

Heverton Barbosa do Nascimento

**Previsão de área alagada de corpos hídricos via
modelos de séries temporais geradas por imagens de
satélite**

**Crato-CE
Maio 2023**



Universidade Federal do Cariri
Centro de Ciências Agrárias e Biodiversidade
Programa de Pós-Graduação em Desenvolvimento Regional Sustentável

Previsão de área alagada de corpos hídricos via modelos de séries temporais geradas por imagens de satélite

Dissertação apresentada ao Programa de Pós-Graduação em Desenvolvimento Regional Sustentável, Centro de Ciências Agrárias e Biodiversidade da Universidade Federal do Cariri, como requisito parcial para obtenção do título de Mestre em Desenvolvimento Regional Sustentável.

Orientador: Francisco Alixandre Avila Rodrigues

Coorientador: Paulo Renato Alves Firmino

Crato-CE
Maio 2023

HEVERTON BARBOSA DO NASCIMENTO

Previsão de área alagada de corpos hídricos via modelos de séries temporais geradas por imagens de satélite.

Dissertação apresentada ao Programa de Pós-Graduação em Desenvolvimento Regional Sustentável, Centro de Ciências Agrárias e Biodiversidade da Universidade Federal do Cariri, como requisito parcial para obtenção do título de Mestre em Desenvolvimento Regional Sustentável.

Orientador: Francisco Alixandre Avila Rodrigues

Coorientador: Paulo Renato Alves Firmino

Prof. Dr. Francisco Alixandre Avila Rodrigues (Orientador)
Universidade Federal do Cariri (UFCA)

Prof. Dr. Paulo Renato Alves Firmino (Coorientador)
Universidade Federal do Cariri (UFCA)

Profa. Dra. Rosilda Benício de Souza
Universidade Federal do Cariri (UFCA)

Prof. Dr. Carlos Wagner Oliveira
Universidade Federal do Cariri (UFCA)

Prof. Dr. Ricardo Holanda Nobre
Universidade Estadual do Ceará (UECE)

Dados Internacionais de Catalogação na Publicação.
Universidade Federal do Cariri.
Sistema de Bibliotecas.

- N244p Nascimento, Heverton Barbosa do.
Previsão de área alagada de corpos hídricos via modelos de séries temporais geradas por imagens de satélite / Heverton Barbosa do Nascimento. – 2023.
53 f.: il. color.30 cm
- Dissertação (mestrado) - Universidade Federal do Cariri, Mestrado em Desenvolvimento Regional Sustentável (PRODER), Crato, 2023.
- Orientação: Prof. Dr. Francisco Alixandre Avila Rodrigues.
Coorientação: Prof. Dr. Paulo Renato Alves Firmino.
1. Processamento digital de imagens. 2. Método de Otsu. 3. Série temporal. 4. Área alagada. 5. Corpos hídricos. I. Título.

CDD 621.367

Bibliotecária: Glacínésia Leal Mendonça
CRB 3/ 925

Agradecimentos

Gostaria de agradecer primeiramente a Deus por permitir que este momento seja possível. Gostaria de expressar minha gratidão ao meu orientador professor Alixandre Avila por me guiar nesta jornada, e ao professor Paulo Renato, meu coorientador, grande profissional, amigo e admirável orientador, desde quando me acolheu no Grupo de Pesquisa MESOR no ano de 2019. Obrigada por todas as orientações, conselhos e palavras amigas. Este trabalho não teria sido possível sem a ajuda de vocês.

Além disso, também gostaria de agradecer:

Os membros da banca examinadora: Rosilda Benício e Ricardo Holanda;

A todos os professores do PRODER que contribuíram para o meu desenvolvimento científico e pessoal.

A Minha esposa (Teresa Amélia) e minhas filhas (Letícia Macêdo e Larissa Macêdo) que são minha zona de conforto e meu porto seguro.

Minha mãe (Maria Asminda), pai (Cosmo do Nascimento) e a minha irmã (Cláudia Micaelle) que foram responsáveis por proporcionar uma base sólida para que eu seguisse em frente com meus sonhos;

Aos membros do grupo de pesquisa MESOR. Em especial, Jair Paulino que sempre me incentivou a seguir em frente, trazendo sempre sua experiência e sua capacidade de resolução de problemas em modelagem estatística;

Meus colegas do PRODER. Em especial, a Tatiane, Ramsés, Vânia e Wal, os quais construímos uma parceria forte através de alguns trabalhos que fizemos juntos. Obrigada a todos;

E por fim, a todos aqueles que acreditaram em mim.

Resumo

Metodologias de processamento digital de imagens podem ser aplicadas a uma ampla classe de problemas em visão computacional. Dentre as aplicações, pode ser citada a avaliação da dinâmica de corpos hídricos a partir de imagens. O monitoramento e a previsão de volumes e áreas alagadas de corpos hídricos são fundamentais para emitir alertas desde inundações à escassez de água. Para prever a dinâmica da área alagada de corpos hídricos, esse artigo propõe o uso de modelos de séries temporais geradas a partir de imagens multitemporais. O corpo hídrico investigado foi o lago Kirwin (USA) no período de 2013 a 2022. As imagens do lago foram adquiridas em cada período e suas áreas alagadas foram estimadas com o auxílio do método de Otsu. Para descrever o comportamento temporal da área alagada do lago Kirwin, foram investigados modelos autorregressivos integrados de médias móveis (ARIMA), autorregressivos integrados de médias móveis com sazonalidade (SARIMA), de suavização exponencial (ETS) e de Holt-Winters. Dentre os modelos investigados para prever a dinâmica da área alagada do corpo hídrico, o de Holt-Winters foi o que obteve as melhores medidas de desempenho, com erro de previsão inferior a 6% em relação à série predita. Sua previsão sugere um cenário que para os 12 meses subsequentes à série sob estudo há uma baixa possibilidade de ocorrer aumento da área alagada na região do lago Kirwin.

Palavras-chave: processamento digital de imagens, método de otsu, série temporal, área alagada, corpos hídricos.

Abstract

Digital Image processing methodologies can be applied to a whole variety of problems at computer view; among them, one finds the dynamic assessment of waterbodies based on images. Monitoring and predicting the volume and flooded area of waterbodies are essential procedures for the emission of warnings about several events, from floods to water shortage. The aim of the present study is to use a time series model generated by images to predict the dynamics of a flooded areas in a waterbody. Kirwin Lake (USA) was the herein assessed waterbody and the study was carried out between 2013 and 2022. Images of the lake were taken in each assessed period-of-time and its flooded area was estimated through the Otsu method. Autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA), exponential smoothing (ETS), and Holt-Winters models were adopted to describe the temporal behavior of the flooded area in the Kirwin Lake. The Holt-Winters model was the one recording the best performance measurements: error prediction lower than 6% in comparison to the target series. Predictions showed low probability of flooded area enlargement in the Kirwin Lake, in the 12 months subsequent to the series.

Keywords: digital processing of images, Otsu method, time series, flooded area, waterbodies.

Lista de ilustrações

Figura 1 – Reservatório Kirwin, Kansas, EUA. Fonte:Google Maps.	30
Figura 2 – Percurso Metodológico. Fonte: Autor	31
Figura 3 – Imagens redimensionadas do reservatório Kirwin adquiridas no site da USGS em momentos distintos. Fonte: USGS	32
Figura 4 – Fluxograma do algoritmo para processamento das imagens. Fonte: Autor.	33
Figura 5 – Fluxograma do algoritmo para calcular área do corpo hídrico. Fonte: Autor.	33
Figura 6 – Imagens do reservatório Kirwin após aplicação do método de Otsu em momentos distintos. Fonte: Autor.	34
Figura 7 – Fluxograma do algoritmo para modelagem estatística de série temporal. Fonte: Autor.	35
Figura 8 – Fluxograma para avaliar o desempenho dos modelos. Fonte: Autor. . .	35
Figura 9 – Box plot com evolução anual da área alagada do reservatório Kirwin. Fonte: Autor.	37
Figura 10 – Série Temporal da área alagada do lago Kirwin. Fonte: Autor.	38
Figura 11 – Série Temporal da área alagada do lago Kirwin após transformação logarítmica. Fonte: Autor.	38
Figura 12 – Decomposição da Série Temporal das áreas alagadas do lago Kirwin. Fonte: Autor.	39
Figura 13 – Separação de dados de treinamento e teste. A linha vertical em vermelho separa os dados de treinamento (à esquerda) e de teste (à direita). Fonte: Autor.	40
Figura 14 – Autocorrelação parcial dos resíduos do modelo ARIMA. Fonte: Autor.	40
Figura 15 – Autocorrelação parcial dos resíduos dos modelos utilizados. Fonte: Autor.	42
Figura 16 – Comparativo das previsões dos modelos avaliados com os dados de teste.	43
Figura 17 – Boxplot comparando as previsões dos modelos avaliados com os dados de teste.	43
Figura 18 – Previsão final da área alagada considerando 12 meses à frente.	44

Sumário

Agradecimentos	3	
Resumo	4	
Abstract	5	
Lista de ilustrações	6	
Lista de símbolos	6	
Sumário	7	
1	INTRODUÇÃO	9
1.1	Objetivos	10
1.1.1	Geral	10
1.1.2	Específicos	10
1.2	Estrutura da Dissertação	11
2	REVISÃO DE LITERATURA	12
2.1	Desenvolvimento Sustentável	12
2.2	Processamento Digital de Imagens	14
2.2.1	Segmentação de imagens	14
2.2.1.1	Método de Otsu	15
2.3	Série temporal	16
2.3.1	Modelos de série temporal	19
2.3.1.1	Autoregressivo integrado de média móvel	19
2.3.1.2	Autoregressivo integrado de média móvel com sazonalidade	20
2.3.1.3	Suavização Exponencial	21
2.3.1.4	Holt-Winters	23
2.4	Medidas de desempenho	24
2.4.1	Critério de Informação de Akaike (AIC)	24
2.4.2	Teste de Shapiro-Wilk	25
2.4.3	Root Mean Squared Error (RMSE)	26
2.4.4	Mean Absolute Percentage Error (MAPE)	26
2.4.5	Coeficiente de Theil	27
3	MATERIAIS E MÉTODOS	29

3.1	Área de estudo	29
3.1.1	Conjunto de dados	30
3.2	Percurso metodológico	30
3.2.1	Coleta de dados	31
3.2.2	Processamento das imagens	32
3.2.3	Calcular área do corpo hídrico	33
3.2.4	Modelagem de série temporal	34
3.2.5	Avaliar desempenho dos modelos	35
4	RESULTADOS E DISCUSSÃO	37
4.1	Modelagem e Previsão	40
5	CONSIDERAÇÕES FINAIS	45
	REFERÊNCIAS	46

1 Introdução

Análises de corpos hídricos através de processamento digital de imagens são destaque no meio científico, para gerar ações em busca do desenvolvimento sustentável. Como exemplo, podem ser citados o mapeamento de deslizamento de terras[1], estudo da estimativa de superfícies impermeáveis[2], detecção de mudanças em florestas[3], monitoramento de inundações de rios[4, 5], entre outros. Uma das vantagens evidentes é que esse tipo de processamento permite que os pesquisadores façam as análises de fenômenos naturais através de imagens de satélites, sem a necessidade de se deslocar até o local de estudo. Este interesse acontece no intuito de melhorar as informações visuais para a interpretação humana e o processamento de imagens para armazenamento, transmissão e representação dos dados obtidos[6].

Neste sentido, existem estudos interessantes que utilizam uma dada série temporal de imagens adquiridas em tempos igualmente espaçados, para a obtenção de uma imagem de tempo futuro[7]. Isso proporciona uma modelagem dos níveis de cinza, indicando o crescimento ou decréscimo de uma região que está sendo ilustrada em uma determinada imagem estudada [7], contribuindo significativamente em análises relacionadas ao desenvolvimento sustentável de uma região em estudo. Portanto, este tipo de estudo estatístico consiste em analisar um conjunto de observações ordenadas no tempo[8] e logo após, tentar prever os próximos resultados dos dados analisados. Este tipo de estudo pode ser utilizado por exemplo para o mapeamento de alta resolução de mudanças florestais através da análise de séries temporais de imagens de satélite de radar[3].

Em paralelo a estes pontos, existe a necessidade de avaliar estes fenômenos afim de buscar soluções preventivas para evitar os danos as populações expostas. O monitoramento da superfície de lagos e rios é fundamental para orientar ações que minimizem danos às populações expostas[9]. Para monitorar corpos hídricos por meio de imagens, se faz necessário o constante imageamento da região que contém o corpo hídrico. Desta forma obtém-se, uma sequência de imagens da mesma região em tempos distintos, chamada de imagens multi-temporais. Um dos aspectos importantes para ser avaliado em um corpo hídrico, a partir de imagens multitemporais, é a dinâmica de sua área alagada. Um caminho natural para estimar a área alagada é fazer uso de métodos de segmentação. Alguns autores utilizaram o método clássico de segmentação conhecido como Otsu[10] para identificar a superfície de lagos[9]. Outros apresentaram um estudo comparando técnicas de segmentação de imagens utilizadas na extração de informações sobre água a partir de

imagens digitais para monitoramento de enchentes na Malásia[5].

O monitoramento e a previsão da área alagada de corpos hídricos são fundamentais para emitir alertas desde inundações à escassez de água. Diante destes pontos, este trabalho está voltado a apresentar uma forma de analisar automaticamente a área de ocupação da água de um corpo hídrico através do processamento de imagens de satélite, calculando sua área alagada em várias datas ordenadas no tempo através de algoritmos de programação. Com a modelagem estatística e análise destas informações, pode-se verificar como o objeto em estudo vai se comportar com relação a sua área ocupado em seu entorno, no decorrer do tempo. Com estas informações, os responsáveis por iniciativas públicas terão a oportunidade de tomar ações preventivas para conter os efeitos prejudiciais a população que residem em tona da área avaliada.

Neste trabalho é proposta uma metodologia para avaliar e prever a dinâmica da área alagada de corpos hídricos. O corpo hídrico investigado aqui foi o lago Kirwin (USA) no período de 2013 a 2016. As imagens do lago foram adquiridas em cada período e suas áreas alagadas foram estimadas com o auxílio do método de Otsu e a informação da resolução espacial das imagens. Para descrever o comportamento temporal e prever a dinâmica da área alagada do lago Kirwin, modelos de séries temporais guiadas pelas imagens do lago foram investigados. Os modelos testados foram o autorregressivo integrado de médias móveis (ARIMA)[11], autorregressivo integrado de médias móveis com sazonalidade (SARIMA)[12], modelo de suavização exponencial (ETS)[13], e o de Holt-Winters [11].

1.1 Objetivos

1.1.1 Geral

Prever área alagada de corpos hídricos através de modelos de séries temporais geradas a partir de processamento de imagens de satélite.

1.1.2 Específicos

- (i) Catalogar fontes disponíveis para obtenção de históricos de imagens digitais de corpos hídricos;
- (ii) Implementar algoritmos para processamento das séries de imagens obtidas de (i);
- (iii) Implementar algoritmo para o cálculo da área do espelho d'água dos corpos hídricos processados a partir de (ii);

- (iv) Desenvolver algoritmos de Modelagem e avaliação estatística das séries temporais obtidas em (iii);
- (v) Avaliar predição da área alagada através do desempenho dos modelos utilizados.

1.2 Estrutura da Dissertação

O presente trabalho está organizado inicialmente com o Capítulo 2, o qual apresenta uma revisão de literatura sobre desenvolvimento sustentável, processamento digital de imagens e análise de séries temporais. Logo em seguida temos o Capítulo 3, o qual discorre sobre os procedimentos metodológicos a serem adotados para execução do presente trabalho. Na sequência são apresentados os resultados e discussões no capítulo 4, no qual as análises são feitas através de dados e gráficos gerados durante a pesquisa. Por fim, no Capítulo 5, são apresentadas as considerações finais do estudo.

2 Revisão de Literatura

2.1 Desenvolvimento Sustentável

Desenvolvimento sustentável é um tema que tem ganhado cada vez mais importância no contexto global. Ele pode ser entendido como um modelo de desenvolvimento que atende às necessidades do presente sem comprometer a capacidade das gerações futuras de atenderem às suas próprias necessidades[14]. Em outras palavras, o desenvolvimento sustentável busca integrar as dimensões social, econômica e ambiental de forma equilibrada, promovendo um desenvolvimento justo e duradouro.

Este tipo de desenvolvimento é fundamental para a preservação da biodiversidade e dos recursos naturais. A utilização excessiva de recursos naturais e a degradação ambiental podem comprometer o desenvolvimento socioeconômico e a qualidade de vida das pessoas, especialmente das comunidades mais vulneráveis[15]. Nesse sentido, é importante que as políticas públicas e as empresas incorporem a sustentabilidade em suas atividades e operações, visando à preservação ambiental e à justiça social.

Uma das principais estratégias para o desenvolvimento sustentável é a transição para uma economia verde. Este tipo de economia busca promover o crescimento econômico e a inclusão social, ao mesmo tempo em que reduz os impactos ambientais negativos[16]. Isso envolve a adoção de tecnologias limpas e eficientes, a utilização de fontes de energia renováveis e a promoção do consumo sustentável. Além disso, é importante que haja uma mudança nos padrões de produção e consumo, estimulando a economia circular e a redução do desperdício.

Para que a transição para uma economia verde seja bem sucedida, é necessário que haja uma cooperação internacional. A mitigação das mudanças climáticas e a promoção do desenvolvimento sustentável exigem a colaboração entre países e setores[17]. Isso envolve a transferência de tecnologia e conhecimento, bem como a promoção de investimentos em energias renováveis e em infraestrutura sustentável. Além disso, é importante que haja um diálogo aberto e transparente entre as partes interessadas, visando à construção de soluções coletivas[17].

Uma outra ferramenta importante para o desenvolvimento sustentável é a educação

. Para o desenvolvimento sustentável, ela busca capacitar as pessoas a compreenderem os desafios socioambientais e a adotarem comportamentos sustentáveis[18]. Isso envolve a promoção da educação ambiental, da educação para a cidadania global e da educação para a paz. Além disso, é importante que haja uma integração da sustentabilidade nos currículos escolares e na formação de profissionais.

Para que isto se torne mais evidente, é fundamental a atuação de uma governança global em busca do desenvolvimento sustentável. Este tipo de governança envolve a coordenação de atores e instituições globais, nacionais e locais para a promoção do desenvolvimento sustentável[19]. Isso envolve a participação ativa da sociedade civil e do setor privado, bem como a promoção de políticas públicas efetivas e a implementação de acordos e tratados internacionais.

Este tipo de desenvolvimento é o que permite que necessidades atuais da população sejam atendidas, não comprometendo a capacidade das gerações futuras em satisfazer suas próprias necessidades[20]. Logo, estes recursos não deverão se esgotar para o futuro quando se fala em organização do desenvolvimento econômico e a conservação ambiental.

Existem alguns pontos importantes que são abordados para o desenvolvimento das próximas gerações em conceitos sustentáveis, devido ao crescimento acelerado. Um dos fatores deste crescimento é o fator da globalização que influencia também no desenvolvimento sustentável [21]. Diante disso, as organizações devem se adaptar as medidas e ações que favorecem o desenvolvimento sustentável. É evidente que se grande parte das empresas atuais resistirem ao desenvolvimento sustentável, perderão mercado e conseqüentemente fecharão as portas[21]. Com isso, busca-se desenvolver uma visão otimista do futuro e das estratégias sociais, econômicas e ambientais necessárias para torná-la realidade[21]. A evidência disto são meios para se calcular, definir e avaliar o capital.

Esta necessidade de mudança nas atitudes consumistas, principalmente de nações do norte, são influenciadas por nações do sul[22]. Eles apresentam uma nova conceituação do desenvolvimento como apropriação efetiva de todos os direitos humanos, políticos, sociais, econômicos e culturais, incluindo-se o direito coletivo ao meio ambiente. Com isso, o desenvolvimento sustentável deve ser conduzido em todo o planeta, promovendo o equilíbrio na utilização dos vários recursos naturais para se viver com dignidade e proporcionar essa mesma oportunidade as gerações futuras[22].

Quando aborda-se temáticas envolvendo o lado econômico e social ligados ao desenvolvimento sustentável, relaciona-se sempre a necessidade de equilíbrio na utilização dos

recursos naturais, mantendo a preservação da biodiversidade[22]. Diante disto, os recursos hídricos que são parte integrante desta biodiversidade, surgem como itens importantes para serem monitorados, sabendo que sua alteração relacionado a mudanças climáticas pode causar danos a sociedade. Isso provoca uma necessidade de se preocupar com a preservação do meio ambiente para que a geração atual tenha uma vida digna e que as próximas gerações também possam usufruir deste recurso[22].

2.2 Processamento Digital de Imagens

O campo do processamento digital de imagens se refere ao processamento de imagens digitais por um computador digital[6]. Quando se refere a imagem digital, existem vários elementos que a compõe onde cada um tem uma localização específica e o seu valor[6]. Um dos elementos mais importantes e o mais utilizado para representar os elementos deste tipo de imagem é o pixel[6].

Em meio a esta necessidade de utilização de ferramentas de tecnologia para estes estudos, tornam-se mais comuns ao passar dos anos, pesquisas com o intuito de desenvolver novas tecnologias com o processamento de imagens[6]. Tendo em vista que a visão computacional faz parte desta evolução, existem níveis de processamento que são caracterizados por suas entradas serem imagens e suas saídas serem atributos extraídos destas imagens[6].

Quando fala-se em campo de atuação do processamento digital de imagens, não existe nenhuma área de empreendimento técnico que não utilize de alguma forma esse tipo de processamento[6]. Para que isso seja possível, é necessário uma fonte de energia para imagens, onde a mais utilizada é o espectro eletromagnético de energia[6]. Este tipo de energia provocado por ondas eletromagnéticas, são ondas senoidais, de vários comprimentos, que se propagam em um fluxo de partículas sem massa, com um padrão ondulatório e se movendo na velocidade da luz[6]. A quantidade de energia de uma partícula sem massa, que é chamada de fóton, é responsável por formar o espectro eletromagnético que é composto por bandas espectrais agrupadas de acordo com a energia do fóton[6]. Os extremos deste espectro varia entre os raios gama que tem uma alta energia e as ondas de rádio que tem a mais baixa energia[6].

2.2.1 Segmentação de imagens

A segmentação de imagens é uma técnica fundamental no processamento de imagens digitais. Esse processo consiste em dividir uma imagem em regiões com características semelhantes, com o objetivo de identificar e extrair informações relevantes de uma imagem.

É utilizada em áreas de estudo que tem a necessidade de identificar com mais exatidão as regiões de interesse contidas em uma imagem[9] Por conta disto, esta técnica faz parte de sistemas de monitoramento de inundações usando uma abordagem de visão computacional[5]. A avaliação dos resultados de segmentação de imagens, desempenha um papel importante em uma análise mais aprofundada da imagem e tomada de decisão[23].

Quando fala-se em segmentação de imagens para monitoramento de rios e lagos, existem métodos utilizados que melhoram o desempenho[24]. Um deles é a limiarização que trata-se de um processo de segmentação de imagens baseado na diferença dos níveis de cinza que compõe diferentes objetos de uma imagem[6]. Como a limiarização tem a possibilidade de ser vista como um problema teórico de decisão estatística[6], para contribuir com minimização do erro médio que ocorre na atribuição de pixels para mais de 1 grupo, pode-se utilizar o método de otsu para maximizar a variância entre estas classes[24].

2.2.1.1 Método de Otsu

O método de Otsu foi proposto por Nobuyuki Otsu em 1979 e é um método de segmentação de imagens que utiliza a análise do histograma dos níveis de cinza da imagem para encontrar o limiar ótimo de segmentação. Logo ele tem uma peculiaridade de se basear em cálculos realizados no histograma de uma imagem[10]. Este método determina o limiar ótimo de segmentação através da maximização da variância intra-classes, que é a variância das intensidades de pixel dentro de cada classe[25].

Este método é muito eficiente na segmentação de imagens em que o objeto de interesse e o fundo possuem diferenças claras de intensidade. Entretanto, quando a imagem apresenta variações de iluminação, ruído ou outras distorções, o método de Otsu pode apresentar resultados insatisfatórios[25]. Nesses casos, é necessário utilizar técnicas de pré-processamento para melhorar a qualidade da imagem antes de aplicar o método de Otsu. Logo, o pré-processamento da imagem é fundamental para melhorar a eficiência do método de Otsu em imagens com distorções[25].

Uma das principais vantagens do método de Otsu é que ele é simples e rápido de implementar, o que o torna amplamente utilizado em diversas aplicações de processamento de imagens. Além disso, o método de Otsu pode ser combinado com outras técnicas de segmentação para obter resultados ainda mais precisos[25]. Porém, é importante destacar que o método de Otsu não é adequado para todos os tipos de imagens e que é necessário avaliar a qualidade do resultado obtido após a aplicação do método[10]. Com isso, a seleção do método de segmentação mais apropriado depende das características da imagem e do objetivo da aplicação[25].

Devido a sua simplicidade e eficiência, ele é um dos métodos mais utilizados para a segmentação de imagens em aplicações práticas[25]. A intensão inicial é realizar a separação dos pixels dentro de uma imagem em dois grupos. O objeto separado é caracterizado por w_0 e u_0 os quais representam a razão entre o número de pixels e o nível de cinza médio[26]. De uma forma similar, também existem dois parâmetros, w_1 e u_1 no plano de fundo da imagem. Desta forma, define-se a média total do nível de cinza da imagem de acordo com a Equação (2.1) [26]:

$$u = w_0(t)u_0(t) + w_1(t)u_1(t) \quad (2.1)$$

onde t denota um nível cinza da imagem. Quando maximizamos a função de otimização $f_s(t)$, como mostra a Equação (2.2), a imagem é binarizada [26].

$$Max f_s(t) = w_0(t)(u_0(t) - u)^2 + w_1(t)(u_1(t) - u)^2 \quad (2.2)$$

O valor do topo do nível de cinza correspondente ao valor máximo de $f_s(t)$ é selecionado como o valor limite para a binarização da imagem. Desta forma, o método de Otsu será de grande importância para determinar os pixels que correspondem a área seca e os que determinam a ocupação da água na área em estudo.

2.3 Série temporal

Uma série temporal é um registro sequenciado de observações realizado em períodos regulares[27]. Elas são amplamente usadas em diversas áreas, incluindo finanças, economia, engenharia e ciências sociais, entre outras. A análise de séries temporais é fundamental para entender padrões temporais, prever comportamentos futuros e tomar decisões importantes para a sociedade em geral. Desta forma, existem vários estudos recentes sobre séries temporais e suas aplicações.

Inicialmente pode-se destacar a análise de séries temporais para prever a demanda de energia elétrica em Taiwan[28]. Neste estudo foi utilizado um modelo ARIMA para modelar a série temporal de demanda de energia elétrica e compararam as previsões com um modelo de regressão linear. Os resultados mostraram que o modelo ARIMA teve um desempenho melhor em prever a demanda de energia elétrica em um horizonte de previsão de uma semana.

Em outro estudo, foi aplicada a análise espectral para analisar a variabilidade da frequência cardíaca em pacientes com doença de Alzheimer[29]. Os autores utilizaram uma transformada de Fourier para analisar a série temporal de frequência cardíaca e identificaram padrões de frequência cardíaca em diferentes estágios da doença. Os resultados

mostraram que a análise espectral pode ser uma ferramenta útil para avaliar a progressão da doença de Alzheimer.

Um estudo recente aplicou a análise de séries temporais para prever a taxa de câmbio entre o dólar americano e a rupia indiana[30]. Os autores utilizaram um modelo ARIMA para modelar a série temporal de taxa de câmbio e compararam as previsões com um modelo de redes neurais artificiais. Os resultados mostraram que o modelo ARIMA teve um desempenho melhor em prever a taxa de câmbio em um horizonte de previsão de um mês.

Um outro exemplo é um estudo em que os autores aplicaram a análise de séries temporais para prever a produção de eletricidade na China[31]. Eles utilizaram um modelo SARIMA para modelar a série temporal de produção de eletricidade e compararam as previsões com um modelo de redes neurais artificiais. Os resultados mostraram que o modelo SARIMA teve um desempenho melhor em prever a produção de eletricidade em um horizonte de previsão de um dia.

Por fim, um estudo realizado também na china, aplicou a análise de séries temporais para modelar a temperatura do solo em um campo de cultivo de milho[32]. Neste caso foi utilizado um modelo de regressão para modelar a relação entre a temperatura do solo e as variáveis ambientais, como a temperatura do ar e a umidade relativa. Os resultados mostraram que o modelo de regressão pode ser uma ferramenta útil para prever a temperatura do solo em um campo de cultivo de milho.

Então, quando fala-se em séries temporais, remete-se a dados que são observados em instantes ordenados de tempo[33]. Geralmente é representada por um gráfico que exhibe padrões, como um movimento para cima ou para baixo (tendência) ou um padrão que se repete (variação sazonal)[13]. Através destas informações que compõe uma observação de série temporal (y_1, \dots, y_n) , pode-se prever uma observação futura no tempo $n + h$ [13].

Logo, seja y_t a observação no tempo t , e seja x_t um “vetor de estado” contendo componentes não observados que descrevem o nível, tendência e sazonalidade da série, um modelo de espaço de estados de inovações lineares pode ser escrito como[13]:

$$y_t = w'x_t - 1 + E_t \quad (2.3)$$

$$x_t = Fx_{t-1} + gE_t \quad (2.4)$$

em que E_t é uma série de ruído branco e F , g e w são coeficientes. A equação de medição ou observação, representada pela Equação (2.3), descreve a relação existente entre os

estados não observados $x_t - 1$ e a observação y_t [13]. Já a Equação (2.4) é conhecida como equação de transição ou estado e descreve a evolução dos estados ao longo do tempo[13]. Com isso, uma série temporal, y , como um modelo aditivo, é representada pela combinação das componentes de tendência T , sazonalidade S e do erro E que é a parte não capturada do modelo no período t , conforme Equação (2.5) [13].

$$y = T + S + E \quad (2.5)$$

Desta forma, T corresponde a direção de longo prazo da série, S ao padrão que se repete com uma periodicidade conhecida e E o comportamento imprevisível da série [13]. Estas séries temporais podem ser classificadas como discretas ou contínuas[34]. Um exemplo de série temporal discreta são os valores diários de poluição em uma determinada cidade e contínuos são os registro de marés em um determinado porto.[34] Sua característica mais importante é que as observações vizinhas podem ser dependentes e o objetivo é analisar e modelar esta dependência. Logo, sempre que existirem dados registrados sequencialmente ao longo do tempo e for considerado o tempo um aspecto importante dos dados, tem-se uma série temporal[35]. Enquanto em modelos de regressão a ordem das observações é irrelevante para a análise, em séries temporais a ordem dos dados é crucial[27].

Os dados de séries temporais podem exibir uma quantidade variada de padrões, sendo interessante dividir estas séries em vários componentes, onde cada um representa uma categoria de padrão subjacente[11]. Essa decomposição pode ser aditiva ou multiplicativa, conforme as Equações (2.6) e (2.7), respectivamente:

$$y_t = S_t + T_t + R_t \quad (2.6)$$

$$y_t = S_t X T_t X R_t \quad (2.7)$$

Onde y_t são os dados, S_t é o componente sazonal, T_t é o componente do ciclo de tendência, e R_t é o componente restante, tudo no período t . O uso de cada uma vai depender da variação no padrão sazonal ou em torno do ciclo de tendência[11]. Uma alternativa ao uso de uma decomposição multiplicativa é transformar os dados até que a variação na série pareça estável ao longo do tempo e, em seguida, usar uma decomposição aditiva[11]. Desta forma, quando uma transformação de log for usada, ela torna-se equivalente a uma decomposição multiplicativa, conforme apresentada na Equação (2.8).

$$\log y_t = \log S_t + \log T_t + \log R_t \quad (2.8)$$

A utilização de séries temporais para analisar técnicas de aumento de dados em aplicativos de reconhecimento e imagem, por exemplo, gera um aprendizado profundo

sobre este assunto[36]. Existem pesquisadores que aproveitam esta metodologia para desenvolver novas estratégias para previsão de vendas, onde é considerado que as vendas dos produtos podem vir de diferentes cidades, distritos ou estados, caracterizando assim a importância do uso de séries temporais hierárquicas para a evolução do estudo[37].

2.3.1 Modelos de série temporal

Quando surge a necessidade de avaliação de uma série temporal, devemos construir modelos para melhor demonstrá-los. Diante deste ponto, um dos enfoques mais utilizados é o de analisar através de um domínio temporal utilizando um modelo paramétrico, ou seja, um modelo com um número finito de parâmetros[34]. Na sequência, serão apresentados os modelos estatísticos utilizados neste trabalho.

2.3.1.1 Autoregressivo integrado de média móvel

Um dos modelos utilizados em larga escala para a previsão de séries temporais é o modelo autoregressivo integrado de média móvel (ARIMA - *Autoregressive Integrated Moving Average*), proposto por George Box e Gwilym Jenkins em 1970 [38]. Quando utilizamos este tipo de modelo, devemos demonstrá-lo como $ARIMA(p, d, q)$, onde a notação p equivale ao número de parâmetros autorregressivos (AR), d representa o número de diferenciações para que a série temporal torne-se estacionária (I) e o q informa o número de parâmetros de média móvel (MA)[39].

As Equações 2.9, 2.10 e 2.11 descrevem, nesta ordem, os modelos AR $AR(p)$, ou seja, $ARIMA(p, 0, 0)$, os modelos MA $MA(q)$, ou seja, $ARIMA(0, 0, q)$, e os modelos de média móvel Autoregressiva $ARMA(p, q)$, ou seja, $ARIMA(p, 0, q)$ [40].

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \epsilon_t, \quad (2.9)$$

$$y_t = c - \epsilon_t - \Theta_1 \epsilon_{t-1} - \Theta_2 \epsilon_{t-2} - \dots - \Theta_q \epsilon_{t-q}, \quad (2.10)$$

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \epsilon_t - \Theta_1 \epsilon_{t-1} - \Theta_2 \epsilon_{t-2} - \dots - \Theta_q \epsilon_{t-q}, \quad (2.11)$$

Onde Φ , Θ e o ϵ_t são, respectivamente, os parâmetros $AR(p)$, os parâmetros $MA(q)$ e o erro aleatório ou ruído branco. O valor de c é a média das mudanças entre observações consecutivas[11]. Diante disto, a ideia principal deste modelo é baseada na suposição de linearidade da série temporal que muitas vezes segue a distribuição estatística normal[41], tendo em vista que foi criado para a análise e previsão de dados estacionários de séries temporais.

Este modelo pode ser utilizado por exemplo, para analisar o comportamento dos preços médios do etanol hidratado no estado de São Paulo, onde criou-se uma previsão para os preços deste produto[42]. Estes, podem também realizar uma previsão de atendimentos em pontos de atenção secundários com alta demanda da rede de assistência à saúde, com o auxílio do ARIMA[43].

2.3.1.2 Autoregressivo integrado de média móvel com sazonalidade

O modelo SARIMA (*Seasonal Autoregressive Integrated Moving Average*) é uma extensão do modelo ARIMA (*Autoregressive Integrated Moving Average*) que incorpora a sazonalidade em seus parâmetros. Este formalismo é comumente utilizado para modelar séries temporais com sazonalidade, sendo amplamente aplicado em diversas áreas, como finanças, economia e marketing[12].

Este modelo trata-se de uma combinação de um modelo autoregressivo (AR), um modelo de média móvel (MA) e um modelo de diferenças sazonais (S). Alguns autores explicam que o modelo ARIMA é estendido para o modelo SARIMA, incluindo um componente sazonal que permite capturar as flutuações sazonais na série temporal[44, 11]. Isso faz com que o modelo SARIMA seja capaz de lidar com séries temporais com diferentes níveis de sazonalidade, incluindo de curto e longo prazo, tendências e flutuações aleatórias.

Uma das principais vantagens do modelo SARIMA é sua capacidade de modelar séries temporais com diferentes níveis de sazonalidade[12]. Além disso, o modelo SARIMA tem sido amplamente utilizado na previsão de séries temporais em diversas áreas[44, 11]. Estudos anteriores aplicaram o modelo SARIMA em previsões de preços de ações, previsões de vendas, previsões de demanda por energia elétrica, entre outras. Como resultado, o modelo SARIMA tem sido eficaz em prever flutuações sazonais, tendências e variações aleatórias em séries temporais.

Neste sentido, o modelo SARIMA foi proposto para previsão de séries temporais sazonais, seguindo o modo de apresentação ARIMA (p, d, q) (P, D, Q), onde os parâmetros (P, D, Q) são os equivalentes sazonais de (p, d, q), que descreve a parte não sazonal[41]. Desta forma, o SARIMA pode ser expresso pela Equação 2.12:

$$\phi_p(B)\Phi_P(B^s)W_t = \theta_q(B)\Theta_Q(B^s)_{wt} \quad (2.12)$$

onde P apresenta a ordem do modelo AR sazonal, D indica o número de diferenças sazonais, Q refere-se à ordem do MA sazonal, s é a periodicidade, o w_t é o valor do ruído branco no período de tempo t e B é o operador de deslocamento em sentido contrário[41].

A Equação 2.12 representa os componentes sazonais de SARIMA que podem ser expandidos matematicamente após substituir o valor de $W_t = \nabla^d(B)\nabla_s^D(B)X_t$ [41].

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D X_t = \theta_q(B)\Theta_Q(B^s)_{wt} \quad (2.13)$$

Logo, os componentes do modelo SARIMA sazonal podem ser escritos como AR não sazonal(2.14), MA não sazonal (2.15), AR sazonal(2.16) e MA sazonal (2.17), conforme suas respectivas equações[41]:

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \dots - \phi_p B^p, \quad (2.14)$$

$$\theta_p(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \dots - \theta_q B^q, \quad (2.15)$$

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \Phi_3 B^{3s} - \dots - \Phi_P B^{Ps}, \quad (2.16)$$

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \Theta_3 B^{3s} - \dots - \Theta_Q B^{Qs} \quad (2.17)$$

e onde

$$B^s X_t = X_{t-s}, \quad (2.18)$$

$$\nabla_s X_t = \nabla_s(B)X_t = (1 - B^s)X_t = X_t - B^s X_t = X_t - X_{t-s}, \quad (2.19)$$

$$\nabla^d(B)X_t = (1 - B)^d X_t, \quad (2.20)$$

$$\nabla_s^D(B)X_t = (1 - B^s)^D X_t \quad (2.21)$$

Considerando a relação com os dados, SARIMA(p,d,q) x (P,D,Q), este tipo de modelo é usado com sucesso com várias séries temporais, pois a ordem do SARIMA é um número muito pequeno[41]. O valor do período da série temporal s (sazonalidade) é baseado no conjunto de dados. É por conta destes fatores que este formalismo torna-se uma ferramenta valiosa para a análise e previsão de séries temporais com sazonalidade, tendências e flutuações aleatórias. Ele oferece a possibilidade de modelar diferentes níveis de sazonalidade, permitindo a previsão de séries temporais em diversas áreas.

2.3.1.3 Suavização Exponencial

O Modelo de suavização ou ETS (Erro, Tendência e Sazonalidade) é um dos modelos mais utilizados para previsão de séries temporais. Este modelo foi proposto por Box e

Jenkins em 1976 e tem sido amplamente utilizado em várias áreas, como previsão financeira, econômica e de demanda de produtos. O modelo ETS é um modelo de espaço de estado que considera a presença de três componentes básicos em uma série temporal: erro, tendência e sazonalidade[12].

A primeira componente do modelo ETS é o erro, que representa a variação aleatória da série temporal. O erro é modelado como um processo estocástico que segue uma distribuição normal com média zero e variância constante[45]. A segunda componente é a tendência, que representa a mudança sistemática ao longo do tempo. A tendência pode ser linear, exponencial ou logística, dependendo do padrão de crescimento da série temporal[46]. A terceira componente é a sazonalidade, que representa a variação cíclica da série temporal. A sazonalidade pode ser mensal, trimestral ou anual, dependendo da frequência da variação cíclica[45].

O modelo ETS é estimado utilizando a abordagem de máxima verossimilhança, que busca encontrar os valores dos parâmetros do modelo que maximizam a probabilidade de observar os dados. Essa abordagem de máxima verossimilhança permite estimar os parâmetros do modelo de forma eficiente e fornecer uma medida de incerteza para as previsões[12].

O modelo ETS pode ser estendido para lidar com séries temporais que apresentam outras características, como não estacionariedade e outliers. Este formalismo pode ser utilizado para prever não apenas o valor futuro da série temporal, mas também as incertezas associadas a essas previsões, o que é fundamental para tomada de decisão em vários contextos práticos[45].

O método de suavização exponencial (ETS) pode ser utilizado para prever dados sem tendência clara ou padrão sazonal[47]. Este método revisa continuamente uma previsão, de acordo com informações recentes sobre os dados estimados. Desta forma, estes modelos se trata de medias ponderadas que atribuem um peso exponencialmente maior as observações de previsões recentes do que as mais antigas[47]. Ele foca nos três componentes que formam uma série temporal, (T), (S) e (E), mostrando que a relação entre eles é considerada aditiva 2.22 ou multiplicativa 2.23, embora ocorram algumas variações 2.24[46].

$$y = T + S + E \quad (2.22)$$

$$y = T X S X E \quad (2.23)$$

$$y = (T + S) X E \quad (2.24)$$

A classificação destes modelos é iniciada com o componente de tendência, que se trata de uma combinação do um termo de nível(l) e o um de crescimento(b). Estes podem ser combinados formando cinco tipos de tendências futuras[46]. Seja T_h denotando a tendência prevista nos próximos h períodos e Φ denotando um parâmetro de amortecimento ($0 < \Phi < 1$), temos os cinco tipos de tendência a seguir[46]:

$$\text{Nenhum} : T_h = l \quad (2.25)$$

$$\text{Aditivo} : T_h = l + bh \quad (2.26)$$

$$\text{AditivoAmortecido} : T_h = l + (\Phi + \Phi^2 + \dots + \Phi^h)b \quad (2.27)$$

$$\text{Multiplicativo} : T_h = lb^h \quad (2.28)$$

$$\text{MultiplicativoAmortecido} : T_h = lb^{(\Phi + \Phi^2 + \dots + \Phi^h)} \quad (2.29)$$

Após definido um componente de tendência, são introduzidos um componente sazonal e um componente de erro, de forma aditiva ou multiplicativa[46]. Caso o erro seja ignorado, teremos quinze métodos de suavização exponencial envolvendo a relação entre tendência e componentes sazonais[46].

Tabela 1 – Modelos ETS com diferentes tipos de tendência e sazonalidade de acordo com HYNDMAN [46].

Tendência	Sazonalidade		
	N (Nenhum)	A (Aditivo)	M (Multiplicativo)
N (Nenhum)	NN	NA	NM
A (Aditivo)	AN	AA	AM
Ad (Aditivo amortecido)	AdN	AdA	AdM
M (Multiplicativo)	MN	MA	MM
Md (Multiplicativo amortecido)	MdN	MdA	MdM

2.3.1.4 Holt-Winters

O Modelo Holt-Winters é um método de previsão de séries temporais que combina a suavização exponencial simples, a suavização exponencial dupla e a suavização exponencial tripla. Foi proposto por Charles Holt e Peter Winters em 1960 como uma extensão da suavização exponencial simples para lidar com séries temporais que apresentam tendência e sazonalidade[48]. Este pode ser aditivo de Holt-Winters (com tendência e sazonalidade

aditiva) e multiplicativo de Holt-Winters (com tendência aditiva e sazonalidade multiplicativa), o qual é considerado um dos modelos utilizados na suavização exponencial desconsiderando o erro[11].

A suavização exponencial simples (SES) é um dos métodos mais simples e populares para previsão de séries temporais, baseado em uma média móvel ponderada dos dados históricos, que atribui mais peso aos dados mais recentes[49]. A suavização exponencial dupla (DES) incorpora a tendência na previsão, permitindo que o modelo preveja um futuro mais distante, enquanto a suavização exponencial tripla (TES) leva em consideração a sazonalidade nos dados, permitindo que o modelo capture os padrões cíclicos da série temporal[50].

O modelo Holt-Winters é amplamente utilizado em diversas áreas, como finanças, economia e previsão de demanda, devido à sua simplicidade e eficácia em prever séries temporais com tendência e sazonalidade[11]. O modelo também permite a incorporação de outras variáveis exógenas que podem influenciar a série temporal, tornando-o ainda mais versátil e útil em diversas aplicações[51].

Uma das principais vantagens do Modelo Holt-Winters é sua capacidade de prever com precisão séries temporais que apresentam tendência e sazonalidade, como dados de vendas ou de tráfego na web[49]. Além disso, o modelo é capaz de lidar com outliers e dados faltantes, o que é uma vantagem significativa em relação a outros modelos de previsão de séries temporais[52]. Sua simplicidade e capacidade de incorporar variáveis exógenas o tornam uma escolha popular em diversas áreas, como finanças e previsão de demanda.

2.4 Medidas de desempenho

Existem várias medidas de desempenho de previsão, dependendo do tipo de problema de previsão e do objetivo específico da previsão. Nesta seção estão as medidas de desempenho de previsão que serão abordadas neste trabalho.

2.4.1 Critério de Informação de Akaike (AIC)

O critério de informação de Akaike (AIC) continua sendo uma medida amplamente utilizada para comparar a qualidade de ajuste de modelos candidatos em diversas áreas, incluindo a bioestatística e a modelagem de sistemas dinâmicos[53, 54]. O AIC é uma medida relativa de qualidade de ajuste que considera a adequação do modelo aos dados e a complexidade do modelo[54].

O AIC é calculado com base na função de verossimilhança do modelo e na quantidade de parâmetros estimados para ajustar o modelo aos dados[53]. Logo, ele penaliza a complexidade do modelo, o que significa que um modelo com muitos parâmetros terá um valor de AIC maior do que um modelo com menos parâmetros, mesmo que ambos ajustem bem os dados[53].

A fórmula para o cálculo do AIC é a seguinte[11]:

$$AIC = T \log \left(\frac{SSE}{T} \right) + 2(k + 2) \quad (2.30)$$

onde T é o número de observações usadas para estimativa e k é o número de preditores no modelo. A última parte da equação ocorre porque existem $k + 2$ parâmetros no modelo: o k coeficientes para os preditores, o intercepto e a variância dos resíduos. A ideia aqui é penalizar o ajuste do modelo (SSE) com o número de parâmetros que precisam ser estimados.

O modelo com o menor valor de AIC é geralmente considerado o melhor modelo para descrever os dados observados, embora outras medidas de ajuste de modelo também possam ser usadas em conjunto com o AIC[47]. Então, o critério de informação de Akaike é uma ferramenta útil para selecionar o melhor modelo em análises estatísticas e modelagem de sistemas complexos.

2.4.2 Teste de Shapiro-Wilk

O teste de Shapiro-Wilk é um método amplamente utilizado para verificar se uma amostra de dados segue uma distribuição normal. Esta verificação é importante para garantir que os testes estatísticos aplicados sejam adequados e confiáveis. O teste de Shapiro-Wilk é baseado na estatística W , que mede a discrepância entre a distribuição dos dados observados e a distribuição normal teórica[55].

Um fato importante evidenciado através de estudos recentes, é que o teste de Shapiro-Wilk tem baixa potência para detectar desvios da normalidade em amostras pequenas[56]. Portanto, é recomendado o uso de outros métodos gráficos, como o gráfico Q-Q plot, em conjunto com o teste de Shapiro-Wilk para avaliar a normalidade dos dados. Além disso, estudos apontam que o teste de Shapiro-Wilk é mais poderoso que outros testes de normalidade, como o teste de Kolmogorov-Smirnov, para amostras com até 50 observações[57]. No entanto, o teste de Shapiro-Wilk pode levar à rejeição da hipótese nula mesmo em casos em que os dados são aproximadamente normais. Portanto, é importante que os pesquisadores avaliem os resultados do teste em conjunto com outras análises exploratórias de dados[58].

Em algumas áreas de pesquisa, como a biologia e a ecologia, é comum lidar com amostras de dados que não seguem uma distribuição normal. Nesses casos, o uso de testes não paramétricos pode ser mais apropriado do que o uso de testes estatísticos paramétricos que assumem normalidade, como o teste t de Student[59]. Com isso, verifica-se que o teste de Shapiro-Wilk é uma ferramenta importante para avaliar a normalidade dos dados, considerando avaliar os resultados do teste em conjunto com outras análises exploratórias de dados.

2.4.3 Root Mean Squared Error (RMSE)

O RMSE (*Root Mean Squared Error*) ou raiz do erro médio quadrático, é uma medida amplamente utilizada para avaliar o desempenho de modelos de previsão em diversas áreas, como estatística, aprendizado de máquina e ciência de dados. Ele é considerado uma das principais métricas de erro para modelos de regressão, pois fornece uma medida da variação dos erros de previsão em relação aos valores observados[60].

A fórmula para calcular o RMSE é relativamente simples, sendo a raiz quadrada da média dos erros ao quadrado. Envolve a subtração do que aconteceu, y_t , do valor que foi projetado, \hat{y}_t , tendo como resultado o erro, e n é o número de períodos considerados[61], conforme apresentado na Equação 2.31.

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (y_t - \hat{y}_t)^2}{N}} \quad (2.31)$$

O benefício de usá-lo é que a escala de erro muda para corresponder a unidade de medida ideal de trabalho, quando calculada a raiz quadrática, possibilitando avaliar o poder de uma previsão de um conjunto de dados[62]. Um RMSE baixo, indica que o modelo teve um bom desempenho[61]. A raiz quadrada é aplicada para garantir que a medida esteja na mesma unidade dos dados originais, o que facilita sua interpretação[63]. Além disso, como o RMSE penaliza mais os erros maiores, ele é particularmente útil em casos em que esses erros são considerados mais críticos.

2.4.4 Mean Absolute Percentage Error (MAPE)

O MAPE (*Mean Absolute Percentage Error*) ou erro percentual absoluto médio, é uma medida de precisão comum em previsão, amplamente utilizada em diferentes áreas, incluindo finanças, economia e engenharia. O MAPE é uma métrica de desempenho que representa a magnitude do erro de previsão em relação ao valor real da variável prevista,

expresso como uma porcentagem. O cálculo do MAPE é uma forma de avaliar a precisão das previsões, pois ele leva em conta tanto o tamanho absoluto quanto a direção do erro.

Desta forma, ele mede o erro absoluto percentual médio, dividindo-se a diferença entre o valor observado e o valor previsto (resultado do erro) pelo valor observado[64] e eliminando-se o sinal. Seu cálculo está apresentado na Equação 2.32.

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (2.32)$$

Diante destes pontos, diversos estudos destacam a importância do MAPE como uma medida precisa e intuitiva de avaliar a acurácia de previsões em diferentes contextos. Trata-se de uma medida de erro que é facilmente interpretada, pois está expressa em uma escala percentual[65]. Ela não é sensível a valores extremos, o que a torna útil em previsões em que os erros podem variar significativamente[66].

Esta medida é amplamente utilizada na prática para avaliar a precisão das previsões em diferentes áreas. A escolha da medida de precisão de previsão deve levar em conta a natureza dos dados e os objetivos do estudo, e uma combinação de medidas pode ser mais apropriada para avaliar a acurácia de previsões em diferentes contextos[67]. Desta forma, o erro percentual absoluto médio (MAPE) é uma medida comum e intuitiva para avaliar a precisão de previsões.

2.4.5 Coeficiente de Theil

O coeficiente de Theil é uma medida de desigualdade que tem sido amplamente utilizada em análises estatísticas, tanto para a distribuição de renda como em outros campos. O coeficiente foi originalmente proposto por Theil (1967) como uma medida de desigualdade de renda baseada na teoria da informação. Desde então, tem havido muitos desenvolvimentos e aplicações do coeficiente.

Este coeficiente pode ser definido como uma medida de entropia relativa entre a distribuição de renda observada e uma distribuição de referência. Uma das principais vantagens do coeficiente de Theil em análises estatísticas é que ele é sensível às mudanças na forma da distribuição de renda[68]. Ele também pode ser utilizado em análises estatísticas para medir as desigualdades regionais[69]. Neste sentido, o coeficiente de Theil é particularmente útil para analisar as desigualdades regionais em países em desenvolvimento.

Esta medida permite uma comparação relativa de métodos formais de previsão com métodos simplistas[70]. Ela é definida de acordo com a Equação 2.33.

$$Theil = \frac{\sum_{t=1}^N (y_t - \hat{y}_t)^2}{\sum_{t=1}^N (y_t - y_{t-1})^2} \quad (2.33)$$

Se o coeficiente de Theil for igual a 0, significa que a previsão é perfeita. Se o coeficiente de Theil for igual a 1, indica que a precisão do método de previsão em estudo é igual à do método simplista, em que a previsão para o próximo instante coincide com a última observação da série. Se for maior que 1, significa que o modelo não é útil para fins preditivos[70]. Desta forma nota-se que esta medida de precisão relativa assume valores não negativos.

Portanto, este coeficiente é uma medida importante e amplamente utilizada em análises estatísticas sobre desigualdade. Tem sido usado em muitos estudos para analisar a desigualdade de renda em vários países[71]. Neste tipo de estudo, é destacado também a importância de levar em consideração a dinâmica temporal da desigualdade de renda ao usar o coeficiente de Theil em análises estatísticas.

3 Materiais e métodos

Trata-se de uma pesquisa quantitativa pela utilização de dados numéricos para encontrar os resultados através de análises estatísticas, e do tipo positivista devido a possibilidade das pessoas utilizarem o modelo que será criado para visualizar e analisar os dados compreendendo melhor os pontos estudados.

Quanto a sua natureza, trata-se de uma pesquisa aplicada, por proporcionar o desenvolvimento de soluções práticas aplicáveis para a sociedade e com análise de dados de forma descritiva, descrevendo e apresentado as características do fenômeno estudado. Esta pesquisa também é do tipo documental com relação aos seus procedimentos, tendo em vista que foram coletados imagens da plataforma USGS, com banco de dados disponíveis de forma gratuita na rede de internet.

3.1 Área de estudo

Este trabalho foi direcionado para coleta de imagens na região do reservatório de Kirwin, que fica no Condado de Phillips, Kansas, Estados Unidos. Ele está localizado próximo à cidade de Kirwin, com as coordenadas 39°39'22"N 99°09'40"W, no norte do Kansas[72]. Em suas margens, fica localizado o Refúgio Nacional de Vida Selvagem de Kirwin, um dos importantes locais de preservação da vida selvagem nos Estados Unidos[73].

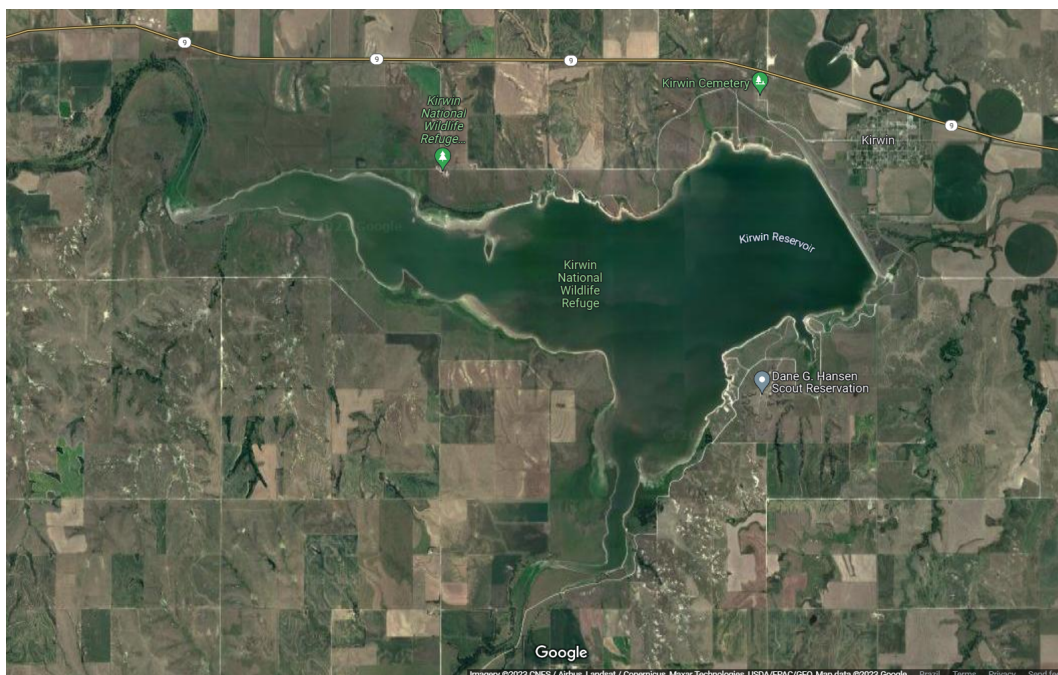


Figura 1 – Reservatório Kirwin, Kansas, EUA. Fonte:Google Maps.

O Departamento de recursos hídricos dos EUA o construiu e continua a operá-lo para fins de controle de enchentes e irrigação da região, com uma área construída de 20,55 km² com uma área de captação de 3.540 km²[73].

3.1.1 Conjunto de dados

O conjunto de dados utilizados tratam-se de imagens de satélite, indexadas no tempo, as quais se referem a região onde se encontra o corpo hídrico estudado. Destas imagens, foram extraídos dados de ocupação da área alagada através dos pixels que compõe o espelho d'água representados na imagem. Estes dados, foram adquiridos por meio de processamento digital de imagens realizado por um algoritmo desenvolvido.

3.2 Percurso metodológico

O percurso metodológico desenvolvido neste trabalho está apresentado em 5 etapas: (i) coleta de dados, (ii) processamento das imagens, (iii) cálculo da área do corpo hídrico, (iv) modelagem estatística e (v) avaliação de desempenho dos modelos. As etapas (ii), (iii), (iv) e (v) são realizadas com a utilização da linguagem R[74]. Elas estão apresentadas na Figura (2).

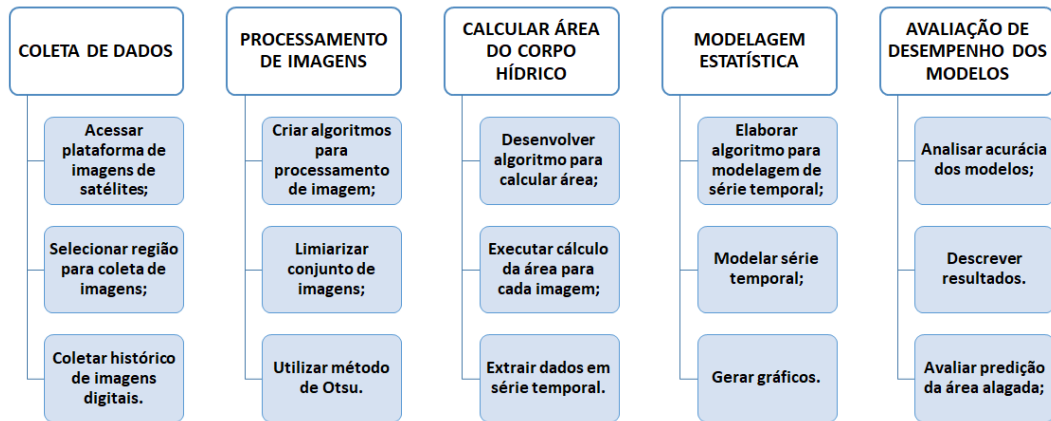


Figura 2 – Percurso Metodológico. Fonte: Autor

3.2.1 Coleta de dados

Atendendo ao Objetivo Específico 1, as imagens foram coletadas na plataforma da Agência Científica United States Geological Survey (USGS), a qual existe um catálogo de dados de imagens de satélite que é disponibilizado em um serviço de computação em nuvem. Isso possibilita o fácil acesso a grande parte dos conjuntos de dados de imagens utilizados em estudos geoespaciais, proporcionando análises de alto rendimento [75]. Estas imagens são obtidas inserindo parâmetros como coordenadas do local, conjunto de dados de satélite, percentual de cobertura e sombreamento de nuvens.

Para a seleção do conjunto de dados, foram utilizadas imagens do satélite Landsat retirados da coleção de produtos científicos, na seção dos dados de extensão dinâmica de água superficial. Para critérios adicionais, incluiu-se o percentual de até 10 por cento para cobertura e sombra de nuvens. Após inseridos estes parâmetros, é solicitado o resultado e na sequência o download da imagem em formato DIAG.TIF, ideal para o uso do método de Otsu já que está em escala de cinza. Devido as imagens terem uma dimensão de 5.000x5.000mm e conterem outros corpos hídricos, foram redimensionadas focando apenas no reservatório a ser estudado. A dimensão ideal para este estudo foi de 350x350mm, pois deixa todo o reservatório no centro da imagem, conforme ilustrado na Figura 3

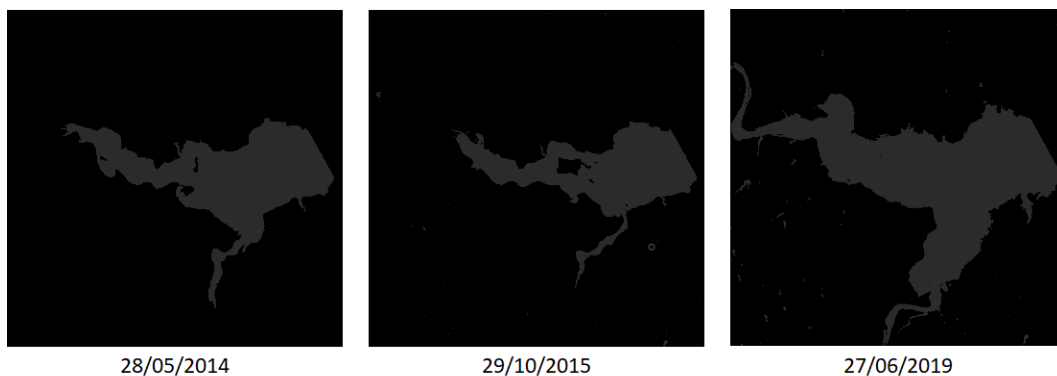


Figura 3 – Imagens redimensionadas do reservatório Kirwin adquiridas no site da USGS em momentos distintos. Fonte: USGS

Foram coletadas 119 imagens do reservatório Kirwin entre os anos de 2013 e 2022, dando condição para que os dados da área alagada desta região sejam estimados. Para isto, foi criado um algoritmo no software R, para se fazer o processamento das imagens.

3.2.2 Processamento das imagens

Na sequência, atendendo ao Objetivo Específico 2, foi criado um algoritmo no programa R 4.2.1[74], que se encontra livre para download. Para isto, as imagens foram salvas em um diretório no computador e processadas em um código de programação baseado no método de Otsu[10]. Devido a necessidade de processamento em massa para determinar os pixels que correspondem o continente e os que determinam a ocupação da água na área em estudo, foi criada uma função a qual aplica-se a limiarização da imagem através do método de Otsu.

O primeiro passo foi carregar estas imagens através da leitura de uma pasta específica no computador. Nesta encontram-se as imagens referidas ao satélite escolhido para estudo e com as referida região onde o corpo hídrico estará situado. Algumas informações adicionais com relação aos parâmetros de coleta das imagens foram separados também neste mesmo diretório, como por exemplo nível de cobertura de nuvem e banda espectral utilizada para melhor definir a área do objeto em estudo.

Em segundo passo, o algoritmo leu todas as imagens e organizou em uma lista. Este ponto é importante para que o processamento das imagens fique organizado, obtendo assim os dados de forma cronológica e com cada informação na sua respectiva imagem.

O último passo desta etapa foi a realização da limiarização da imagem. Nesta, foram

utilizados os conceitos do método de Otsu para realizar a separação dos pixels que representam o continente e os que representam o corpo hídrico em estudo. Os pixels brancos tem o valor de 1 e os pretos tem o valor de 0, os quais são representados respectivamente o corpo hídrico e o continente. Estas etapas podem ser apresentadas no fluxograma da Figura 4.

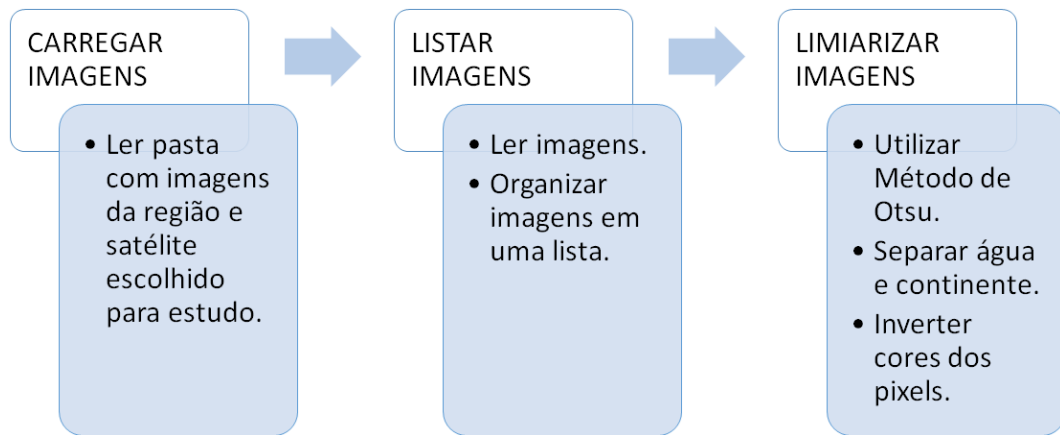


Figura 4 – Fluxograma do algoritmo para processamento das imagens. Fonte: Autor.

3.2.3 Calcular área do corpo hídrico

Para determinar a área de ocupação do corpo hídrico, atendendo o Objetivo Específico 3, foi necessário desenvolver um algoritmo para verificar a quantidade de pixels brancos e multiplicar este valor pelo número que representa a resolução da imagem. Estas etapas podem ser apresentadas no fluxograma da Figura 5.

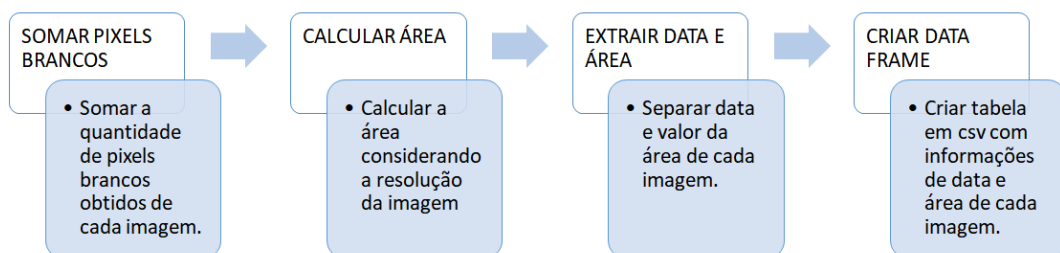


Figura 5 – Fluxograma do algoritmo para calcular área do corpo hídrico. Fonte: Autor.

Levando em consideração que a resolução da imagem da nossa pesquisa é de 30m, 1 pixel tem uma área quadrada de 900m². Logo, o somatório dos pixels multiplicado por este valor, leva à estimativa da área alagada do corpo hídrico que está representada na Figura6.

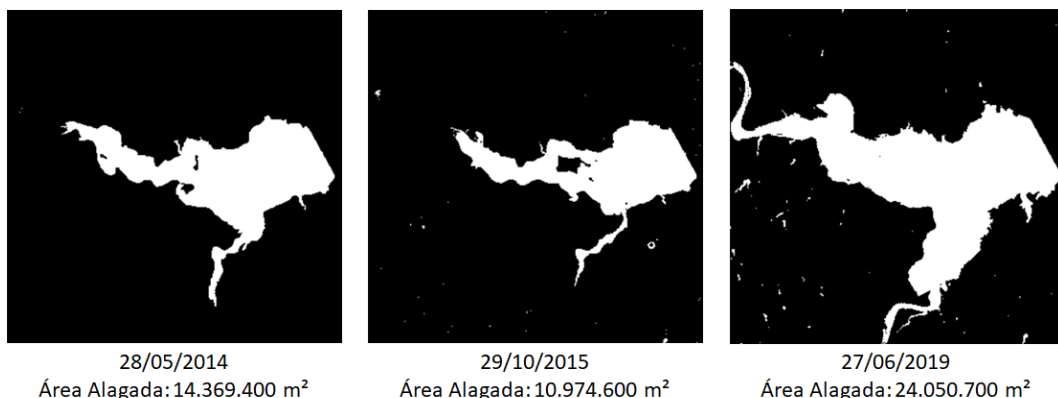


Figura 6 – Imagens do reservatório Kirwin após aplicação do método de Otsu em momentos distintos. Fonte: Autor.

Na sequência, todas as imagens foram lidas e organizadas em uma lista com informações cronológicas de data e valor de área, as quais permitiram a criação de um *dataframe* que fica disponível para as próximas etapas. Com a aplicação do código nas imagens adquiridas, foi possível inferir o valor de área alagada em metros quadrados, de forma mensal, com início em janeiro de 2013 e término em novembro de 2022, possibilitando a criação de uma série temporal.

3.2.4 Modelagem de série temporal

Para atender ao Objetivo Específico 4, foi adicionado ao algoritmo de processamento de imagem um novo script em linguagem R para realizar as análises estatísticas de séries temporais. Foi realizada a transformação dos dados em série temporal e uma decomposição desta para início das análises da área alagada em estudo. Isso possibilitou uma maior detalhe para análise dos seus componentes.

Devido a esta necessidade de avaliação de uma série temporal para gerar uma previsão dos dados de área alagada do corpo hídrico em estudo, foram construídos modelos para melhor demonstrá-los. Os modelos utilizados foram o ARIMA, SARIMA, ETS e Holt-Winters. Estas etapas podem ser apresentadas no fluxograma da Figura 7.

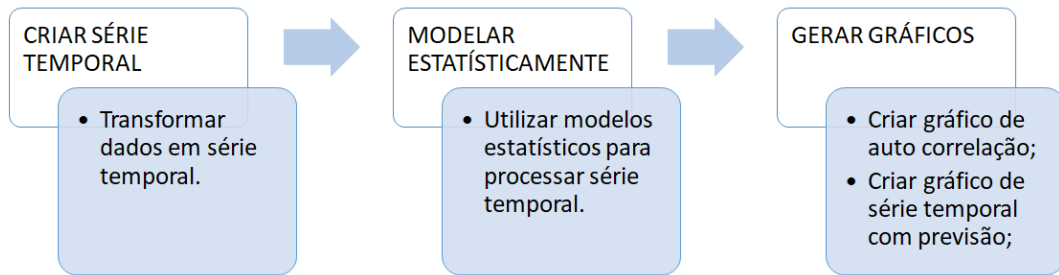


Figura 7 – Fluxograma do algoritmo para modelagem estatística de série temporal. Fonte: Autor.

Em busca desta modelagem para se ter uma previsão de alterações na área alagada do corpo hídrico em estudo, foram separados 107 dados para treinamento e 12 dados para teste. Desta forma pode-se avaliar o quão próximo os modelos a serem testados chegam do resultado real. Na sequência, foram desenvolvidos gráficos que auxiliaram na análise dos resultados e avaliações futuras das alterações de área na região em estudo.

3.2.5 Avaliar desempenho dos modelos

Atendendo o Objetivo Específico 5, foi feita uma avaliação de desempenho de todos os formalismos utilizados na etapa de modelagem dos dados das áreas alagadas das imagens através de uma avaliação detalhada dos seus resíduos. Isto se fez necessário para possibilitar a criação de uma previsão de possíveis alterações na área alagada do corpo hídrico estudado, por meio do melhor modelo analisado. Estas etapas podem ser apresentadas no fluxograma da Figura 8.

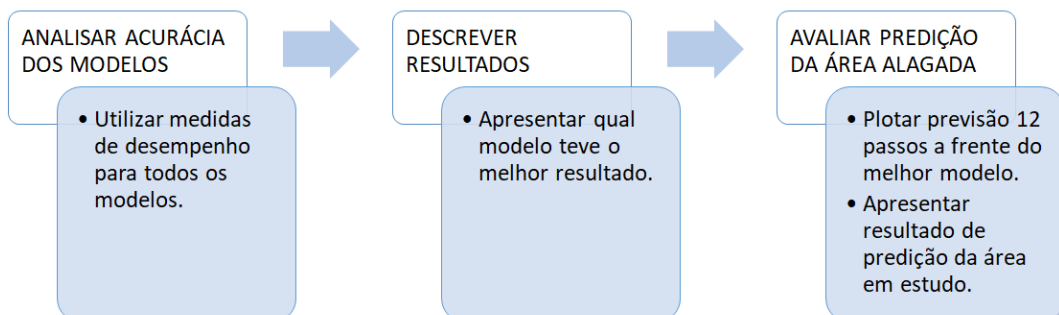


Figura 8 – Fluxograma para avaliar o desempenho dos modelos. Fonte: Autor.

Após isso, consolida-se os dados em estudo e plota-se uma previsão 12 passos à frente com os parâmetros do melhor modelo analisado para identificação das possíveis alterações da área alagada, permitindo a avaliação de alterações da área alagada em estudo.

4 Resultados e Discussão

Analisando a série mensal da área alagada do corpo hídrico de uma forma detalhada, foi criado uma box plot com a evolução anual que está representada na Figura 9.

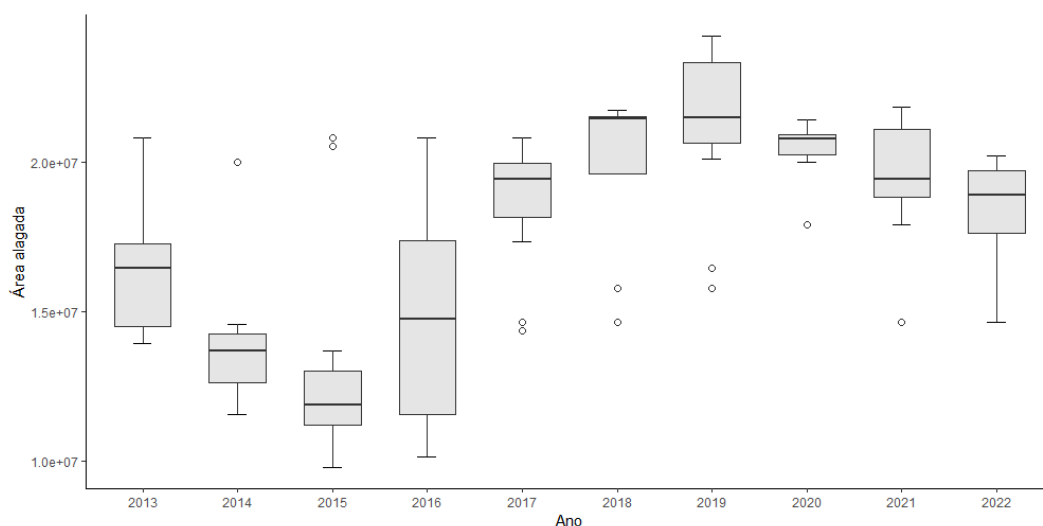


Figura 9 – Box plot com evolução anual da área alagada do reservatório Kirwin. Fonte: Autor.

Nota-se que a partir de 2016, há uma maior variação, seguida por um considerável aumento de sua mediana. A partir de 2017 até 2019 ocorre um aumento considerável da mediana, evidenciando uma alteração significativa na área alagada neste período. Em 2020, nota-se a menor variabilidade da área alagada e o início de uma nova redução da mediana dos dados, seguidos pelos anos de 2021 e 2022, os quais demonstram também uma variabilidade menor, em relação ao ano de 2016. Os pontos fora da curva, apresentam em maior quantidade nos anos de 2015, 2017, 2018 e 2019, consequência de uma evolução no aumento da área alagada do corpo hídrico.

Na sequência, conforme Figura 10, os dados são colocados em um gráfico de linha para a identificação de mais alguns padrões não aleatórios desta série. Desta forma, observa-se o comportamento durante o período avaliado, onde pode-se criar previsões futuras de aumento ou diminuição da área alagada na região estudada.

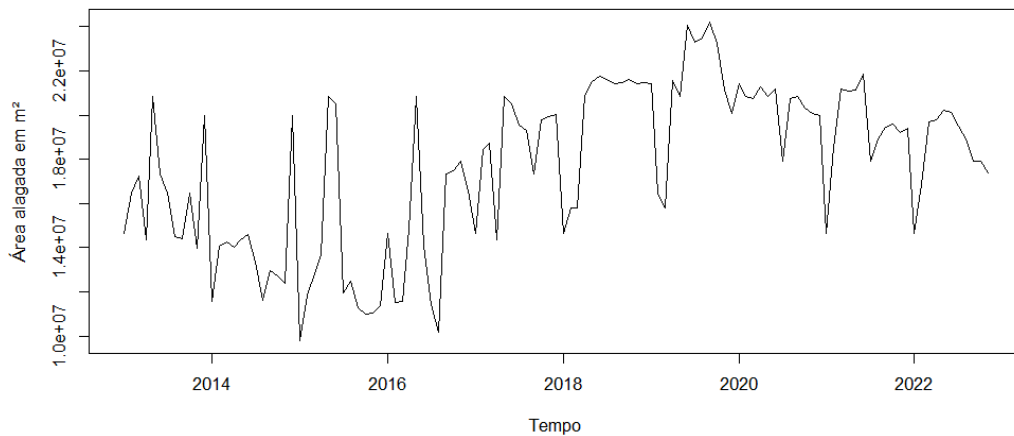


Figura 10 – Série Temporal da área alagada do lago Kirwin. Fonte: Autor.

Observa-se uma tendência crescente média na área alagada no período avaliado, porém estes dados ainda podem conter algumas medidas que serão influenciadas pelo pico da distribuição e ainda não é possível avaliar. Para reduzir este efeito, foi feita uma transformação logarítmica nos dados, conforme Figura11.

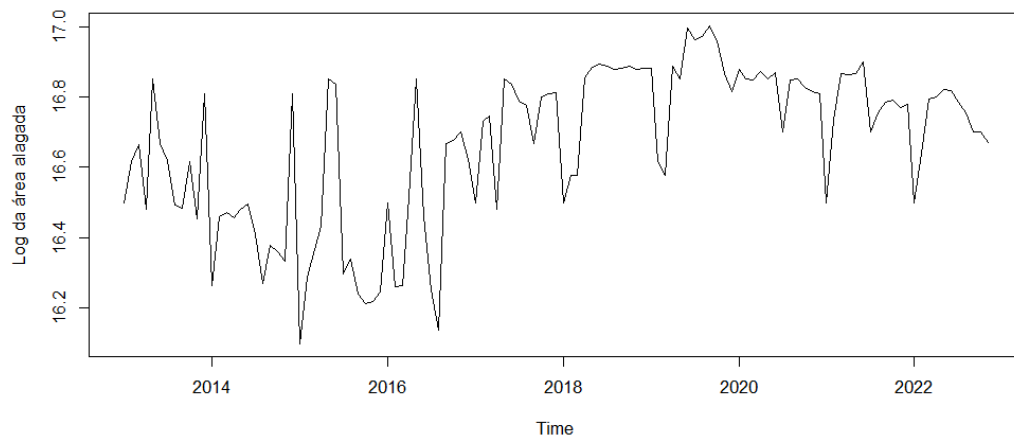


Figura 11 – Série Temporal da área alagada do lago Kirwin após transformação logarítmica. Fonte: Autor.

Para uma melhor avaliação da série temporal, foram analisados os seus padrões de forma decomposta, através de uma decomposição de seus componentes que está representada na Figura 12.

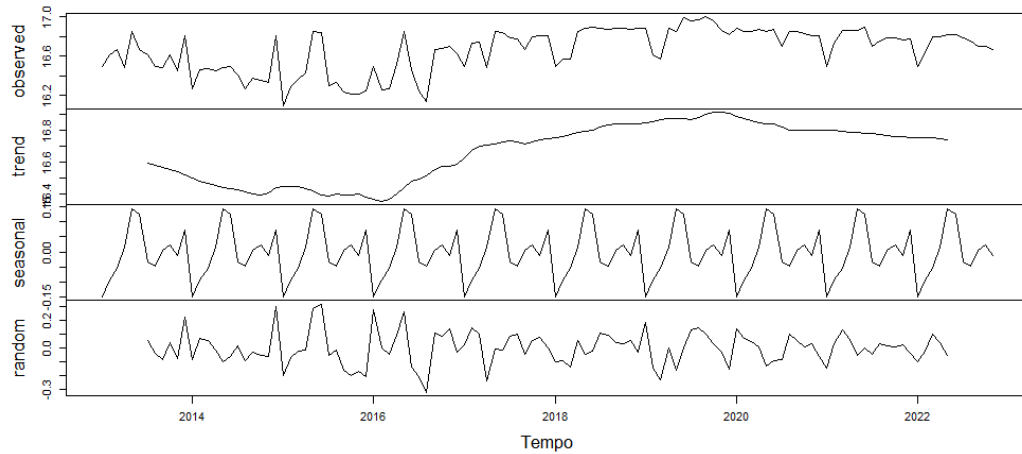


Figura 12 – Decomposição da Série Temporal das áreas alagadas do lago Kirwin. Fonte: Autor.

A Figura 12 mostra uma representação dos dados observados que se configura a série temporal observada. Em seguida, foi extraída a tendência desta série a qual mostra uma tendência de crescimento da área alagada no período de 2016 a 2020. Na sequência, nota-se a sazonalidade anual existente nesta série, refletindo a influência de um determinado fator externo. A seção final, ilustra o comportamento errático provocado por interrupções na série em questão.

Na sequência, a Figura 13 mostra exatamente a separação dos dados em treinamento e teste, onde os valores de área alagada separados para teste estão após a linha vermelha da série temporal estudada. Os modelos são treinados apenas com os dados a esquerda da reta vertical.

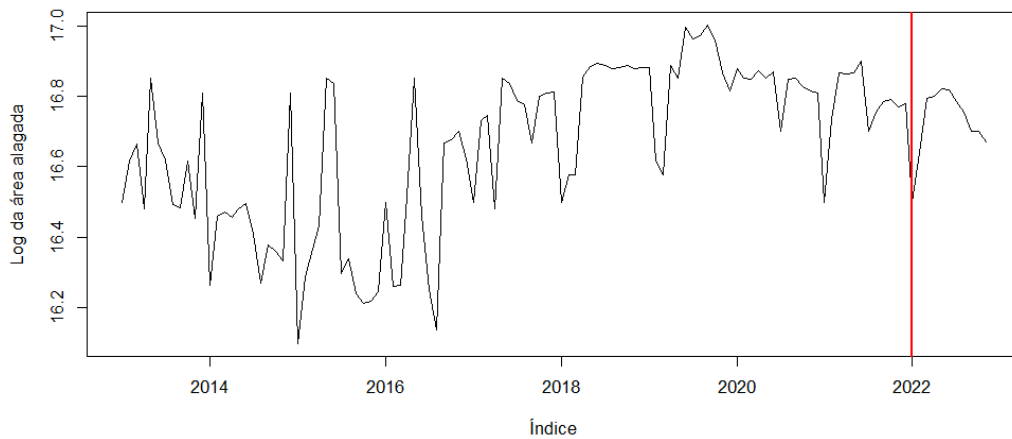


Figura 13 – Separação de dados de treinamento e teste. A linha vertical em vermelho separa os dados de treinamento (à esquerda) e de teste (à direita). Fonte: Autor.

4.1 Modelagem e Previsão

Em análise preliminar do modelo ARIMA executando a função *autoarima* para os dados de treinamento, foi gerado o melhor modelo ARIMA que foi o de ordem (0,1,2), apresentando um menor AIC de -79,15. Porém, ao avaliar seus resíduos, que se trata da diferença entre as séries original e a ajustada[76], foi identificado que ele poderia ser melhorado devido a existência de uma autocorrelação significativa em 12 como evidenciado no autocorrelograma da Figura 14.

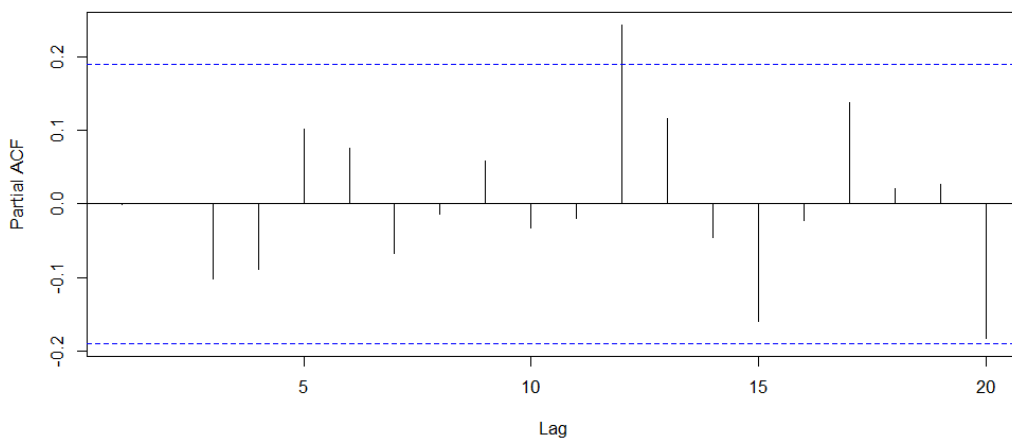


Figura 14 – Autocorrelação parcial dos resíduos do modelo ARIMA. Fonte: Autor.

Desta forma, alguma informação relevante está ainda contida nos resíduos, o qual sua projeção pode ser melhorada. Como este modelo não apresentou sazonalidade, foi necessário trabalhá-lo de uma forma melhor para apresentar uma sazonalidade que não estava sendo identificada. Logo, foi considerada uma sazonalidade 12 e avaliado o melhor modelo sazonal executando novamente a função autoarima. O melhor modelo testado foi o SARIMA (0,1,2) (1,0,0)12 o qual apresentou um AIC de -83,9, sendo o menor entre todos os modelos sazonais avaliados nesta etapa.

Devido a isto, utilizou-se também o modelo SARIMA, que apresentou também uma autocorrelação significativa no seu correlograma. Em virtude disto e devido a identificação de sazonalidade, foi incluído nas análises o terceiro modelo chamado de ETS e um quarto modelo, chamado Holt-Winters. A partir de então, todas as análises e resultados foram baseados na avaliação dos resíduos destes modelos, com o intuito de definir qual é o melhor entre eles, para possibilitar a criação de uma previsão da área alagada naquela região. A Tabela 2 mostra o resultado dos parâmetros encontrados para cada modelo avaliado.

Tabela 2 – Resultado dos modelos para a série temporal da área alagada.

MODELO	DESCRIÇÃO
ARIMA (0,1,2)	ma1=-0,595 ma2=-0,185
SARIMA (0,1,2) (1,0,0)12	ma1=-0,623 ma2=-0,166 sar1=0,261
ETS (A, N, A)	Alpha=0,293 l=16,68 Sigma=0,148
Holt-Winters	Alpha = 0,239 beta = 0,060 gama = 0,255
	a=16,759 b= -0,002
	s1=0.0745 s2=-0.1294 s3=-0.0630 s4=-0.0284
	s5=0.0411 s6=0.1162 s7=0.1261 s8=-0.0040
	s9=0.0009 s10=0.0231 s11=0.0367 s12=-0.0065

Com o intuito de melhorar as análises com o objetivo de encontrar o melhor modelo, foram utilizados os correlogramas para analisarmos a aleatoriedade dos resíduos para cada modelo avaliado. Na Figura 15, é apresentado o correlograma de autocorrelação parcial dos resíduos dos modelos analisados.

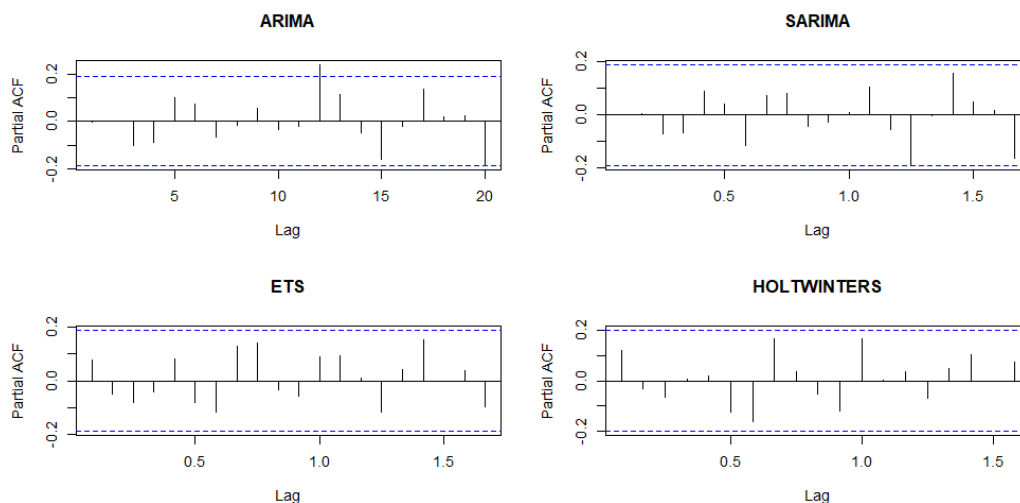


Figura 15 – Autocorrelação parcial dos resíduos dos modelos utilizados. Fonte: Autor.

Ao analisar estes gráficos, nota-se que existe uma correlação significativa para os modelos ARIMA e SARIMA. Este também apresenta um resultado interessante para as análises pelo fato dos dados dispostos para avaliação dos modelos ETS e Holt-Winters, estarem dentro dos limites de especificação.

Em seguida, foi realizado o teste de normalidade para saber se os resíduos de treinamento podem estar normalmente distribuídos. Isso foi possível através do teste de Shapiro-Wilk, o qual retornou o valor de p para cada modelo, que estão apresentados na Tabela 3.

Tabela 3 – Resultado do teste de Shapiro-Wilk para os resíduos de treinamento, onde estão apresentados o valor de p de cada modelo.

MODELO	VALOR DE p
ARIMA (0,1,2)	0,057
SARIMA (0,1,2) (1,0,0) ₁₂	0,212
ETS (A, N, A)	0,045
Holt-Winters	0,045

Considerando o nível de significância de 0,05, nota-se que os pressupostos teóricos dos testes de hipóteses foram satisfatórios para os modelos.

Para se ter uma melhor análise destes pontos, foram geradas as previsões e comparadas com os dados separados para teste, através de uma análise baseada em uma previsão com horizonte de tempo igual a 12 passos à frente, conforme apresentado na Figura 16.

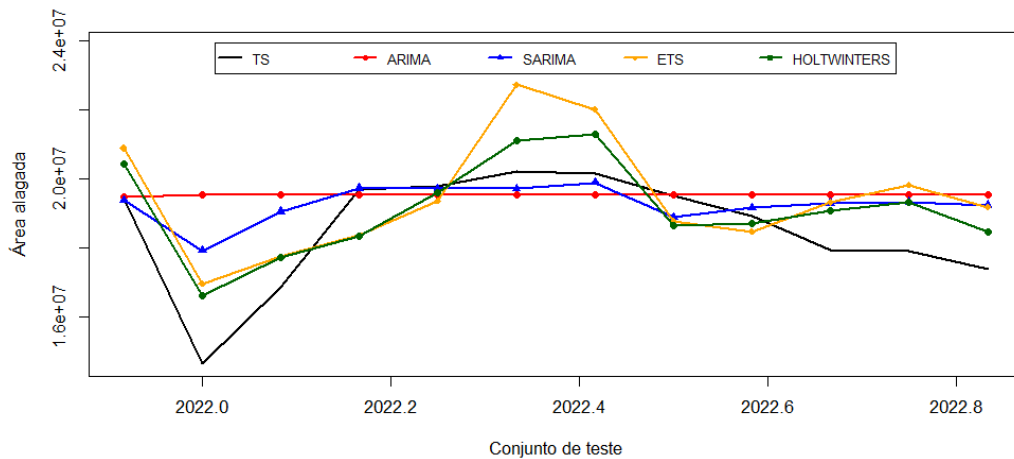


Figura 16 – Comparativo das previsões dos modelos avaliados com os dados de teste.

Como observado na Figura 16, as previsões geradas pelos modelos SARIMA e Holt-Winters, performaram próximos aos dados de teste da série temporal em estudo. Na Figura 17, é apresentado através do boxplot a previsão dos modelos em comparação as observações da série temporal real.

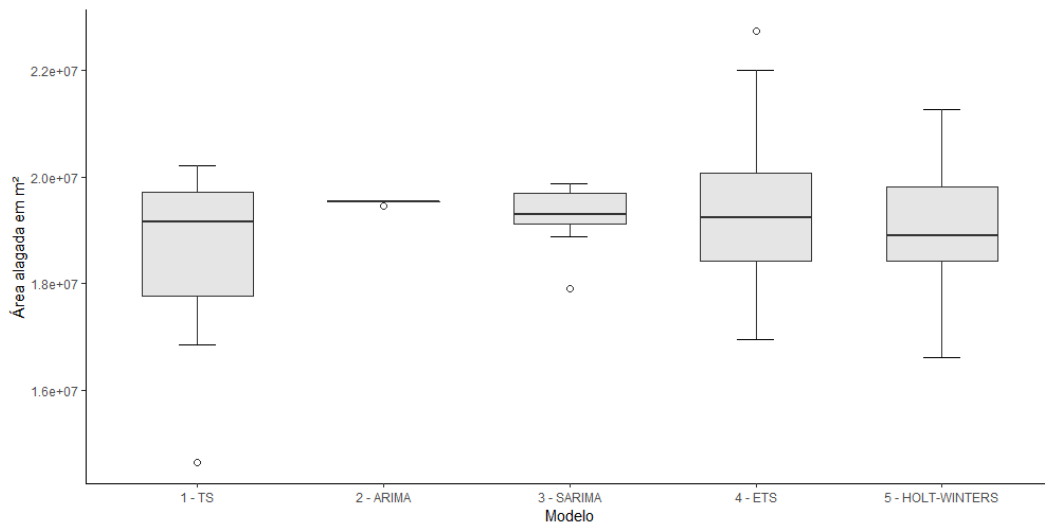


Figura 17 – Boxplot comparando as previsões dos modelos avaliados com os dados de teste.

Para analisar com detalhes a acurácia dos modelos avaliados, foram utilizadas algumas medidas apresentadas na Tabela 4 para demonstrar qual o modelo teve melhor performance.

Tabela 4 – Medidas de avaliação dos modelos de previsão estudados.

MODELO	RMSE (km ²)	MAPE (Percentual)	COEFICIENTE DE THEIL
ARIMA	1.879.976	7,672	1,126
SARIMA	1.400.625	5,872	0,625
ETS	1.573.033	7,868	0,730
Holt-Winters	1.118.389	5,669	0,371

Entre os modelos avaliados, o que apresenta uma melhor medida de precisão para modelos de previsão de series temporais da área alagada do corpo hídrico estudado é o Holt-Winters, por que apresenta os menores valores para as medidas de desempenho avaliadas. Um destaque é que seu RMSE representa 7,77% da amplitude dos dados, evidenciando ser o modelo com melhor desempenho relativo.

Desta forma, foram considerados todos os dados da série temporal e criado uma previsão baseada no melhor modelo avaliado pelas medidas de acurácia. A Figura 18 mostra a previsão final considerando um horizonte de tempo de 12 passos à frente.

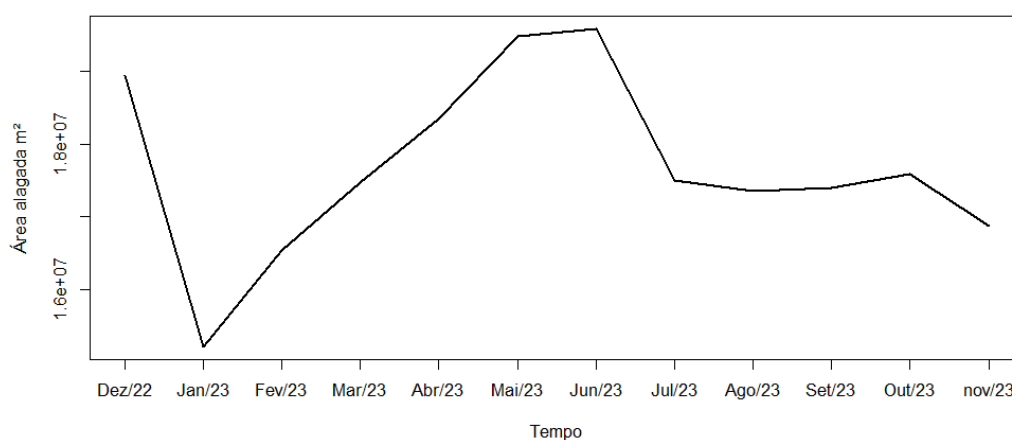


Figura 18 – Previsão final da área alagada considerando 12 meses à frente.

É possível notar que os dados de área alagada seguem semelhantes aos últimos dados coletados nos últimos 12 meses. Com isso, define-se que há evidências de uma baixa possibilidade de ocorrer aumento da área alagada para o próximo ano na região do lago Kirwin. Um destaque desta imagem está no mês de janeiro de 2023 que demonstra uma baixa significativa na área alagada devido a dinâmica da região do lago Kirwin, que nesta época do ano passa por um período de nevasca, tendo grande parte de sua área coberta por gelo.

5 Considerações Finais

Neste trabalho sugeriu-se um algoritmo para o mapeamento e previsão da variação da área alagada de um corpo hídrico. Tomou-se como caso de estudo o lago Kirwin (USA), com série de imagens de satélite Landsat 8, de janeiro de 2013 a novembro de 2022. Neste sentido, destaca-se que encontrar uma quantidade razoável de imagens indexadas no tempo e de boa qualidade mostrou-se como um dos principais desafios de aplicação do método proposto.

Após processamento das imagens, identificou-se uma sazonalidade anual, evidenciada ao decompor a série temporal e no correlograma. Baseando-se nisso, modelos que incorporam sazonalidade apresentaram previsões atraentes para futuras alterações da área alagada do corpo hídrico. Com base nos resultados obtidos neste estudo, é possível sinalizar que a modelagem de séries temporais utilizando o modelo de Holt-Winters é uma técnica promissora para previsão da área alagada em corpos hídricos. A avaliação da medida de desempenho do modelo indicou que ele é capaz de produzir previsões mais próximas da realidade, o que pode auxiliar governantes na tomada de decisões em relação a possíveis cheias ou escassez de água.

Diante dos resultados encontrados, é fundamental que o poder público esteja atento a essas informações e utilize-as para planejar e executar ações que minimizem os impactos das enchentes e secas em regiões afetadas. Uma das medidas que podem ser adotadas é a implantação de sistemas de alerta que permitam a previsão e acompanhamento de eventos climáticos extremos. Além disso, é importante que sejam realizados investimentos em pesquisa e desenvolvimento de novas tecnologias e técnicas para previsão destes eventos. Nesse sentido, uma possibilidade interessante é o uso de outros modelos de previsão de séries temporais em outros tipos de corpos hídricos, para aprimorar ainda mais a precisão das previsões, contribuindo para uma gestão mais eficiente dos recursos hídricos.

Por fim, é importante ressaltar que a gestão dos recursos hídricos é uma questão complexa e que exige esforços conjuntos de governos, sociedade civil e setor privado. É fundamental que sejam adotadas medidas de planejamento e gestão sustentáveis, que considerem aspectos socioeconômicos, ambientais e culturais. A utilização de técnicas de previsão mais precisas e eficientes, como o modelo de Holt-Winters, pode ser um importante aliado nesse processo, contribuindo para uma gestão mais eficiente e sustentável dos recursos hídricos.

Referências

- [1] E. Intrieri, F. Raspini, A. Fumagalli, P. Lu, S. Del Conte, P. Farina, J. Allievi, A. Ferretti, and N. Casagli, “The maoxian landslide as seen from space: detecting precursors of failure with sentinel-1 data,” *Landslides*, vol. 15, no. 1, pp. 123–133, 2018.
- [2] Z. Lei and Q. Weng, “Impervious surface estimation by integrated use of landsat and modis time series in wuhan, china,” *Remote Sensing Time Series Image Processing*, pp. 143–156, 2018.
- [3] Y. Xu, Zhihuo e Wang, “Análise de séries temporais de imagens de satélite de radar para mapeamento de alta resolução de mudanças florestais provocadas pelo homem em chongming eco-island,” *Detecção Remota*, vol. 12, 2020.
- [4] S. I. Khan, Y. Hong, J. J. Gourley, M. U. Khattak, and T. De Groeve, “Multi-sensor imaging and space-ground cross-validation for 2010 flood along indus river, pakistan,” *Remote Sensing*, vol. 6, no. 3, pp. 2393–2407, 2014.
- [5] N. A. Muhadi, A. F. Abdullah, S. K. Bejo, M. R. Mahadi, and A. Mijic, “Image segmentation methods for flood monitoring system,” *Water*, vol. 12, no. 6, p. 1825, 2020.
- [6] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Upper Saddle River, NJ: Prentice Hall, 2008.
- [7] H. I. A. Bustos, D. N. Dantas, and D. J. Aloise, “Método de estimativa e previsão de movimento em séries temporais de imagens,” *Brazilian Journal of Development*, vol. 6, no. 9, pp. 72476–72483, 2020.
- [8] P. Morettin and C. Toloí, *Análise de séries temporais: modelos lineares univariados*. BLUCHER., 2018.
- [9] S. Zhou, P. Kan, J. Silbernagel, and J. Jin, “Application of image segmentation in surface water extraction of freshwater lakes using radar data,” *ISPRS International Journal of Geo-Information*, vol. 9, no. 7, p. 424, 2020.
- [10] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [11] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2018.

- [12] G. E. Box, G. M. Jenkins, G. C. Reinsel, and L. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [13] R. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder, *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media, 2008.
- [14] J. D. Sachs, *Dez desafios para o desenvolvimento sustentável*. São Paulo: Editora Cia das Letras, 2019.
- [15] S. Lele, “Reconciling sustainability and development: reflections on the concept of sustainable development,” *Sustainability Science*, vol. 14, no. 5, pp. 1435–1444, 2019.
- [16] UNDESA, *World Economic and Social Survey 2018: Frontier Issues in Economic Development*. New York: United Nations, 2018.
- [17] IPCC, “Global warming of 1.5c,” *Special Report*, 2018.
- [18] UNESCO, *Educação para o Desenvolvimento Sustentável: Objetivos de aprendizagem*. 2018.
- [19] F. Biermann and J. Gupta, *The Politics of the Global Environment: Issues, Strategies, and Actors*. Cambridge: MIT Press, 3 ed., 2018.
- [20] S. W. S. WCED, “World commission on environment and development,” *Our common future*, vol. 17, no. 1, pp. 1–91, 1987.
- [21] J. Elkington, “Cannibals with forks. the triple bottom line of the 21st century. capstone,” 1997.
- [22] I. Sachs, *Caminhos para o desenvolvimento sustentável*. Editora Garamond, 2000.
- [23] G. Kazakeviciute-Januskeviciene, E. Janusonis, R. Bausys, T. Limba, and M. Kiskis, “Assessment of the segmentation of rgb remote sensing images: A subjective approach,” *Remote Sensing*, vol. 12, no. 24, p. 4152, 2020.
- [24] S. Zhou, P. Kan, J. Silbernagel, and J. Jin, “Application of image segmentation in surface water extraction of freshwater lakes using radar data,” *ISPRS International Journal of Geo-Information*, vol. 9, no. 7, 2020.
- [25] P. Pandey, P. K. Singh, and S. Chandra, “A review on image segmentation techniques,” *Pattern Recognition and Image Analysis*, vol. 27, no. 2, pp. 259–284, 2017.
- [26] N.-D. Hoang, “Detection of surface crack in building structures using image processing technique with an improved otsu method for image thresholding,” *Advances in Civil Engineering*, vol. 2018, 2018.

- [27] R. S. Ehlers, “Análise de séries temporais,” *Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná*, vol. 1, pp. 1–118, 2007.
- [28] Y.-C. Lin, Y.-T. Yang, and C.-T. Hsu, “Short-term electricity demand forecasting using arima and regression models,” *Energy*, vol. 228, p. 120401, 2021.
- [29] Y.-Y. Lai, C.-C. Huang, Y.-C. Yeh, J.-L. Hsu, W.-C. Chiu, and T. B. Kuo, “Spectral analysis of heart rate variability in alzheimer’s disease,” *Journal of Alzheimer’s Disease*, vol. 76, no. 4, pp. 1561–1571, 2020.
- [30] H. Kaur and S. Singh, “Predicting usd/inr exchange rate using time series models and artificial neural networks,” *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 3, pp. 338–346, 2021.
- [31] X. Guo, Y. Zheng, and W. Wang, “Electricity consumption forecasting using sarima and deep learning models,” *Applied Energy*, vol. 275, p. 115290, 2020.
- [32] Y. Zhao, W. Liu, D. Li, J. Li, W. Li, H. Li, and Y. Li, “Modeling and predicting soil temperature in maize fields using regression analysis,” *Soil and Tillage Research*, vol. 205, p. 104747, 2021.
- [33] P. MORETTIN and W. BUSSAB, *ESTATÍSTICA BÁSICA*. Saraiva Educação S.A., 2017.
- [34] C. M. Morettin, Pedro A e Tolo, *Análise de séries temporais: modelos lineares univariados*. Editora Blucher, 2018.
- [35] N. R. Sharpe, *Estatística aplicada: administração, economia e negócios*. Grupo A-Bookman, 2000.
- [36] J. Rashid, Khandakar M e Louis, “Aprimoramento de dados de séries temporais e aprendizado profundo para reconhecimento de atividades de equipamentos de construção,” *Informática de Engenharia Avançada*, vol. 42, 2019.
- [37] S. Karmy, Juan Pablo e Maldonado, “Previsão de séries temporais hierárquicas por meio de regressão de vetor de suporte no setor de varejo de viagem europeu,” *Sistemas Especialistas com Aplicativos*, vol. 137, 2019.
- [38] G. Box, “Box and jenkins: time series analysis, forecasting and control,” *A Very British Affair, ser. Palgrave Advanced Texts in Econometrics. Palgrave Macmillan UK*, pp. 161–215, 2013.
- [39] J. P. d. Sales, “Modelling and forecasting sustainable development time series via single and combined approaches,” 2019. Dissertação (Mestrado) – Universidade Federal do Cariri, Centro de Ciências Agrárias e Biodiversidade, Mestrado em Desenvolvimento Regional Sustentável, Crato.

- [40] M. R. Abonazel and N. M. Darwish, “Forecasting confirmed and recovered covid-19 cases and deaths in egypt after the genetic mutation of the virus: Arima box-jenkins approach,” *Commun. Math. Biol. Neurosci.*, vol. 2022, pp. Article–ID, 2022.
- [41] M. Farsi, D. Hosahalli, B. R. Manjunatha, I. Gad, E.-S. Atlam, A. Ahmed, *et al.*, “Parallel genetic algorithms for optimizing the sarima model for better forecasting of the ncdc weather data,” *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 1299–1316, 2021.
- [42] C. A. G. da Silva, “Análise da previsão do preço do etanol hidratado no estado de são paulo: uma aplicação do modelo arima,” *Brazilian Journal of Development*, vol. 5, no. 10, pp. 17763–17778, 2019.
- [43] F. G. d. Brito *et al.*, “Aplicação do modelo arima na previsão de atendimentos em pontos de atenção com alta demanda da rede de assistência à saúde do município de monte carmelô, mg,” 2019.
- [44] C. Chatfield, *The analysis of time series: an introduction*. CRC Press, 2019.
- [45] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, “A state space framework for automatic forecasting using exponential smoothing methods,” *International Journal of Forecasting*, vol. 24, no. 3, pp. 439–454, 2008.
- [46] R. J. Hyndman and Y. Khandakar, “Automatic time series forecasting: the forecast package for r,” *Journal of statistical software*, vol. 27, pp. 1–22, 2008.
- [47] R. J. Hyndman and G. Athanasopoulos, “Forecasting: principles and practice, otexts: Melbourne, australia; 2018,” *OTexts. com/fpp2*. Accessed, vol. 19, 2021.
- [48] C. C. Holt and P. R. Winters, “Forecasting seasonal and trended time series: some extensions to the holt-winters approach,” *Journal of the American Statistical Association*, vol. 55, no. 292, pp. 786–793, 1960.
- [49] A. Jain and S. Kumar, “Forecasting with holt-winters exponential smoothing and its variants: A review,” *International Journal of Research in Engineering, Science and Management*, vol. 4, no. 4, pp. 119–124, 2021.
- [50] W. Enders, *Applied econometric time series*. John Wiley & Sons, 2019.
- [51] Y. Bao, G. Zhang, and X. Dang, “A review of the development of the holt–winters forecasting methodology for application in supply chain management,” *International Journal of Production Research*, vol. 56, no. 14, pp. 4659–4678, 2018.
- [52] S. Kundu, S. Saha, and S. Pal, “On selection of model parameters for holt-winters exponential smoothing method,” *Journal of Applied Statistics*, vol. 47, no. 5, pp. 832–853, 2020.

- [53] L. S. Silva, J. O. Cerdeira, and F. H. Quina, “Análise de sobrevivência de pacientes com câncer de mama utilizando modelos de regressão de cox,” *Revista Eletrônica de Biologia*, vol. 17, no. 1, p. e20212837, 2021.
- [54] A. K. Sangaiah, K. Subramanian, and S. Vinodh, “A review of various evaluation metrics and performance comparison of machine learning models,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 12, pp. 5543–5558, 2020.
- [55] S. Shapiro and M. Wilk, “Analysis of variance test for normality (complete samples),” *Biometrika*, vol. 52, no. 3-4, pp. 591–611, 1965.
- [56] J. Rojas-Lima, M. Cruz-Aceves, and O. Garcia-Cueto, “Shapiro-wilk test: interpretation and implementation in statistical software,” *Revista de la Facultad de Ingeniería*, vol. 33, no. 1, pp. 1–10, 2018.
- [57] J. Kim, “Power comparison of shapiro-wilk test and kolmogorov-smirnov test in small samples,” *Journal of Korean Data and Information Science Society*, vol. 29, no. 1, pp. 41–49, 2018.
- [58] U. Bodenhofer, A. Kothmeier, and S. Hochreiter, “Evaluation of visual and statistical goodness-of-fit measures for normal distributions,” *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 68, no. 4, pp. 1009–1027, 2019.
- [59] N. J. Gotelli and A. M. Ellison, *A primer of ecology*. Sinauer associates, 2019.
- [60] A. Kuznetsov, R. Borges, W. C. da Silva, and A. von Wangenheim, “Performance of regression models on simulation data from studies with continuous endpoints: A simulation study,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 9, p. 4728, 2021.
- [61] T. O. Hodson, “Root-mean-square error (rmse) or mean absolute error (mae): when to use them or not,” *Geoscientific Model Development*, vol. 15, no. 14, pp. 5481–5487, 2022.
- [62] A. Kumar, A. Alsadoon, P. Prasad, S. Abdullah, T. Rashid, D. T. H. Pham, and T. Q. V. Nguyen, “Generative adversarial network (gan) and enhanced root mean square error (ermse): deep learning for stock price movement prediction,” *Multimedia Tools and Applications*, pp. 1–19, 2022.
- [63] L. S. Lima, A. R. Santos, and G. B. Cardoso, “Forecasting monthly electricity demand in northeastern brazil using a time series approach,” *Energy*, vol. 195, p. 116957, 2020.
- [64] M. Elsaraiti and A. Merabet, “Solar power forecasting using deep learning techniques,” *IEEE Access*, vol. 10, pp. 31692–31698, 2022.

- [65] M. S. Rahman, M. Saleem, and A. Maqbool, “Forecast accuracy evaluation using mape and wmape metrics: a comparative study,” *Procedia Computer Science*, vol. 148, pp. 622–629, 2019.
- [66] S. C. Koh and C. K. Chan, “Comparison of mape and wmape for demand forecasting with intermittent demand patterns,” *International Journal of Production Research*, vol. 57, no. 23, pp. 7424–7438, 2019.
- [67] F. Petropoulos and N. Kourentzes, “Forecasting with multivariate temporal aggregation: The case of promotional modelling,” *Journal of Business Research*, vol. 115, pp. 361–372, 2020.
- [68] G. Jasso and H. Suh, “Information and entropy measures for categorical data,” *Sociological Methodology*, vol. 45, no. 1, pp. 1–48, 2015.
- [69] C. Oyvat, “Regional inequality in brazil: a multidimensional analysis,” *Review of Development Economics*, vol. 22, no. 1, pp. 145–170, 2018.
- [70] M. López-García, M. Martínez-Damián, and J. Arana-Coronado, “Predictores del precio de maíz blanco en jalisco y michoacán,” *Revista mexicana de ciencias agrícolas*, vol. 13, no. 2, pp. 261–272, 2022.
- [71] A. Fakhri and P. L. Ghazalian, “Income inequality and economic growth: empirical evidence from developing countries,” *Applied Economics Letters*, vol. 26, no. 15, pp. 1279–1284, 2019.
- [72] United States Geological Survey, “Provisional landsat 8 surface reflectance code (lasrc) product. versão 3.3,” Acesso em outubro de 2022.
- [73] D. G. Jorgensen and L. E. Stullken, *Hydrology and Model of North Fork Solomon River Valley, Kirwin Dam to Waconda Lake, North-central Kansas*, vol. 6. University of Kansas Publications, 1981.
- [74] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022.
- [75] K. Lin, D. Wald, C. Kircher, D. Slosky, K. Jaiswal, and N. Luco, “Usgs shakecast system advancements,” in *Proceedings of the 11th national conference on earthquake engineering*, pp. 3458–3468, June 2018.
- [76] S. Atique, S. Noureen, V. Roy, V. Subburaj, S. Bayne, and J. Macfie, “Forecasting of total daily solar energy generation using arima: A case study,” in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 0114–0119, IEEE, 2019.