportados pelo ambiente e do relacionamento entre a taxa de crescimento e o tamanho da população. [6]

No entanto, diferente do modelo de Verhulst, considera que o crescimento populacional se torna cada vez mais lento ao longo do tempo à medida que se aproxima da capacidade de suporte do ambiente. [6]

A equação diferencial do modelo de Gompertz é dada por:

$$\frac{dP}{dt} = -r_0 \cdot P \cdot \ln\left(\frac{P}{K}\right) \tag{3.4}$$

- P é o tamanho da população;
- r_0 é a taxa de crescimento populacional inicial;
- K é a quantidade máxima de indivíduos que o ambiente pode sustentar, conhecido também como a capacidade de suporte;
- $\ln\left(\frac{P}{K}\right)$ é o fator de crescimento

3.3 Modelo de Bertalanffy-Richards

O modelo de Bertalanffy-Richards pode ser considerado uma extensão dos modelos anteriores, além de incorporar os conceitos de um limite máximo para o número de indivíduos suportados pelo ambiente e do relacionamento entre a taxa de crescimento e o tamanho da população, apresenta uma maior flexibilidade na forma da curva de crescimento populacional, com parâmetros que controlam a suavidade ou acentuação da curva. [6]

A equação diferencial do modelo de Bertalanffy-Richards é dada por:

$$\frac{dP}{dt} = (a \cdot P^{\beta}) - (b \cdot P)$$

(3.5)

$$a = -\frac{r_0}{q \cdot K^q}$$

(3.6)

$$b = -\frac{r_0}{q}$$

(3.7)

$$\beta = q + 1$$

(3.8)

- P é o tamanho da população;
- r_0 é a taxa de crescimento populacional inicial;
- K é a quantidade máxima de indivíduos que o ambiente pode sustentar, conhecido também como a capacidade de suporte;
- β e q não possuem um significado físico no aspecto macroscópico, conforme Ribeiro (2016) esses fatores estão relacionados a um caráter microscópico de interação entre os "organismos". No entanto para valor de $\beta = 1$ temos o modelo de Gompertz e para $\beta = 2$ temos o modelo de Verhulst.

4 Formas de Caracterização

Para BRAGA et al. (2015) os sistemas complexos não possuem uma definição precisa a respeito, o que se tem é a utilização de determinadas propriedades como indícios de caracterização de um sistema como tal.

> Embora saber se o sistema apresenta ou não dinâmica caótica seja ainda inconclusivo, uma vez que o conceito de dinâmica caótica está ligado a se demonstrar a existência de certas propriedades em um sistema, e desejando-se um íındice de confiança de 100%, ao longo dos anos, técnicas foram desenvolvidas permitindo que um experimentalista tenha "fortes evidências" da presença da dinâmica caótica em seu sistema sob análise." (Santos, 2010).

Nesse contexto, faz-se necessário a busca por mais formas de apontar em um conjunto de dados presença de características que possam indicar que os dados são oriundos de um sistema dinâmico.

4.1: Autocorrelação

Segundo BRAGA et al. (2015) "Medidas de autocorrelação indicam o quão correlacionados estão os termos de uma série temporal. Com medidas como essa, pode-se intuir muitos aspectos diretamente ligados à dinâmica do sistema sob investigação".

Dessa forma, a autocorrelação pode apontar a dependência de um valor da variável dependente no tempo *t* com um dos valores anteriores e assim indicar a presença de uma das propriedades do sistema dinâmico que é a sensibilidade aos valores iniciais.

4.2: Estacionariedade

A estacionariedade é definida segundo Feijó (2015) como a manutenção das medidas estatísticas de uma série temporal ao longo do tempo. As observações não apresentam padrões de crescimento, sazonalidade e ciclos.

Santos (2010) aplica o teste denominado Janelas Móveis, que busca avaliar os critérios da estacionariedade ao longo da série a partir de um subconjunto de dados e intervalo de passo parametrizado, o método busca verificar se a média e a variância da janela móvel (parcial) está numa distância de até um desvio padrão da média e variância total respectivamente.

O resultado aplicado a uma série do sistema de Lorenz aponta estacionariedade na série temporal experimental, isto é, o sistema mantém o seu comportamento e características não contendo grandes irregularidades para nenhum subconjunto.

4.3: Teorema de Takens

O Teorema de Takens, desenvolvido por Floris Takens, é uma técnica que permite mapear uma série temporal unidimensional em um espaço de estados de mais dimensões podendo revelar a estrutura referente ao sistema e assim através do atrator gerado pela técnica observar a presença de comportamento não linear e/ou caótico.

Para a reconstrução do espaço de estados é necessário que haja a escolha de um "atraso" a ser aplicado na série temporal e assim gerar o mapeamento.

Segundo Campanharo (2006) "Um método utilizado frequentemente é o método das coordenadas de atraso temporal [16], onde cada ponto no espaço de fase é uma n-tupla de consecutivos valores de uma série: $(X(t), X(t+\tau), X(t+2\tau), \ldots, X(t+(m-1)\tau))$. O tempo τ e algum múltiplo do espaçamento Δt entre os pontos da série temporal ".

Para escolha dessa medida de atraso Campanharo (2006) argumenta que para um conjunto finito de observações τ não pode ser definida arbitrariamente e a escolha é fundamental para que a reconstrução seja assertiva.

Uma das técnicas corresponde a utilizar a autocorrelação para a escolha da medida de atraso, observando os picos ou quedas significativas na plotagem da métrica, em suma o primeiro grande pico marca a transição do comportamento da trajetória.

4.4 Expoente de Lyapunov

O expoente de Lyapunov mede a sensibilidade do sistema às condições iniciais e permite verificar a dinâmica caótica ou não linear do sistema.

Se o sistema apresentar comportamento caótico o expoente de Lyapunov será positivo. Por outro lado, se o sistema apresentar comportamento estável ou regular, o expoente de Lyapunov será igual ou próximo de zero.

Campanharo (2006) diz que o método de Wolf utiliza média da divergência entre pontos de trajetórias próximas, obtidas também com a abordagem de reconstrução do espaço de fases discutido no subcapítulo anterior, para estimar os expoentes da série temporal sendo uma forma de medir o grau de sensibilidade do sistema.

5 Análise Exploratória e Experimental

5.1 Análise Exploratória

A base de dados foi extraída do repositório *World Bank Open Data* do Banco Mundial, por se tratar de um estudo em que o tamanho da série temporal seria um fator relevante, a base de dados contendo o tamanho populacional e outros dados socioeconômicos apresentou maior quantidade de registros que a localizada no site do IBGE. Mesmo não sendo o órgão oficial brasileiro a base do Banco Mundial é alimentada a partir dos registros dos órgãos governamentais e organizações internacionais em cada país.

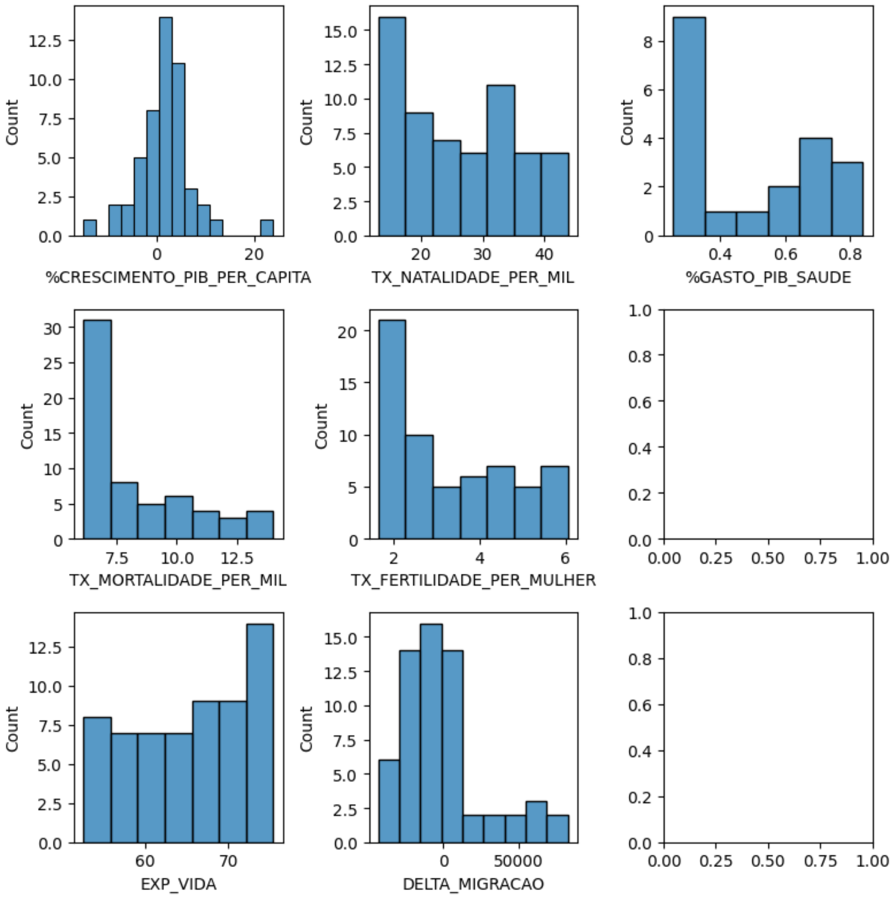
A base contém sete atributos levantados, além do tamanho da população, que poderiam interferir no crescimento populacional, são eles: Percentual de Crescimento do PIB por Habitante, Taxa de Natalidade por Mil Habitantes, Percentual de Gasto do PIB em Saúde, Taxa de Mortalidade por Mil Habitantes, Taxa de Fertilidade por Mulher, Expectativa de Vida ao Nascer e o Delta de Migração (Imigração - Emigração).

Os atributos Percentual de Crescimento do PIB por Habitante e Percentual de Gasto do PIB em Saúde apresentaram valores faltantes para doze e quarenta e dois registros respectivamente, os demais com exceção do tamanho da população apresentaram um registro faltante.

Para ambos os casos o tratamento foi substituir a ausência desses dados pelo valor da mediana a fim de preservar a distribuição, uma vez que a média é sensível a valores extremos e eliminar os registros iria reduzir muito o tamanho da base de dados.

Analisando a distribuição dos dados observa-se certa assimetria em alguns atributos como nas taxas de natalidade, mortalidade, fertilidade e no percentual de gasto em saúde, essa assimetria a esquerda indica a presença de uma maior quantidade de valores baixos para essas taxas.

FIGURA 5 Histogramas dos atributos da base de dados.



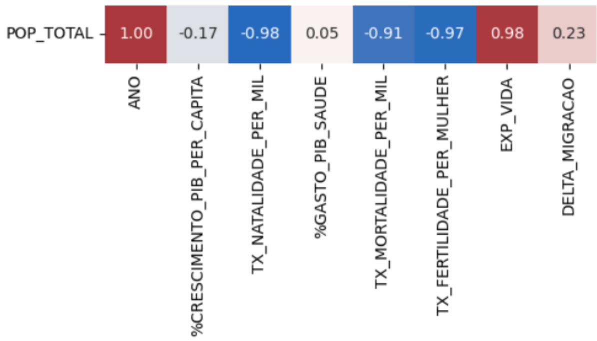
Fonte: Elaborado pelo autor.

Ao analisar a correlação observa-se que o tamanho da população possui uma relação forte e direta com o ano e com a expectativa de vida, indicando uma relação próxima da linear, observa-se também uma relação inversa e forte com as taxas de natalidade, mortalidade e fertilidade.

Esse fator de diminuição nas taxas que compõem o crescimento populacional à medida que a população cresce vai de encontro com as ideias discutidas no capítulo 3.

Os demais atributos apresentam uma relação fraca indicando que pode existir uma relação não linear para o delta de migração e para o crescimento do PIB por habitante e o percentual de gasto do PIB com saúde. Existe a possibilidade do tratamento dos dados faltantes terem influência na relação dos atributos , a substituição pela mediana pode ter impactado na variabilidade dos dados.

FIGURA 6 Linha da matriz de correlação entre os atributos da base de dados.



Fonte: Elaborado pelo autor.

5.2 Caracterização da Base de Dados

Nesta etapa o objetivo era coletar evidências de que o conjunto de dados pode ser considerado como oriundo de um sistema dinâmico. Nas pesquisas iniciais pretendia-se utilizar os atributos socioeconômicos da base de dados como a composição da taxa de crescimento para ser utilizado em um modelo matemático escolhido, inicialmente o mapa logístico, que trata da dinâmica populacional.

Após a definição desse parâmetro, iterar o mapa e comparar a curva gerada com a curva real da série temporal.

Para tal atividade foi aplicada a técnica de Análise de Componentes Principais a fim de selecionar os atributos que garantem o maior grau de explicabilidade da variância do conjunto de dados.

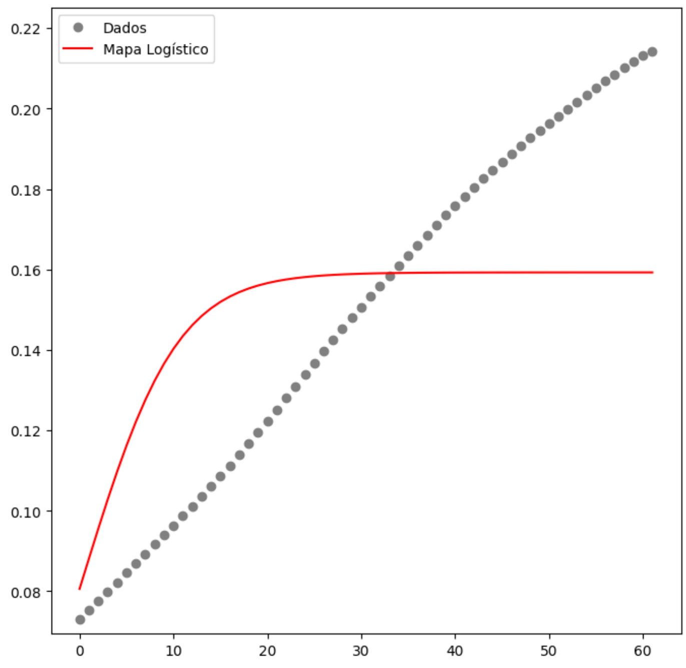
No entanto, ao utilizar a seleção dos atributos como forma de obter um valor para a taxa de crescimento através da média dos valores normalizados, ponderada pela importância na explicação da variância, obtida pela técnica de Análise de Componentes

Principais, não gerou uma curva que se ajustasse bem aos dados.

A partir disso a tarefa se tornou uma atividade de ajuste de curva e um problema de otimização de parâmetros, para tal atividade além do modelo pensado inicialmente (mapa logístico) foram selecionados outros modelos matemáticos encontrados através de pesquisas.

Foi realizado o ajuste de curva para o mapa logístico e as equações diferenciais do Modelo de Gompertz e do Modelo de Bertalanffy-Richards utilizando todo o conjunto de dados[1], uma vez que a intenção é apenas validar o quão bem os modelos teóricos definem bem o conjunto de dados, em nenhum dos casos o ajuste se deu de forma satisfatória.

FIGURA 7 Plotagem do resultado do ajuste à curva - Mapa Logístico.



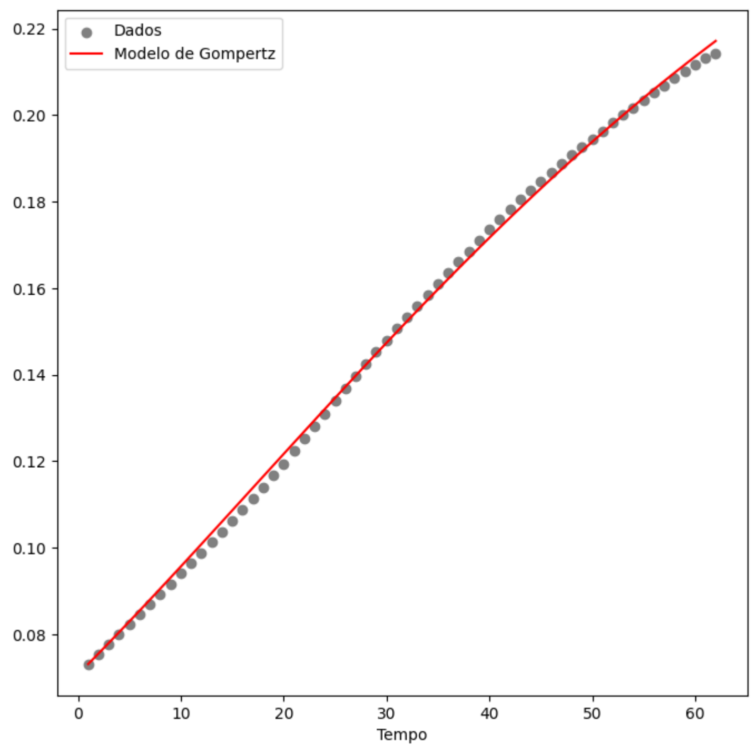
Fonte: Elaborado pelo autor.

Visualmente é perceptível que o ajuste da curva do mapa logístico não foi satisfatório,

[1] O tamanho da população foi ajustado em uma escala de bilhões de pessoas, em virtude do Fator de crescimento

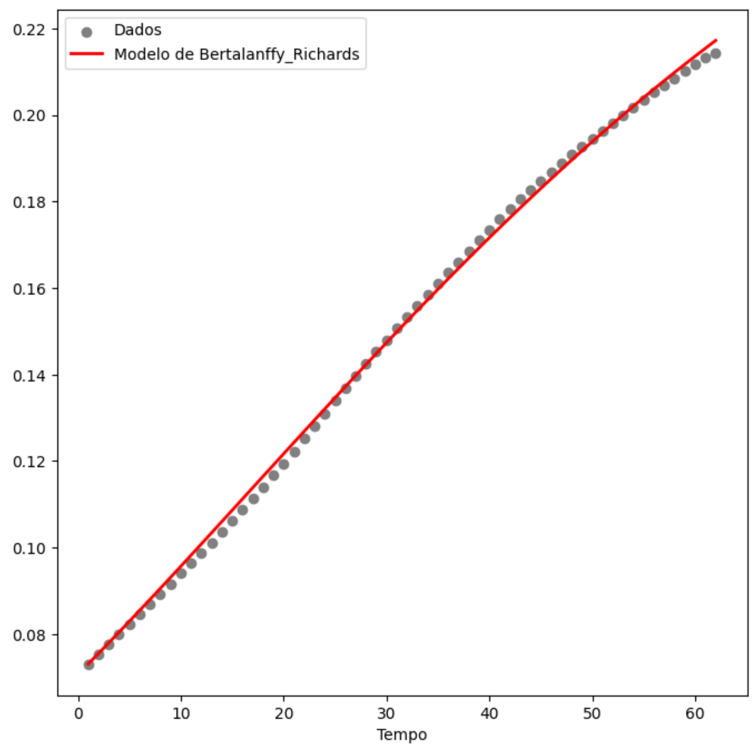
há um distanciamento grande para a maior parte do intervalo, além da diferença do comportamento das curvas.

FIGURA 8 Plotagem do resultado do ajuste à curva - Modelo de Gompertz.



Fonte: Elaborado pelo autor.

FIGURA 9 Plotagem do resultado do ajuste à curva - Modelo de Bertalanffy-Richards.



Fonte: Elaborado pelo autor.

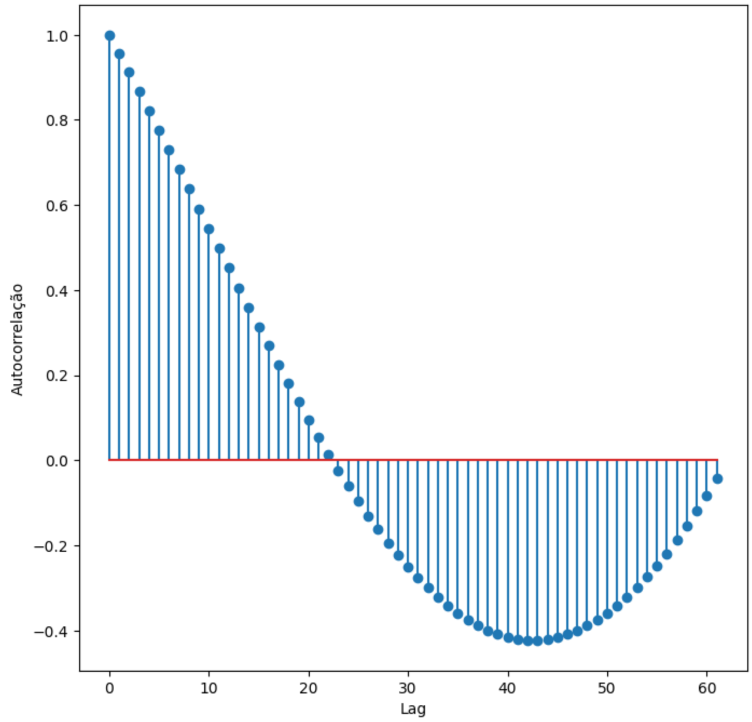
O resultado do ajuste aos Modelos de Gompertz e de Bertalanffy-Richards, são bem similares, no entanto eles não capturam a totalidade dos dados embora apresentem uma boa aproximação, para ambos o Erro Absoluto gira em torno de 0.013 e o Erro Quadrático Médio 2.23x10^-6 e 2.29x10^-6 respectivamente.

5.3 Análise Experimental

5.3.1: Autocorrelação

Avaliando a autocorrelação presente na base de dados observa-se uma queda gradual à medida que o número do *lag,* períodos de tempo cada vez mais distantes, cresce o que sugere a existência de dependência dos valores atuais em relação aos valores passados, vê se também que os valores são significativos para a estatística calculada.

FIGURA 10 Plotagem do valor da autocorrelação em função do *'lag'*.

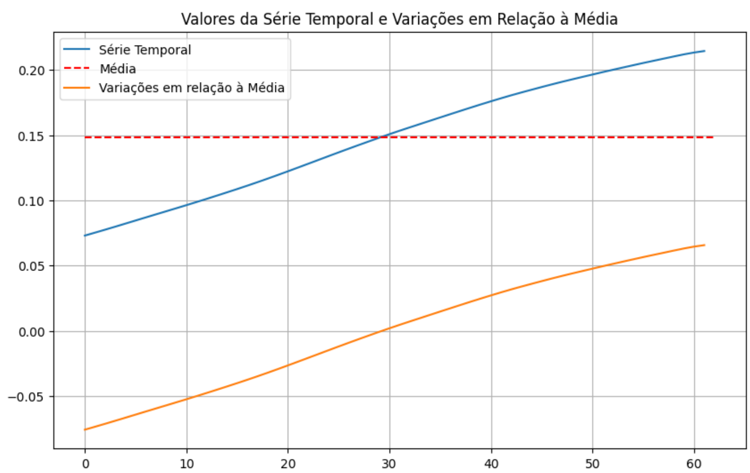


Fonte: Elaborado pelo autor.

5.3.2: Estacionariedade

No que diz respeito a estacionariedade, observa-se que a série temporal do tamanho da população apresenta uma tendência de crescimento, tal característica é critério excludente para a presença de estacionariedade conforme discutido no subcapítulo 4.2.

FIGURA 11 Plotagem da série temporal para avaliação de tendência.



Fonte: Elaborado pelo autor.

Fazendo a validação pelo método de Janelas Móveis discutido na subseção 4.2 para diversos intervalos de janela o teste retorna que a série não estacionária.

FIGURA 12 Teste de estacionariedade - Método de Janelas Móveis (Média) Iteração 1.

```
Media Total: 0.1485603418387097
Media Total - 1Dp: 0.10480459224014321
Media Janela: 0.078803972833333333
Media Total + 1Dp: 0.19231609143727618
```

Fonte: Elaborado pelo autor.

FIGURA 13 Teste de estacionariedade - Método de Janelas Móveis (Variância) Iteração 1.

```
Variancia Total: 0.0019145656229324515
Variancia Total - 1Dp: -0.04184118397563404
Variancia Janela: 1.5534907763475805e-05
Variancia Total + 1Dp: 0.04567031522149894
```

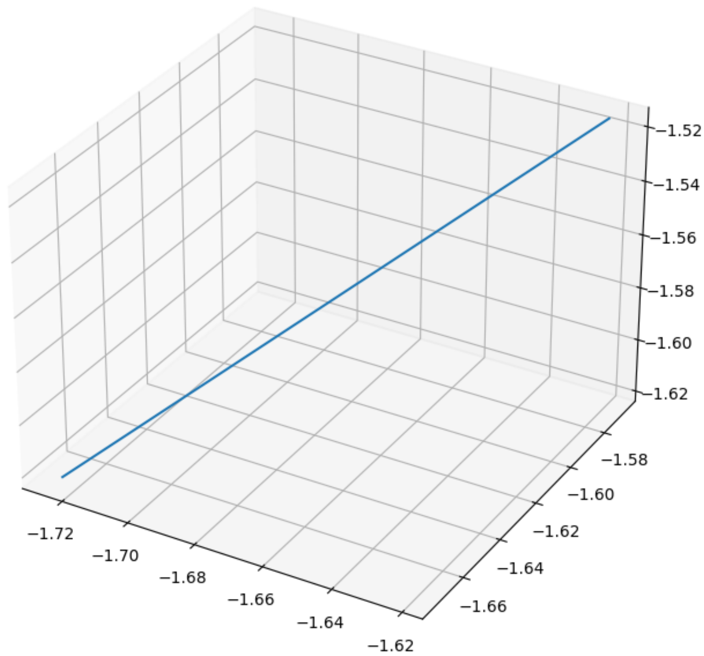
Fonte: Elaborado pelo autor.

5.3.3: Teorema de Takens

A aplicação do Teorema de Takens e a trajetória produzida através do espaço de fases não apontou a presença de nenhum comportamento complexo ou caótico para nenhum valor de atraso empiricamente testado no intervalo de 1 a 62, intervalo correspondente ao *lag* do cálculo da autocorrelação.

Para todos os plots gerados o comportamento subjacente é equivalente ao percebido na análise da série temporal e em ambos os casos a uma tendência de crescimento.

FIGURA 14 Aplicação do Teorema de Takens.



Fonte: Elaborado pelo autor.

5.3.4: Expoente de Lyapunov

O cálculo do expoente de Lyapunov não se faz útil no contexto de sua aplicação na base de dados obtidas uma vez que a trajetória produzida pelo espaço de fases aparenta ter um comportamento de tendência crescente e não caótico como os Atratores de Lorenz e Rossler utilizados por Wolf (1984).

As demais propriedades dos sistemas dinâmicos como estabilidade, equilíbrio, atratores, e presença de caos dependiam do sucesso integral da aplicação do ajuste à algum dos modelos matemáticos para descrever o conjunto de dados uma vez que são necessárias a utilização das equações diferenciais ou de diferenças para que essas propriedades pudessem ser exploradas com exatidão, como não houve um ajuste livre de erro quaisquer análises a respeito dessas propriedades não estariam bem fundamentadas.

6 Conclusão

A atividade principal de caracterização de um conjunto de dados populacionais "ao acaso" como originário de um sistema dinâmico foi um trabalho complexo, sendo necessária muita busca por referências na tentativa de se obter critérios bem fundamentados que permitissem o cumprimento da tarefa de caracterização.

Dessa forma, realizou-se pesquisa na teoria demográfica e na matemática a fim de entender como é a abordagem dessas duas áreas no estudo e compreensão da dinâmica populacional. Através desses estudos chegou-se a alguns modelos matemáticos que tentam explicar o comportamento de crescimento e prever a curva populacional.

Nesse sentido, técnicas de ajuste de curva a equações diferenciais foram exploradas utilizando a linguagem de programação python na tentativa de encontrar os melhores parâmetros de cada modelo e avaliar se é possível obter a função originadora. Embora a tarefa tenha sido concluída com relativo êxito para dois dos três modelos, o fato de não termos um grau de acurácia satisfatório limitou a exploração de algumas das propriedades que seriam úteis na caracterização.

A partir disso, foi necessário explorar outros caminhos de análise experimental tanto de séries temporais quanto de testes relacionados a conjuntos de dados de sistemas dinâmicos e para a maioria desses testes não foi possível estabelecer a caracterização, alguns falharam e outros caíram em casos excludentes de aplicação.

Portanto pode-se concluir que a busca por mecanismos e por desenvolvimento de metodologia para a caracterização de uma base de dados populacionais como sendo de um sistema dinâmico falhou e que para trabalhos futuros faz-se necessário a utilização de outras abordagens mais avançadas na tentativa de explorar outras visões ou propriedades que por ventura não foram cobertas.

7 Referências Bibliográficas

[1] Strogatz, S. H. Nonlinear Dynamics and Chaos - With Applications to Physics, Biology, Chemistry, and Engineering. Nova Iorque. Perseus Books. 1994. 505;

[2] Alligood, K.; Sauer, T.; Yorke, J. CHAOS: An Introduction to Dynamical Systems. Nova Iorque. Springer. 1996. 612;

[3] EQUAÇÃO DE LOTKA-VOLTERRA. In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2023. Disponível em: <https://pt.wikipedia.org/wiki/Equação_de_Lotka-Volterra>. Acesso em 11 de mai. de 2023.;

[4] Domingues, A. Aula 4 - Circuitos CA e Caos. 08 abr.2013. Apresentação do Power Point. Disponível em: <http://www.fap.if.usp.br/~hbarbosa/uploads/Teaching/LabAberto2013Fis4/Circuitos_Aula04_2013_Discussao.pdf> . Acesso em 12 de mai. de 2023.;

[5] Malthus, T. An Essay on the Principle of Population. Londres. Capítulo 2. 1798.;

[6] Ribeiro, F. An attempt to unify some population growth models from first principles. Revista Brasileira De Ensino De Física, 39 (1), 2017.;

[7] Braga, A.; Ribeiro, A.; de Jesus, M.; & Ribeiro, H.. Sobre a Detecção de Autocorrelações em Séries Temporais: Uma Comparação Objetiva entre Análise de Flutuações, Transformações Wavelet e Análise Entrópica.III CMAC-SE - Congresso de Matemática Aplicada e Computacional Sudeste, 2015.;

[8] CAMPANHARO, A.; Ramos, F.; Macau, E. Análise de sinais turbulentos na copa da Floresta Amazônica: em busca de comportamento caótico e estruturas coerentes. 2006. 124 p.

IBI: <6qtX3pFwXQZGivnJSY/N62ey>. (INPE-14604-TDI/1184). Dissertação (Mestrado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2006.;

[9] Takens's theorem . In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2023. Disponível em: <https://en.wikipedia.org/wiki/Takens%27s_theorem >. Acesso em 14 de mai. de 2023.;

[10] Dos Santos, L.; Macau.E. Caracterização da dinâmica caótica em séries temporais. Workshop dos Cursos de Computação Aplicada do INPE. Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos,, 2010.;

[11] Feijo, Carmen. Conceitos Básicos de Séries Temporais para Modelagem Macroeconômica. 12 abr.2005. Apresentação do Power Point. Disponível em: <http://www.icad.puc-rio.br/cfeijo/pdf/revisão%20básica%20séries%20temporais_material%20de%20apoio_curso%20teoria%20macroeconomica_PPGE%20UFF.pdf> . Acesso em 01 de jun. de 2023.;

[12] Galdino, V. Técnicas para estimação de expoentes de Lyapunov em sistemas dinâmicos não-lineares 2018. 60 p. Dissertação (Mestrado em Modelagem Matemática e Computacional). Universidade Federal da Paraíba, João Pessoa, 2018.