

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

ANDRÉ LUIS QUIOSI

**CARACTERIZAÇÃO E ANÁLISE DE SÉRIES TEMPORAIS A PARTIR DE
WAVELETS PARA PREDIÇÃO DE PRECIPITAÇÃO**

TOLEDO

2023

ANDRÉ LUIS QUIOSI

**CARACTERIZAÇÃO E ANÁLISE DE SÉRIES TEMPORAIS A PARTIR DE
WAVELETS PARA PREDIÇÃO DE PRECIPITAÇÃO**

**Characterization and analysis of time series using wavelets for precipitation
prediction**

Trabalho de Conclusão de Curso de Graduação
apresentado como requisito para obtenção do
título de Tecnólogo em Tecnologia em Sistemas
para Internet do Curso Superior de Tecnologia
em Sistemas para Internet da Universidade
Tecnológica Federal do Paraná.

Orientador: Prof. Dr. Fabio Alexandre Spanhol

Coorientador: Prof. Dr. Márcio Paulo de Oliveira

TOLEDO

2023



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

ANDRÉ LUIS QUIOSI

**CARACTERIZAÇÃO E ANÁLISE DE SÉRIES TEMPORAIS A PARTIR DE
WAVELETS PARA PREDIÇÃO DE PRECIPITAÇÃO**

Trabalho de Conclusão de Curso de Graduação
apresentado como requisito para obtenção do
título de Tecnólogo em Tecnologia em Sistemas
para Internet do Curso Superior de Tecnologia
em Sistemas para Internet da Universidade
Tecnológica Federal do Paraná.

Data de aprovação: 20/novembro/2023

Fabio Alexandre Spanhol
Doutor
Universidade Tecnológica Federal do Paraná

Márcio Paulo de Oliveira
Doutor
Universidade Tecnológica Federal do Paraná

Ivan Luiz Salvadori
Doutor
Universidade Tecnológica Federal do Paraná

**TOLEDO
2023**

Dedico este trabalho à minha família, em especial mãe e irmã pelos momentos de ausência, apoio e compreensão.

AGRADECIMENTOS

Agradeço ao meu orientador Prof. Dr. Fabio Alexandre Spanhol, pela grande sabedoria e paciência, também o coorientador Prof. Dr. Márcio Paulo de Oliveira, que me auxiliou muito nesta trajetória.

Agradecimento também ao meu antigo orientador, que mudou de instituição, Prof. Dr. Thiago Henrique Pereira Silva que me inspirou muito com seu vasto conhecimento no começo dessa jornada.

Aos meus colegas de sala, pelos vários trabalhos realizados em conjunto e experiência compartilhada.

Também aos meus amigos e colegas de graduação, Eduardo, Diego, Paula, Lucas, Gustavo, Cristian e Vinícius que sempre estiveram presentes e ajudando com suas experiências.

Por último, mas não menos importante, imensa gratidão a todos os professores da graduação, por todo conhecimento passado, experiências de vida compartilhadas e paciência em nossas dificuldades.

RESUMO

Nos últimos anos, temos observado um aumento na intensidade das mudanças climáticas, o que ressalta a importância de prever e compreender os fenômenos naturais. A análise de dados de séries temporais provenientes de modelos de previsão desempenha um papel fundamental em várias áreas científicas, como estatística, meteorologia, agricultura, economia e ciência de dados. A transformada *wavelet* é uma ferramenta matemática amplamente usada na análise desses dados. Ela permite decompor as séries em diferentes escalas e frequências, revelando padrões e comportamentos ocultos, sendo possível a extração de informações relevantes, como tendências, sazonalidades, variações de curto prazo e anomalias. Ao combinar a análise de séries temporais com dados espaciais, podemos obter informações valiosas sobre a distribuição espacial de fenômenos, identificar padrões geográficos e compreender as relações entre variáveis espaciais e temporais. O seguinte trabalho propôs compreender os dados de precipitação pluviométrica, considerando suas características espaciais e temporais, para obter conhecimentos sobre os padrões de precipitação, contribuindo assim para uma melhor compreensão dos fenômenos relacionados e auxiliando na tomada de decisões embasadas em informações confiáveis. Como resultado, as representações visuais detalhadas das séries temporais, utilizando a transformada *wavelet*, facilitaram uma compreensão mais aprofundada de seus padrões e mudanças, permitindo a extração de conclusões a partir das evidências apresentadas.

Palavras-chave: transformada wavelet; séries temporais; previsão de precipitação; periodograma; mudança climática.

ABSTRACT

In recent years, we have observed an increase in the intensity of climate change, underscoring the importance of predicting and understanding natural phenomena. The analysis of time series data from forecasting models plays a crucial role in various scientific fields, such as statistics, meteorology, agriculture, economics, and data science. The wavelet transform is a widely used mathematical tool in the analysis of this data. It allows for the decomposition of time series into different scales and frequencies, revealing hidden patterns and behaviors, enabling the extraction of relevant information such as trends, seasonal variations, short-term fluctuations, and anomalies. By combining the analysis of time series with spatial data, we can obtain valuable insights into the spatial distribution of phenomena, identify geographic patterns, and understand the relationships between spatial and temporal variables. The following work aimed to understand precipitation data, considering its spatial and temporal characteristics. The expectation is to gain knowledge about precipitation patterns, thus contributing to a better understanding of related phenomena and aiding decision-making based on reliable informations. As a result, detailed visual representations of time series, using the wavelet transform, facilitated a deeper understanding of their patterns and changes, allowing for the extraction of conclusions from the presented evidence.

Keywords: wavelet transform; time series analysis; precipitation forecast; periodogram; climate change.

LISTA DE FIGURAS

Figura 1 – Seções GRIB	18
Figura 2 – Localização do pixel na área de estudo	19
Figura 3 – Sequência das etapas do trabalho	22
Figura 4 – Detalhamento das atividades realizadas	22
Figura 5 – Figura de exemplo da localização espacial dos dados de previsão de precipitação	25
Figura 6 – Exemplo dos dados de previsão de precipitação localizados em coordenadas geográficas correspondentes ao estado do Paraná	26
Figura 8 – Serie Temporal de previsão de precipitação decendial para Toledo e Cascavel - PR (<i>pixel</i> 94)	29
Figura 9 – Periodogramas das series temporais de previsão de precipitação decendial para Toledo e Cascavel - PR (<i>pixel</i> 94), obtidos com a <i>wavelet</i> de Morlet	31

LISTA DE TABELAS

Tabela 1 – Tabela demonstrativa dos dados encontrados na Figura 9	34
--	-----------

LISTAGEM DE CÓDIGOS FONTE

Listagem 1 – Exemplo de requisição de dados da API <i>European Centre for Medium-Range Weather Forecasts</i> (ECMWF)	20
Listagem 2 – Cálculo do espectro de potência <i>wavelet</i> de série temporal univariada	23

LISTA DE ABREVIATURAS E SIGLAS

Siglas

API	<i>Application Programming Interface</i>
EAP	Estrutura Analítica de Projeto
ECMWF	<i>European Centre for Medium-Range Weather Forecasts</i>
GRIB	<i>Gridded Binary</i>
LOESS	<i>Locally Estimated Scatterplot Smoothing</i>
NetCDF	<i>Network Common Data Form</i>
ONU	Organização das Nações Unidas
TIGGE	<i>The Interactive Grand Global Ensemble</i>
UTC	Coordenada de Tempo Universal
WMO	<i>World Meteorological Organization</i>
WWRP	<i>World Weather Research Programme</i>

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Considerações iniciais	11
1.2	Objetivos	12
1.2.1	Objetivo geral	12
1.2.2	Objetivos específicos	12
1.3	Justificativa	12
2	REFERENCIAL TEÓRICO	14
2.1	Séries Temporais	14
2.2	Transformada Wavelet	15
2.3	Dados Geoespaciais	17
3	MATERIAIS E MÉTODOS	18
3.1	Materiais	18
3.1.1	Arquivo GRIB	18
3.1.2	<i>Dataset</i>	19
3.1.3	Séries Temporais	20
3.1.4	Extração dos dados utilizando Python	20
3.1.5	Linguagem R	21
3.2	Métodos	21
3.2.1	<i>Scripts</i>	23
4	RESULTADOS	25
5	CONCLUSÃO	35
	REFERÊNCIAS	36

1 INTRODUÇÃO

1.1 Considerações iniciais

Este trabalho aborda o tema da análise de previsões de precipitação, usando séries temporais de dados pluviométricos georreferenciados para uma análise espacial. O conhecimento relacionado às previsões de precipitação pluviométricas influenciam na tomada de decisões em diversas atividades sociais, econômicas, gestão de recursos hídricos, planejamento urbano e agricultura.

Um dos desafios do trabalho com séries temporais é a seleção adequada de técnicas que possam lidar com a sua complexidade e não linearidade, para isso as *wavelets* têm mostrado eficácia em decompor as séries temporais em diferentes escalas e revelar padrões ocultos.

As *wavelet*, que podem ser traduzidas como “pequenas ondas” tiveram sua primeira aparição na literatura de geofísica, no trabalho de (MORLET *et al.*, 1982). Contudo as *wavelets* foram também desenvolvidas em várias outras áreas da ciência, como estatística, processamento de imagens e hidrologia (BOLZAN, 2006; NORDEMANN, 1998; VETTERLI; HERLEY, 1992).

As bases de dados, de onde foram extraídos os objetos da pesquisa, fazem parte do *World Weather Research Programme* (WWRP), sendo uma iniciativa internacional de pesquisa patrocinada pela *World Meteorological Organization* (WMO). Seu objetivo é impulsionar avanços nas previsões meteorológicas de curto prazo, abrangendo um período de um a quinze dias, especialmente em eventos de grande impacto. Um dos principais componentes do programa é o *The Interactive Grand Global Ensemble* (TIGGE), que desempenha um papel fundamental nesse avanço.

O Centro Europeu de Previsões Meteorológicas a Médio Prazo ECMWF disponibiliza o acesso ao modelo TIGGE, que por sua vez disponibiliza os dados de precipitação pluviométrica, sendo um parâmetro temporal climático altamente variável e de características sazonais. Para tal, foi realizado o tratamento, visualização e criação de mapas temáticos. Além disso, caracterizou-se os dados obtidos por meio de gráficos, proporcionando uma visualização clara da interpretação dos resultados. Também realizou-se a incorporação de dados espaciais georreferenciados, permitindo uma análise abrangente e contextualizada no espaço. A expectativa para este trabalho é os resultados contribuírem para um melhor entendimento das séries temporais de previsão de precipitação de alcance decenal do modelo ECMWF com o uso de *wavelets* na localização interna da mesorregião oeste do estado do Paraná.

1.2 Objetivos

1.2.1 Objetivo geral

Avaliar o método de transformada *wavelet* para previsão de precipitação, com vistas a contribuir para um melhor entendimento das informações contidas no método, informações essas que podem ser utilizadas no planejamento de diversas atividades. Com isso, melhorar as tomadas de decisões em relação a eventos climáticos extremos, impactando positivamente na qualidade de vida da população.

1.2.2 Objetivos específicos

O escopo deste estudo abrange uma série de etapas metodológicas, iniciando pela aquisição de dados meteorológicos de fontes oficiais como o ECMWF¹. O processo segue com o tratamento e limpeza dos dados coletados, assegurando que a qualidade dos mesmos seja adequada para análises subsequentes. A representação das séries temporais é efetuada através da aplicação de transformadas de *wavelets*, que permitem a decomposição dos dados em componentes de frequência-tempo para uma análise mais refinada. A etapa final consiste na visualização e interpretação dos dados por meio de tabelas, gráficos e mapas temáticos, facilitando a compreensão e a identificação de padrões ou tendências significativas nos fenômenos meteorológicos investigados.

1.3 Justificativa

De acordo com a Organização das Nações Unidas (ONU)², as mudanças climáticas, sendo elas naturais, tal qual as variações no ciclo solar, ou atividades humanas, como a queima de combustíveis fósseis, refletem transformações ao longo do tempo nos padrões de temperatura e clima. Tais mudanças percebidas e analisadas nas últimas décadas trouxeram uma problemática para diversas atividades humanas, como a agricultura, transporte de mercadorias, transporte humano, viagens, entre outras. Essas transformações tornaram mais desafiadora a tarefa de analisar dados climáticos para prever fenômenos naturais. As ferramentas e métodos que foram empregados por um longo período agora demonstram menor eficácia (MALHI; KAUR; KAUSHIK, 2021).

Tomando como base de estudo a região oeste do estado do Paraná, temos como peça fundamental o agronegócio, que movimenta a economia regional³ e contém grandes cooperativas que exportam suas produções.(ALMEIDA, 2020) Além disso, outro ponto de interesse é

¹ <https://www.ecmwf.int/>, acessado em 25/05/2023.

² <https://shorturl.at/hrERV>, acessado em 02/05/2023.

³ Agência Estadual de Notícias: <https://shorturl.at/abkFM>, acessado em 28/05/2023.

o rio Paraná e seus afluentes, que cobre toda a divisa oeste do estado do Paraná, fornecendo água e a capacidade de produção de energia elétrica na usina hidrelétrica de Itaipu.

Este trabalho contribui para a interdisciplinaridade dos aspectos socioeconômicos, desastres naturais, análises matemáticas resolvidas por métodos avançados de programação dinâmica (transformada de *wavelet*), engenharia hídrica, pesquisa geofísica, pesquisa ambiental, agronomia, entre outros.

2 REFERENCIAL TEÓRICO

2.1 Séries Temporais

Séries temporais referem-se a conjuntos de dados coletados, registrados ou observados em intervalos regulares ao longo do tempo, organizados cronologicamente para analisar padrões, tendências e variações. Amplamente aplicadas em setores como finanças, economia, meteorologia, ciência ambiental e engenharia, essas séries proporcionam percepções cruciais. Conceitos essenciais associados incluem observações temporais, vinculando cada ponto de dados a um momento específico, tendências que representam mudanças de longo prazo (ascendentes, descendentes ou constantes), padrões sazonais refletindo variações periódicas, flutuações aleatórias como variações não sistemáticas e ciclos representando padrões recorrentes de aumento e diminuição sem duração fixa. Essa análise é fundamental para prever comportamentos futuros, sendo aplicada estrategicamente em diversas áreas (MORETTIN; TOLOI, 2018).

A teoria de Fourier, desenvolvida por Joseph Fourier antes de 1930, é fundamental para a compreensão da análise de frequência de funções periódicas. Fourier postulou que qualquer função periódica $f(x)$ pode ser representada como a soma infinita de senos e cossenos, conforme expresso na Equação 1.

$$f(x) = a_0 + \sum_{k=1}^{\infty} (a_k \cos(kx) + b_k \sin(kx)) \quad (1)$$

Essa formulação significa que a função periódica pode ser decomposta em componentes fundamentais senoidais. Os coeficientes a_0 , a_k e b_k são determinados pelas integrais definidas de $f(x)$, como indicado nas Equações 2, 3 e 4.

$$a_0 = \frac{2}{\pi} \int_0^{2\pi} f(x) dx \quad (2)$$

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos(kx) dx \quad (3)$$

$$b_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin(kx) dx \quad (4)$$

Essas fórmulas permitem calcular os coeficientes que descrevem a contribuição de cada termo senoidal na representação da função periódica. Assim, essa abordagem permite decompor funções periódicas complexas em componentes mais simples, facilitando a compreensão e manipulação dessas funções em contextos analíticos e computacionais (BOLZAN, 2004).

No entanto, a abordagem de Fourier tem limitações quando se trata de representar funções não periódicas ou funções com descontinuidades. É nesse contexto que a transformada de

wavelet (ou transformada em ondeletas) se destaca, oferecendo uma abordagem mais flexível e adaptável para analisar sinais em diferentes escalas de frequência e localizações no tempo.

2.2 Transformada Wavelet

Existem diferentes transformadas, entre as mais utilizadas e conhecidas temos o grupo de Wavelet Haar, sendo as *wavelets* mais simples e foram as primeiras a serem introduzidas na teoria das *wavelets*, tendo um tamanho fixo e são usadas para a decomposição de sinais discretos (LEPIK; HEIN, 2014); Wavelet Daubechies, são um conjunto de *wavelets* ortogonais que possuem um número finito de coeficientes, essas *wavelets* são amplamente utilizadas na compressão de imagens e em outras aplicações de processamento de sinais (LINA; MAYRAND, 1995); Wavelet Coiflet, que possuem um bom desempenho em aplicações que envolvem a detecção de bordas em imagens (WEI, 1998) e Wavelet de Morlet, que são moduladas por uma função senoidal. Elas são usadas em aplicações que envolvem a análise de sinais não-estacionários (MORLET *et al.*, 1982).

Dessa forma, neste estudo, pressupomos que a análise dos dados de previsão de séries temporais não estacionárias seria realizada por meio da aplicação da transformada *wavelet* de Morlet. De forma aprofundada, essas transformadas são funções matemáticas importantes na análise de sinais e processamento de dados, pois permitem decompor um sinal em diferentes frequências ou resoluções, revelando informações sobre sua estrutura e comportamento em diferentes escalas. À medida que a resolução aumenta, detalhes mais finos tornam-se visíveis (BOLZAN, 2004). Este processo de decomposição ocorre da seguinte maneira.

Iniciamos considerando uma função f pertencente ao espaço de funções quadraticamente integráveis $L_2(\mathbb{R})$. Portanto, a energia da função f é limitada ao longo do eixo real, conforme a equação:

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt < \infty \quad (5)$$

O produto escalar, representado por (\bullet) , e a norma para o espaço $L_2(\mathbb{R})$ são definidos como:

$$f \bullet g = \int_{-\infty}^{+\infty} f(t) \cdot g(t) dt \quad (6)$$

$$|f| = \sqrt{f \bullet f} \quad (7)$$

Se a função g tiver uma norma unitária, então o produto escalar entre f e g corresponderá à operação de projeção de f sobre g .

Agora, considere um conjunto de funções ortonormais $\Psi_k(t)$, $k \in \mathbb{N}$, ou seja, ortogonais e de módulo 1, formando uma base para o espaço $L_2(\mathbb{R})$. Pela condição de ortogonalidade:

$$\Psi_m \bullet \Psi_n = 0 \quad (8)$$

para todo m diferente de n . A expansão da função f em uma série ortonormal é então definida como a combinação linear das funções base ponderadas pelas projeções de f sobre cada uma delas:

$$f(t) = \sum_{k=-\infty}^{\infty} \langle f, \Psi_k \rangle \Psi_k(t) \quad (9)$$

Isso pode ser expresso como:

$$f(t) = \sum_{k=-\infty}^{\infty} F_k \Psi_k(t) \quad (10)$$

onde F_k representa a transformada da função $f(t)$. Imbuído nessa abordagem está o fato de a base $\Psi_k(t)$ ter suporte em toda a extensão do eixo real, ou seja, cada função base $\Psi_k(t)$ decai para zero apenas nos limites de $\pm\infty$. No entanto, se a base tiver suporte limitado, decaindo para zero muito rapidamente, a melhor maneira de cobrir todo o eixo real com essa base seria por meio de translações do tipo:

$$\Psi_{k,l}(t) = \Psi_k(t - l), l \in \mathbb{N} \quad (11)$$

Assim, o termo “ondeleta” refere-se a um conjunto de funções com a forma de pequenas ondas geradas por dilatações, onde $\Psi(t) \rightarrow \Psi(2t)$, e translações, onde $\Psi(t) \rightarrow \Psi(t + 1)$, de uma função base geradora simples $\Psi(t)$, chamada de ondeleta-mãe. Esta ondeleta-mãe deve ser quadraticamente integrável dentro de um intervalo de tempo real ou espaço $[L_2(\mathbb{R})]$, ou seja, deve apresentar energia finita. A condição de admissibilidade da função é imposta pela exigência de que sua energia média seja zero. Isso é expresso pela seguinte equação:

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{2}} \Psi\left(\frac{t-b}{a}\right) \quad (12)$$

Assim, a expansão em série de ondeletas e a transformada são definidas como:

$$\langle f, \Psi_{a,b} \rangle = \int_{-\infty}^{+\infty} \Psi\left(\frac{t-b}{a}\right) dt. \quad (13)$$

2.3 Dados Geoespaciais

Dados geoespaciais, também conhecidos como "geodados", representam informações que estão diretamente relacionadas a pontos ou regiões na superfície terrestre, permitindo a representação cartográfica de objetos e eventos em áreas geográficas específicas, identificadas por meio de coordenadas de latitude e longitude. A visualização geoespacial frequentemente se vale de tecnologia geoespacial para criar mapas interativos e representações gráficas tridimensionais. Essa prática desempenha um papel fundamental na obtenção de uma compreensão mais profunda dos padrões e tendências que ocorrem na superfície terrestre, contribuindo significativamente para o acúmulo de conhecimento sobre o impacto ambiental nas vidas das pessoas e na natureza (LOPES; DELBEM; SOUSA, 2021).

A utilização de geodados e ferramentas de geolocalização tem se tornado cada vez mais crucial nas áreas de planejamento urbano, gestão de recursos naturais, prevenção de desastres, navegação, monitoramento ambiental, e muitas outras. Essas informações espaciais permitem não apenas a representação de dados geográficos de forma eficaz, mas também auxiliam na tomada de decisões informadas e na criação de políticas públicas mais eficazes.

Graças aos avanços tecnológicos na aquisição e análise de geodados, é possível realizar análises mais detalhadas e precisas, permitindo o acompanhamento de mudanças ao longo do tempo e a identificação de áreas com necessidades especiais de intervenção. Portanto, os geodados desempenham um papel crucial na compreensão do nosso planeta, bem como na preservação e na gestão sustentável dos recursos naturais.

3 MATERIAIS E MÉTODOS

3.1 Materiais

3.1.1 Arquivo GRIB

O ECMWF fornece um grande conjunto de dados e uma *Application Programming Interface* (API) que permite desenvolvedores e pesquisadores acessarem e recuperarem de maneira padronizada dados meteorológicos e climáticos dessa instituição. Na meteorologia e ciências climáticas comumente são usados os formatos *Gridded Binary* (GRIB) ou *Network Common Data Form* (NetCDF). Neste trabalho utilizamos dados do modelo TIGGE, disponibilizados pela API no formato GRIB. Esses arquivos contêm informações em forma de grades regulares, onde os dados são organizados em uma estrutura hierárquica.

Figura 1 – Seções GRIB

SECTION 0 - Indicator Section
SECTION 1 - Identification Section
SECTION 2 - [Local Use Section]
SECTION 3 - Grid Definition Section
SECTION 4 - Product Definition Section
SECTION 5 - Data Representation Section
SECTION 6 - Bitmap Section
SECTION 7 - Binary Data Section
SECTION 8 - End Section

Fonte: GRIB decoding Computer User Training Course 2017.

A estrutura básica de um arquivo GRIB consiste em seções lógicas, cada uma com uma finalidade específica. As principais seções demonstradas na Figura 1 incluem:

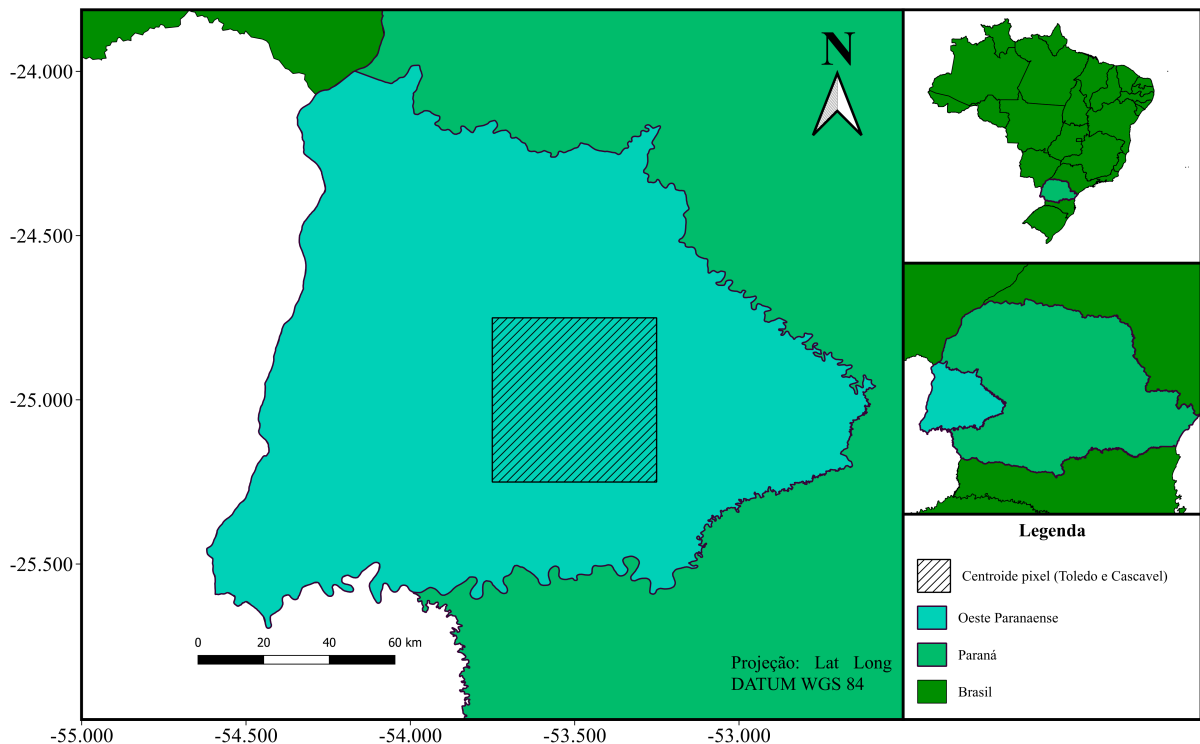
1. **Seção de Cabeçalho** (*SECTION 0*): Esta seção contém informações gerais sobre o arquivo GRIB, como a versão do formato, tamanho do arquivo, data de criação e informações adicionais.
2. **Seção de Identificação** (*SECTION 1*): Essa seção contém informações sobre o produto ou mensagem contida no arquivo GRIB, como o tipo de dados meteorológicos (por exemplo, temperatura, umidade, vento) e a grade espacial utilizada.
3. **Seção de Grade** (*SECTION 3*): Nesta seção, são especificados os detalhes da grade espacial utilizada para representar os dados. Isso inclui informações sobre a latitude e longitude dos pontos da grade, bem como os intervalos de grade em termos de latitude e longitude.

4. **Seção de Registros de Dados** (*SECTION 7*): Esta é a seção que contém os dados meteorológicos ou oceânicos propriamente ditos. Os valores são armazenados em formato binário e organizados de acordo com a grade espacial especificada na seção anterior.
5. **Seção de Fim de Arquivo** (*SECTION 8*): Esta seção marca o final do arquivo GRIB e contém informações adicionais, como um código de verificação para verificar a integridade dos dados.

3.1.2 Dataset

Neste estudo, foi conduzida uma análise de série temporal para previsão de precipitação pluviométrica usando o modelo ECMWF, com um horizonte de previsão de 240 horas, equivalente a um período de dez dias (decêndio). Na fase inicial, será realizada a estratificação espacial dos dados, selecionando as coordenadas geográficas correspondentes a um retângulo que abrange o estado do Paraná (55°W, 48°W, 27°S, 22°S), conforme mostrado na Figura 2. Os *pixels* têm uma dimensão de $0,5^\circ \times 0,5^\circ$, totalizando 165 *pixels*, onde os valores em cada um deles são obtidos por interpolação realizada com base nos próprios dados disponíveis. A região de estudo selecionada nesse trabalho corresponde ao *pixel* 94 dentro da mesorregião Oeste do Paraná, que abrange as cidades de Toledo e Cascavel.

Figura 2 – Localização do pixel na área de estudo



Fonte: Autoria própria (2023).

3.1.3 Séries Temporais

O próximo passo envolve a estratificação temporal, que requer uma análise detalhada dos dados divididos em decêndios referentes aos anos de 2018, 2019, 2020, 2021 e 2022. Além disso, é crucial levar em consideração a Coordenada de Tempo Universal (UTC) para garantir uma interpretação precisa dos eventos no contexto temporal. Destaca-se que 00:00 UTC corresponde a 03:00 no horário de Brasília (BRT), o que implica uma diferença horária importante a ser ajustada ao analisar e correlacionar os dados ao longo desse período.

3.1.4 Extração dos dados utilizando Python

Para realizar a extração de dados meteorológicos do ECMWF, utilizou-se os parâmetros definidos e armazenados em um arquivo no formato GRIB. A Listagem 1 mostra um trecho de código Python especificando os parâmetros da requisição de dados através da API ECMWF, correspondendo à descrição da Seção 3.1.2.

Listagem 1 – Exemplo de requisição de dados da API ECMWF

```

1 from ecmwfapi import ECMWFDataServer
2 server = ECMWFDataServer()
3 server.retrieve({
4     "class": "ti",
5     "dataset": "tigge",
6     "date": "2018-01-01/to/2018-01-31",
7     "stream": "ENFO",
8     "grid": "0.5/0.5",
9     "levtype": "sfc",
10    "number": "0",
11    "origin": "ecmf",
12    "param": "228",
13    "step": "240",
14    "time": "00:00:00",
15    "type": "cf",
16    "area": "-22/305/-27/312",
17    "target": "ECMWF_Step_240_201801.grib",
18 })

```

Fonte: Autoria própria (2023).

Aqui temos uma breve descrição de cada parâmetro.

- `class`: Define a classe de dados como “ti” (*Temporal Interpolation*).
- `dataset`: Especifica origem do conjunto de dados como “tigge”.
- `date`: Define o intervalo de datas para recuperar os dados, de 1 de janeiro de 2018 a 31 de janeiro de 2018.
- `stream`: Indica o fluxo de dados como “ENFO” (*Ensemble Forecast*).
- `grid`: Especifica a grade de dados como Lat/Lon “0.5/0.5”.

- `levtype`: Tipo de nível, neste caso, “sfc” (*surface*).
- `origin`: Origem dos dados, aqui é “ecmf”.
- `param`: Parâmetro a ser recuperado, neste caso, “228” (*Total Precipitation*).
- `step`: Passo de tempo, aqui é “240” horas.
- `time`: Tempo da previsão, definido como “00:00:00”.
- `type`: Tipo de dados como “cf” (*Forecast*).
- `area`: Área geográfica para a qual os dados são solicitados.

Esses e outros parâmetros estão descritos na documentação do usuário¹ no ECMWF.

3.1.5 Linguagem R

A linguagem R² foi desenvolvida especificamente para análise estatística e visualização de dados. Essa linguagem possui uma ampla variedade de pacotes estatísticos e gráficos, tornando-se uma escolha popular entre estatísticos e pesquisadores, também oferece uma sintaxe especializada para análise estatística. As principais bibliotecas que serão usadas são `rgdal`³, `raster`⁴, `sp`⁵, `maptools`⁶, `rworldmap`⁷, `rgeos`⁸ e `WaveletComp`⁹.

Ambas as linguagens têm suas vantagens e são amplamente adotadas pela comunidade de análise de dados. Python é conhecido por sua versatilidade e popularidade geral, sendo usado tanto em empresas como no meio acadêmico. R, por sua vez, é altamente especializado em estatística e análise de dados, sendo frequentemente utilizado em pesquisas científicas e projetos que requerem análise estatística avançada e gráficos complexos.

3.2 Métodos

O trabalho foi desenvolvido conforme o planejamento de atividades demonstrado na Figura 3. Inicialmente, definiu-se o escopo e objetivos do trabalho. Os objetivos foram tratados em ordem, desde a coleta de dados, seguindo o desenvolvimento em relação ao tratamento de

¹ User Documentation: <https://confluence.ecmwf.int/display/UDOC/Keywords+in+MARS+and+Dissemination+requests>

² The R Project for Statistical Computing: <https://www.r-project.org/>, acessado em 20/05/2023.

³ `rgdal`: <https://cran.r-project.org/package=rgdal>, acessado em 21/05/2023.

⁴ `raster`: <https://cran.r-project.org/package=raster>, acessado em 21/05/2023.

⁵ `sp`: <https://cran.r-project.org/package=sp>, acessado em 21/05/2023.

⁶ `maptools`: <https://cran.r-project.org/package=maptools>, acessado em 21/05/2023.

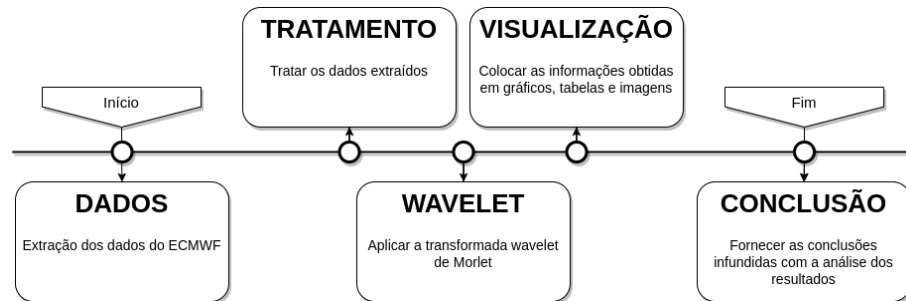
⁷ `rworldmap`: <https://cran.r-project.org/package=rworldmap>, acessado em 21/05/2023.

⁸ `rgeos`: <https://cran.r-project.org/package=rgeos>, acessado em 21/05/2023.

⁹ `WaveletComp`: <https://cran.r-project.org/package=WaveletComp>, acessado em 01/10/2023.

dados e a representação de séries temporais, e, por fim, a visualização de dados. Em paralelo, uma revisão sistemática da literatura ao longo do período letivo, bem como a escrita da redação final do TCC nos últimos quatro meses.

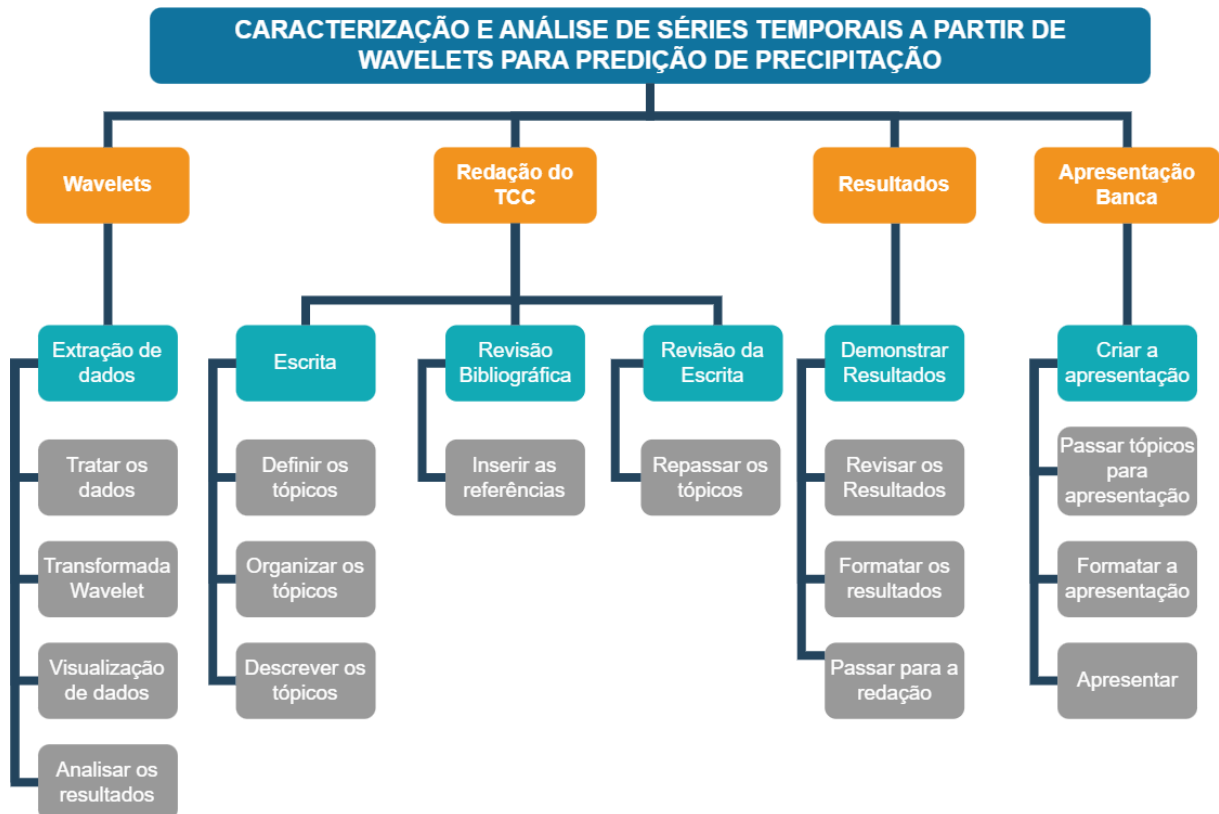
Figura 3 – Sequência das etapas do trabalho



Fonte: Autoria própria (2023).

Para demonstrar de forma mais detalhada todo o trabalho realizado durante a pesquisa e desenvolvimento, foi criado uma Estrutura Analítica de Projeto (EAP), Figura 4, que proporciona uma organização visual dos entregáveis do trabalho em diversos níveis, fundamentando-se nas dependências entre eles. Essencialmente, a EAP representa o plano do projeto em um formato visual, com o objetivo principal no topo e as dependências e sub dependências dispostas hierarquicamente abaixo.

Figura 4 – Detalhamento das atividades realizadas



Fonte: Autoria própria (2023).

3.2.1 Scripts

Os *scripts* desenvolvidos com a linguagem R para realizar as transformadas em ondaleta e gerar seus gráficos, foram baseados no documento desenvolvido por (ROESCH; SCHMIDBAUER, 2018), disponível na documentação da biblioteca WaveletComp. A seguir estão demonstrados os *scripts* juntamente com a descrição de seus parâmetros.

Listagem 2 – Cálculo do espectro de potência *wavelet* de série temporal univariada

```

1  wavelet <- analyze.wavelet (
2    dados,
3    "Precipitacao",
4    loess.span = 0,
5    dt = 1 / 12,
6    dj = 1 / 250,
7    lowerPeriod = 0.150,
8    upperPeriod = 8,
9    make.pval = TRUE,
10   n.sim = 10
11 )

```

Fonte: Autoria própria (2023).

Na Listagem 2, a função `analyze.wavelet()`, pertencente à biblioteca WaveletComp, é empregada para a realização de uma análise *wavelet* de uma série temporal. Os argumentos da função são configurados da seguinte maneira:

- `dados`: Representa a série temporal que será submetida à análise, devendo ser univariada.
- `"Precipitacao"`: Nome da variável dentro do conjunto de dados a ser analisada pela função.
- `loess.span = 0`: Define o grau de suavização *Locally Estimated Scatterplot Smoothing* (LOESS) para atenuar tendências na série temporal, onde o valor 0 indica a ausência de suavização.
- `dt = 1 / 12`: Especifica o intervalo entre as observações na série temporal, sendo 1/12 representativo de observações mensais.
- `dj = 1 / 250`: Determina o passo de frequência na transformada de *wavelet*, onde um valor menor proporciona maior resolução de frequência.
- `lowerPeriod = 0.150`: Menor período a ser considerado na análise, permitindo a detecção de ciclos curtos dentro da série.
- `upperPeriod = 8`: Maior período a ser analisado, possibilitando a identificação de ciclos mais longos.

- `make.pval = TRUE`: Indica se os valores (p-valores¹⁰), para testar a significância estatística, devem ser calculados.
- `n.sim = 10`: Quantidade de simulações que serão realizadas para a estimativa dos p-valores.

O resultado da função `analyze.wavelet()` permitem a análise detalhada da variável "Precipitacao" em diferentes escalas temporais e frequências, bem como a avaliação da significância estatística dos padrões encontrados.

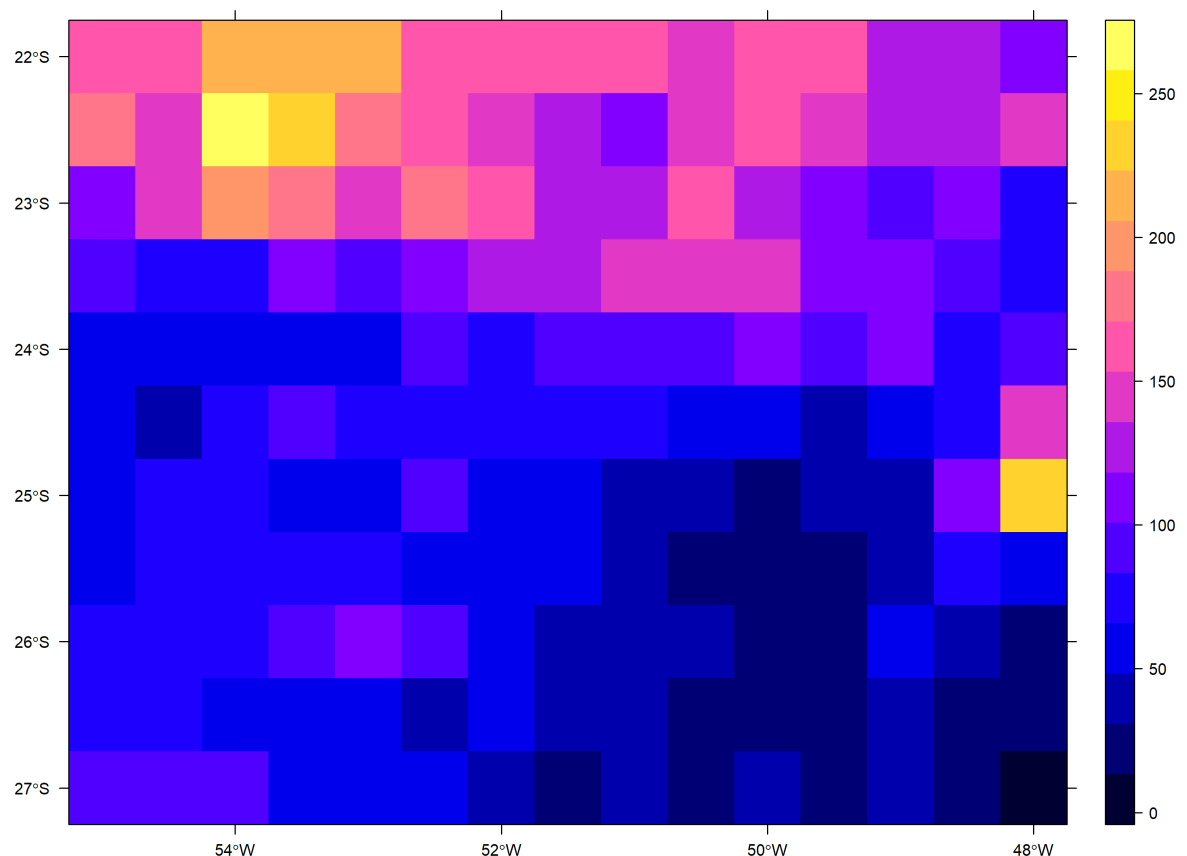
¹⁰ p-valor: No contexto da análise de wavelet, os p-valores são usados para determinar se os coeficientes de wavelet são significativamente diferentes de zero. Se o p-valor for pequeno, isso indica que o coeficiente de wavelet é significativamente diferente de zero, o que sugere que há uma característica significativa na série temporal nesse ponto de tempo e escala.

4 RESULTADOS

Como entrada, o problema recebeu arquivos no formato GRIB, que contêm dados meteorológicos. Em seguida, foram aplicadas as transformadas *wavelets* para extrair características relevantes desses dados, conforme detalhado na Seção 3.2.

A partir dos dados extraídos do arquivo GRIB, pôde-se demonstrar através da Figura 5 um gráfico com coordenadas geoespaciais (55°W, 48°W, 27°S, 22°S) e escala de cores representando a previsão em milímetros, os quadrados com escala de cores e resolução de $0,5^\circ \times 0,5^\circ$, são os *pixels* de cada superfície de previsão de precipitação com alcance de 240h.

Figura 5 – Figura de exemplo da localização espacial dos dados de previsão de precipitação

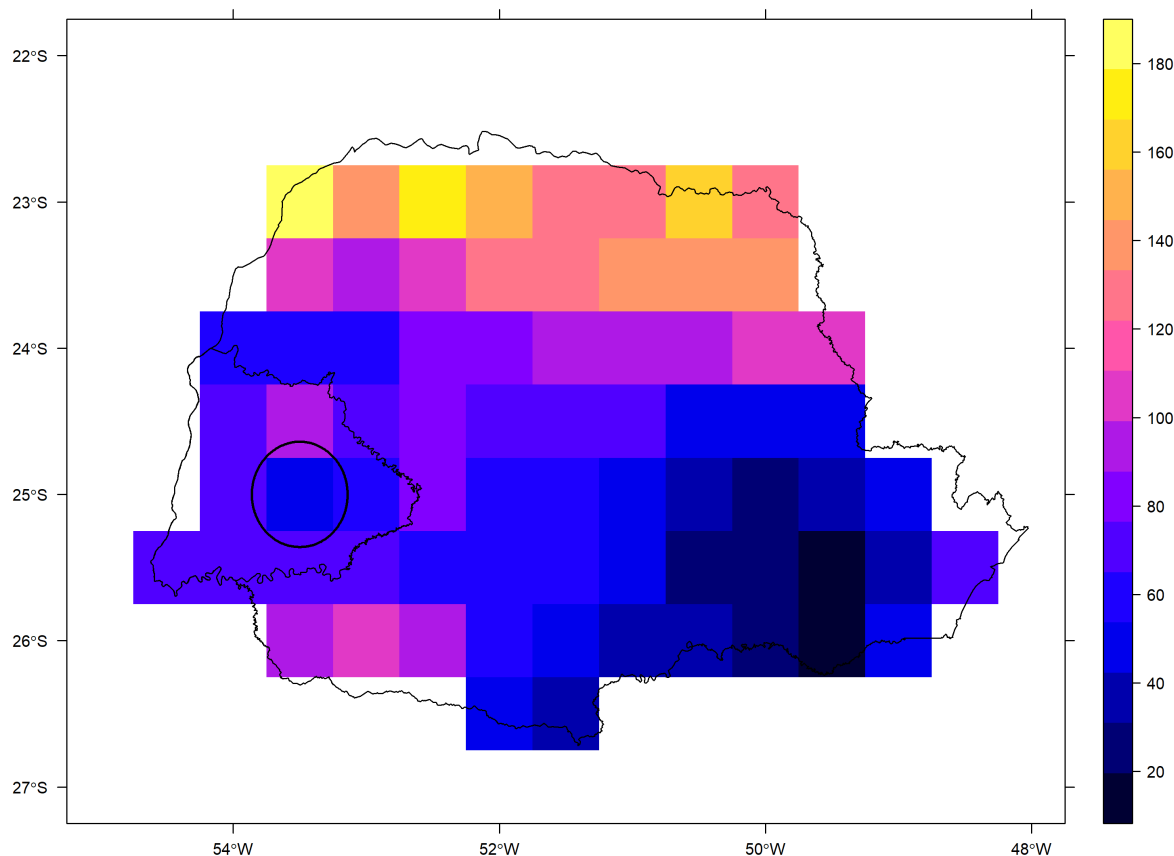


Fonte: Autoria própria (2023).

Como sequência da demonstração dos dados, temos a Figura 6, sendo uma melhoria na apresentação dos dados mostrados anteriormente na Figura 5. Aqui temos os dados para o primeiro decêndio de Janeiro de 2018, com a aplicação do recorte para o estado do Paraná, juntamente com a região Oeste destacada, onde se encontram diversas cidades banhadas pelos afluentes do rio Paraná e com grande produtividade agrícola. Por fim, temos um círculo onde se encontra o *pixel* 94 de uma sequência de 165, que se encontram duas das principais cidades

do território, sendo elas Toledo e Cascavel, sendo esse *pixel* nossa área de estudo. Pode-se visualizar melhor essas localizações na Figura 2.

Figura 6 – Exemplo dos dados de previsão de precipitação localizados em coordenadas geográficas correspondentes ao estado do Paraná



Fonte: Autoria própria (2023).

Para os dados da Figura 7 temos os anos de 2019 a 2022, divididos em cinco gráficos dentro da figura. Cada mês está dividido em decêndios, demonstrados pelos vários mapas com o mesmo recorte exemplificado na Figura 6.

As previsões de precipitação pluviométrica para cada uma das datas foram obtidas por meio do modelo do ECMWF, considerando um intervalo de 240 horas, correspondente aos decêndios (dez dias) e ao horário 00:00 UTC. Os valores dos *pixels* exibem variabilidade tanto temporal quanto espacial, destacando-se especialmente aqueles circundados, que compõem a série temporal em análise.

Na área de estudo, é notável uma ampla flutuação na variável de previsão de precipitação pluviométrica, variando de zero durante períodos de estiagem a valores próximos de 250 mm. Essa significativa oscilação denota um padrão não estacionário na série de dados. Os dois extremos nessa série apontam potenciais problemas, pois períodos prolongados de estiagem

podem impactar a disponibilidade de água nos municípios, enquanto episódios de precipitações intensas podem ocasionar danos tanto em áreas urbanas quanto rurais.

Nas zonas urbanas, chuvas intensas frequentemente resultam em inundações, afetando o tráfego e causando danos às residências, acarretando prejuízos para os moradores. Em regiões rurais, os principais danos estão relacionados às atividades agrícolas, com possíveis impactos negativos nas colheitas, diminuindo a produtividade e, consequentemente, afetando a renda das famílias dependentes dessas atividades.

Figura 7 – Localização espacial dos dados da série temporal. Os círculos indicam os *pixels* em cada superfície de previsão que contém os dados da série temporal

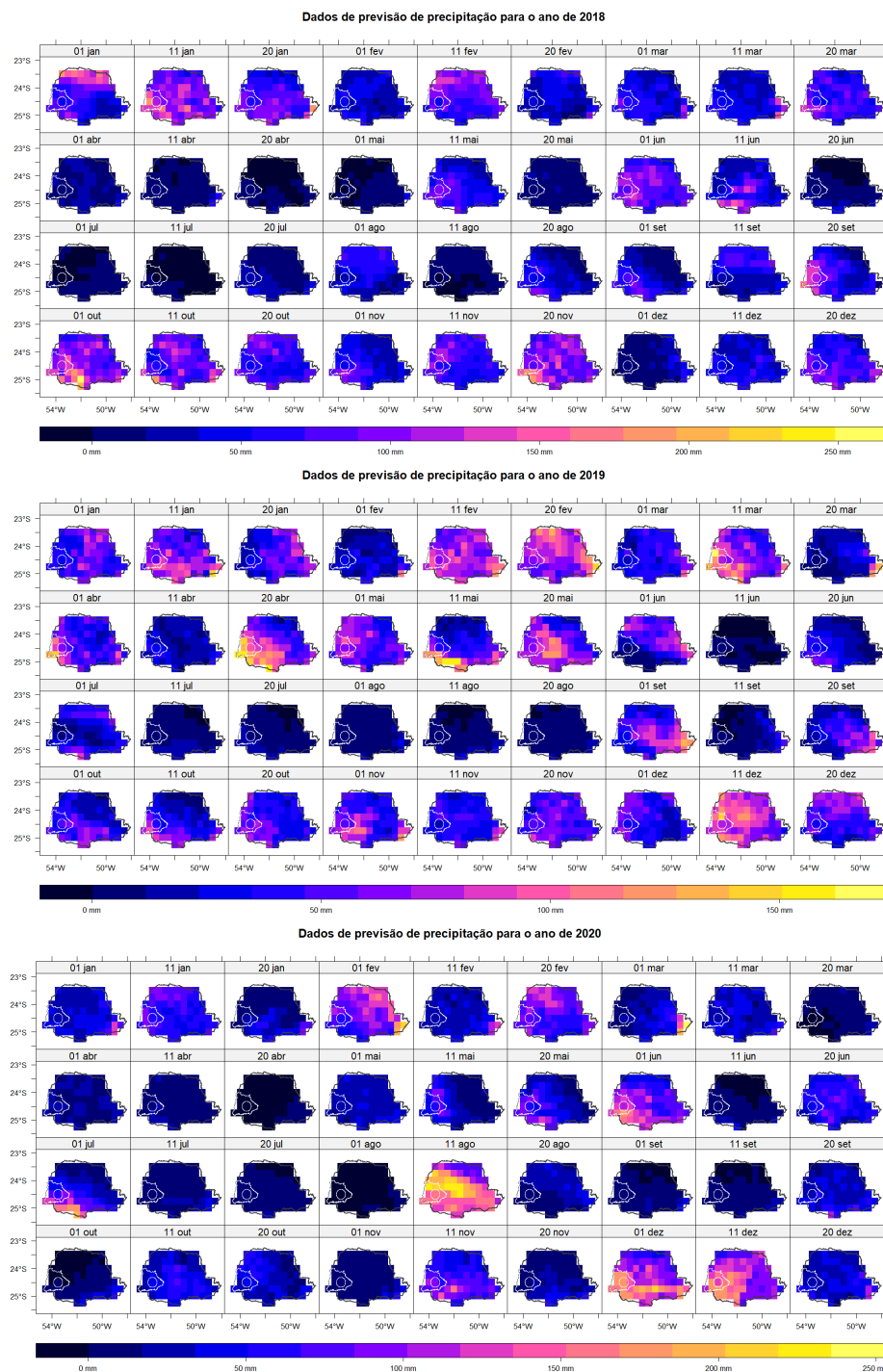
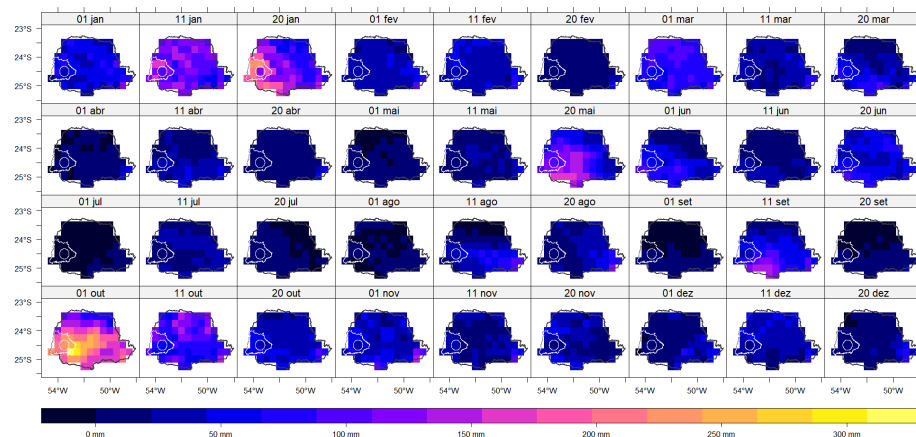
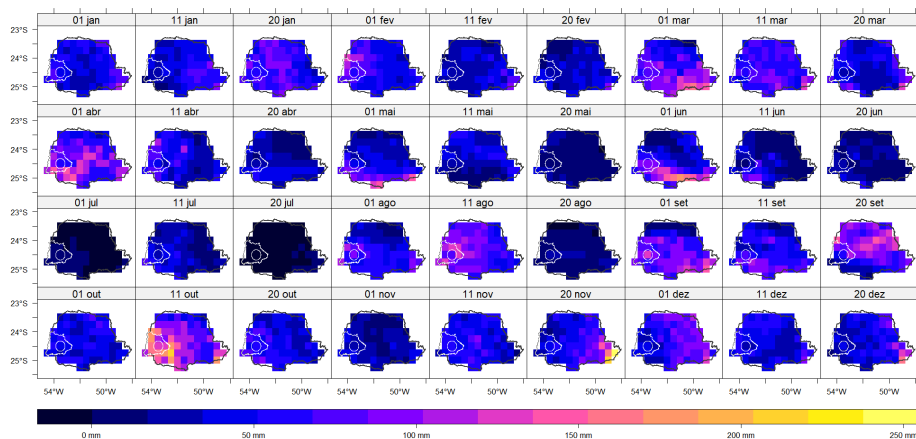


Figura 7 – Continuação

Dados de previsão de precipitação para o ano de 2021



Dados de previsão de precipitação para o ano de 2022

**Fonte: Autoria própria (2023).**

Na Figura 8 é apresentada a série temporal proveniente dos dados dos *pixels* indicados na Figura 7. Essas séries consistem de 36 valores, representando três previsões decendiais para cada mês do ano. Observa-se um padrão não estacionário nessas séries de dados de precipitação pluviométrica, um fenômeno comum em séries temporais provenientes de sistemas físicos, caracterizado pela não constância nos momentos estatísticos, como média e variância, ao longo dos segmentos da série (BOLZAN, 2004). Podemos assim analisar as séries de cada ano, descrevendo cada uma da seguinte forma:

- 2018 teve cinco picos de precipitação acumulada, sendo eles, 11 de Janeiro, 11 de Maio, 01 de Junho, 20 de Setembro e 01 de Outubro. Por outro lado, tivemos períodos de baixo acúmulo de precipitação em 11 de Abril, 20 de Abril, 01 de Maio, 01 de Julho, 11 de Julho, 11 de Agosto e 01 de Dezembro.
- 2019 tem uma boa precipitação acumulada em 11 de Março, 20 de Abril e 11 de Dezembro e períodos de baixa precipitação em 11 de Junho, 01 de Agosto e 11 de Setembro.

- 2020 temos picos de precipitação acumulada em 01 Junho, 11 de Agosto e 01 de Dezembro. Períodos de seca em 20 de Março, 20 de Abril, 11 de Junho, 01 de Agosto, 01 de Outubro e 01 de Novembro.
- 2021 tendo seis períodos de estiagem, sendo eles 01 de Abril, 01 de Maio, 01 de Julho, 01 de Agosto, 01 de Setembro e 20 de Setembro. Porém também temos o período com o decêndio de maior acúmulo de precipitações, 01 de Outubro.
- 2022 pode ser observado um pico no decêndio de 01 Junho, e nos decêndios 01 de Setembro e 11 de Outubro, que representam altas precipitações acumuladas. Também pode ser observado em 01 de Julho e 20 de Julho como sendo os períodos de estiagem.

Figura 8 – Serie Temporal de previsão de precipitação decendial para Toledo e Cascavel - PR (pixel 94)

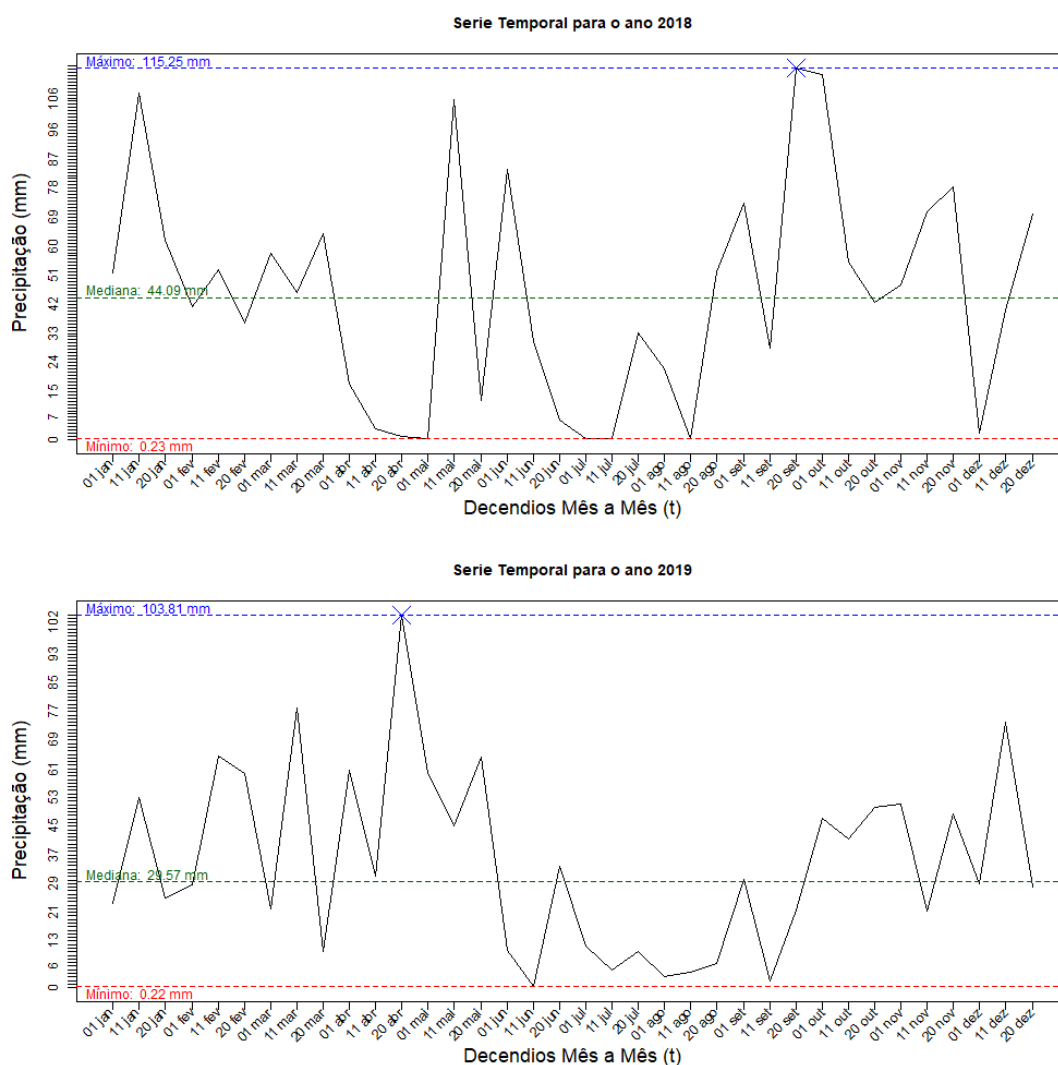
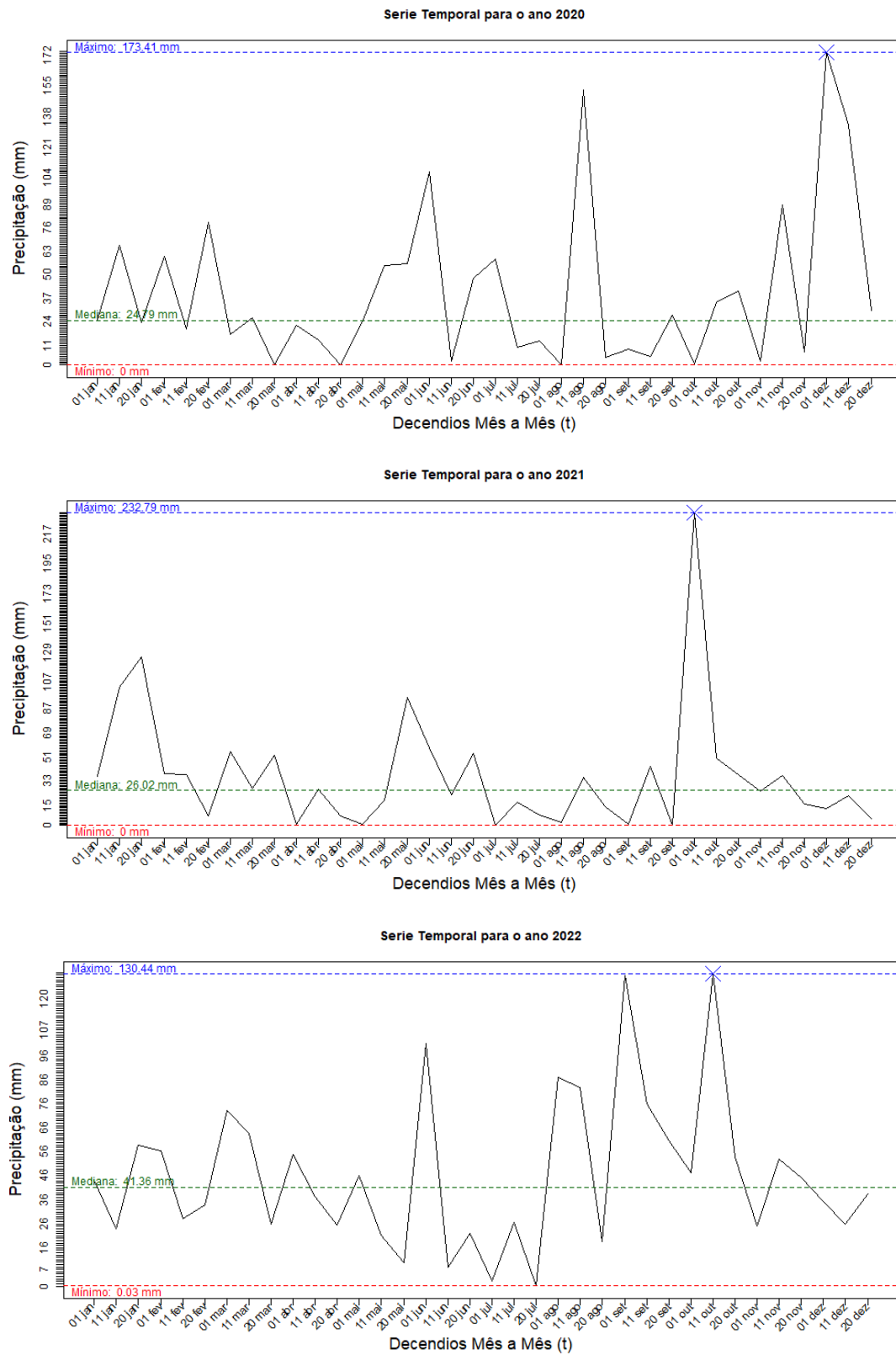


Figura 8 – Continuação



Fonte: Autoria própria (2023).

Os gráficos do espectro de potência de *wavelet*, ilustrados na Figura 9, são uma representação visual que resulta da aplicação da análise de *wavelet* a uma série temporal. Eles fornecem informações sobre a distribuição de energia em diferentes escalas de tempo, ajudando a identificar padrões e variações temporais em dados complexos.

A análise de *wavelet* divide a série temporal em diferentes componentes de frequência e permite analisar como a amplitude dessas frequências varia ao longo do tempo. O gráfico do espectro de potência de *wavelet* representa essa variação de amplitude em um plano bidimensional, onde o eixo horizontal representa o tempo e o eixo vertical representa as diferentes escalas de frequência (ou períodos). A intensidade das cores ou a altura das curvas no gráfico indicam a intensidade da variabilidade em uma determinada escala de tempo. Áreas mais intensamente coloridas ou picos mais altos indicam períodos em que a série temporal exibe maior energia ou variabilidade naquela escala específica (BOLZAN, 2006).

A interpretação destes gráficos (Figura 9) envolve a identificação de padrões temporais em diferentes escalas. Por exemplo, picos em escalas mais longas podem representar padrões sazonais, enquanto picos em escalas mais curtas podem indicar variações de curto prazo. A análise desses padrões pode ajudar a entender a dinâmica temporal dos dados e identificar ciclos ou eventos significativos.

Figura 9 – Periodogramas das series temporais de previsão de precipitação decendial para Toledo e Cascavel - PR (*pixel* 94), obtidos com a *wavelet* de Morlet

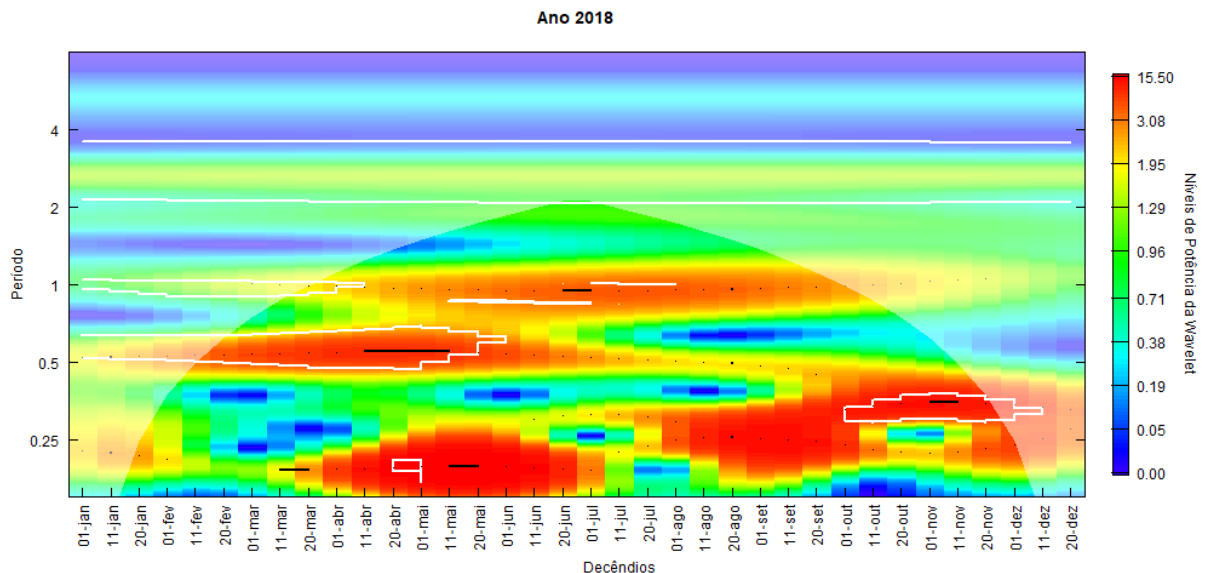


Figura 9 – Continuação

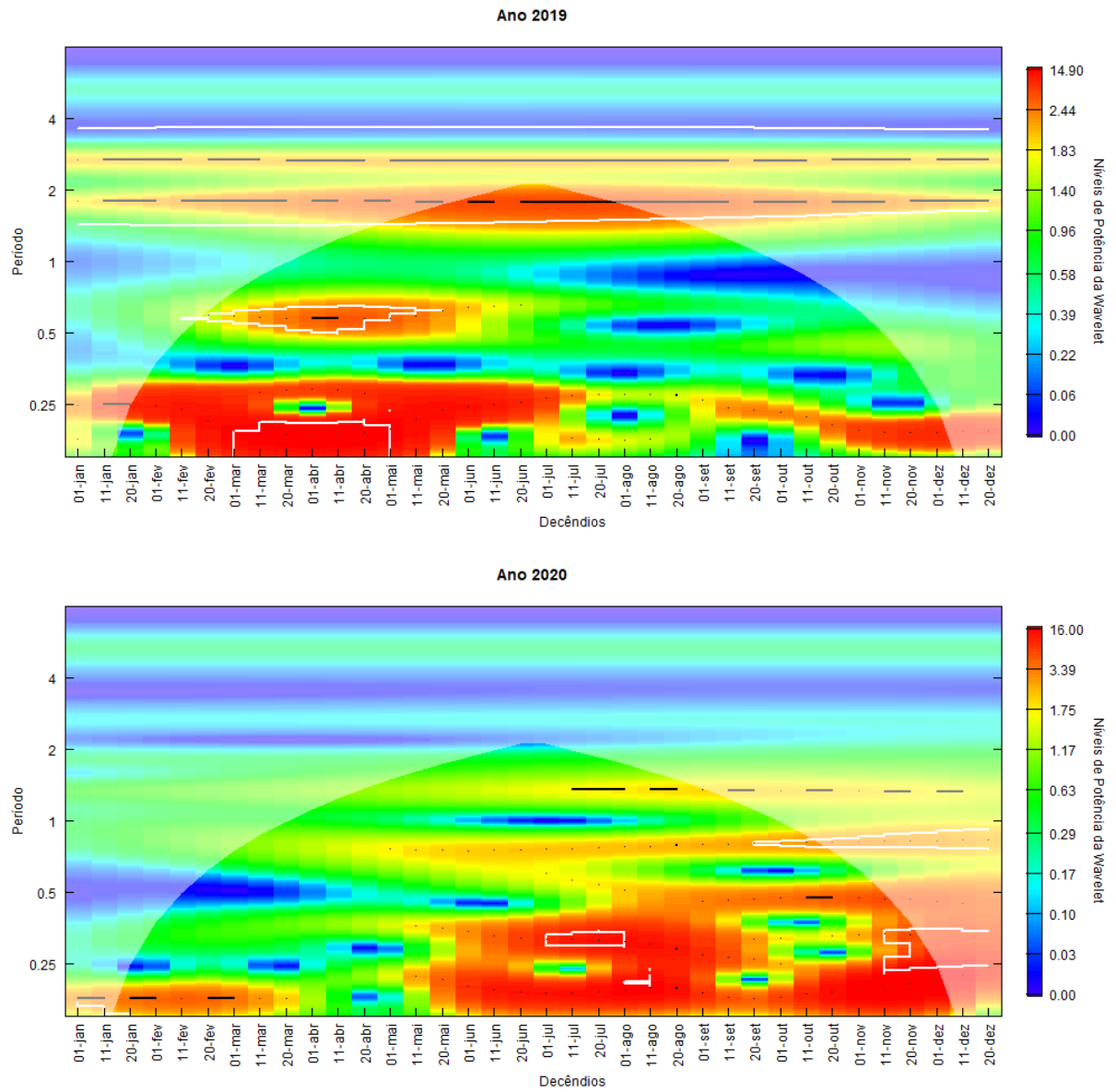
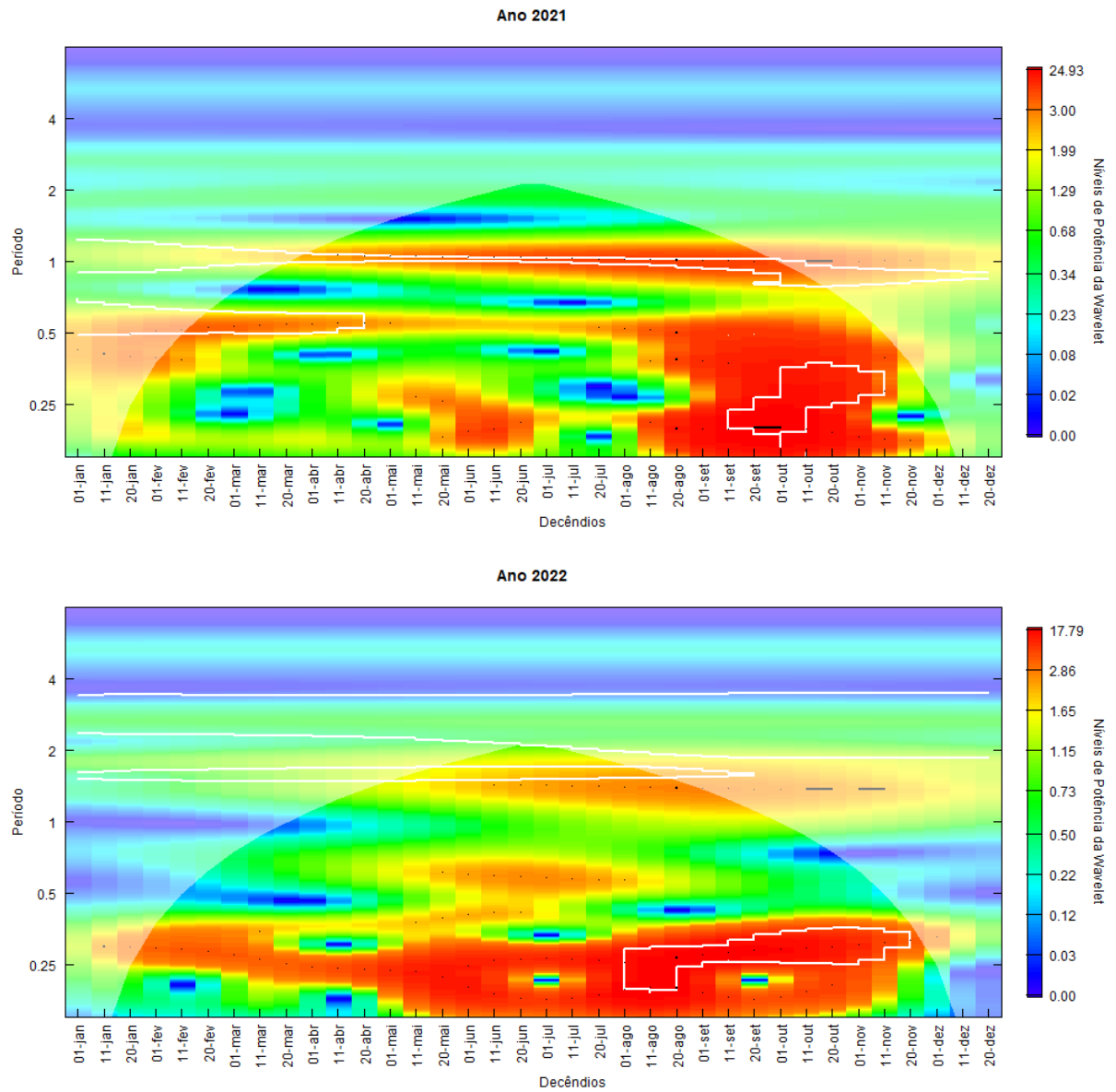


Figura 9 – Continuação



Fonte: Autoria própria (2023).

Ao avaliar os gráficos na Figura 9, identifica-se o cone de influência, que sinaliza a área de confiabilidade dos dados — elementos fora dessa zona, representados pela parte clareada, são omitidos na interpretação. Observa-se ainda que os períodos de maior intensidade, destacados em vermelho, mudaram de posição durante os cinco anos analisados, migrando do início para o meio e, em seguida, do meio para o final dos anos, indicando as épocas de previsões de chuvas mais intensas na área estudada. Além disso, contornos definidos por linhas brancas demarcam as regiões em que os valores de potência da transformada de wavelet são estatisticamente significativos, como descrito na documentação da biblioteca *WaveletComp* por Roesch e Schmidbauer (2018), usada para gerar os periodogramas.

A Tabela 1 fornece uma visão organizada dos resultados da análise do espectro de potência de wavelet ao longo de vários anos, destacando os períodos de significância estatística associadas a um determinado *pixel*. As células da tabela contêm informações detalhadas sobre os períodos de interesse. Por exemplo, na célula correspondente a 2018 e “Nível de Significância”, três períodos diferentes são listados, indicando que nesses momentos o espectro de potência foi estatisticamente significativo, onde as previsões indicam precipitações mais atuantes na região de estudo.

Tabela 1 – Tabela demonstrativa dos dados encontrados na Figura 9

Ano	Períodos com Nível de Significância
2018	11 de fevereiro a 01 de Junho
	20 de Abril a 01 de Maio
	01 de Outubro a 20 de Novembro
2019	01 de Março a 01 de Maio
2020	01 de Julho a 11 de Agosto
	01 a 11 de Agosto
2021	11 de Setembro a 11 de Novembro
2022	01 de Agosto a 20 de Novembro

Fonte: Autoria própria (2023).

Como resultado desse trabalho, obteve-se uma melhor compreensão dos dados meteorológicos relacionados à previsão de precipitação pluviométrica, considerados para a região específica (Seção 3.1.2) em termos de suas características geoespaciais e temporais. Para isso, dominou-se suficientemente a linguagem de programação (Seção 3.1.5) e as bibliotecas usadas para o tratamento desses dados, assim como se compreendeu o formato utilizado pelo ECMWF (Seção 3.1.1) na disponibilização dos dados meteorológicos. Também, pode-se obter uma melhor compreensão dos gráficos do espectro de potência de *wavelet*, que por sua vez fornecem informações importantes para a análise de séries temporais não-estacionárias.

5 CONCLUSÃO

A investigação de séries temporais derivadas de modelos preditivos é vital em vários campos científicos, tais como estatística, meteorologia, agricultura, economia e ciência de dados. Quando essa análise é conjugada com dados espaciais, ela fornece perspectivas valiosas acerca da distribuição espacial dos fenômenos, contribuindo para o reconhecimento de padrões geográficos e para a compreensão das interações entre as variáveis ao longo do tempo e do espaço. A transformada de *wavelet* emerge como um instrumento matemático fundamental nessa integração, pois permite a decomposição de informações em diversas escalas e frequências, desvendando assim, padrões e tendências que não são imediatamente aparentes.

Este estudo apresentou uma análise de séries temporais não estacionárias de previsão de chuvas, aplicando a transformada *wavelet* conforme proposto por Morlet *et al.* (1982) (Capítulo 3, Seção 2.2). As séries foram obtidas através do modelo TIGGE, que fornece previsões decendiais (de 10 dias) e estão acessíveis na plataforma do Centro Europeu de Previsões Meteorológicas a Médio Prazo, via um *script* em Python detalhado na Listagem 1 (Seção 3.1.4). Os dados, uma vez extraídos pelo método previamente descrito, foram processados utilizando-se *scripts* elaborados na linguagem R (Seção 3.1.5) e disponíveis integralmente através do endereço <https://github.com/andrelquiosi/WaveletsTimeSeries>, com o intuito de serem exibidos em tabelas, gráficos e figuras no Capítulo 4. As visualizações gráficas fornecidas facilitaram uma compreensão aprofundada do comportamento das séries temporais, permitindo derivar conclusões substantivas a partir das evidências apresentadas. Os gráficos, em particular, com suas escalas meticulosas e detalhamento minucioso, contribuem significativamente para a identificação de alterações e padrões emergentes.

A complexidade do processamento dos dados varia com o volume a ser analisado, podendo resultar em um processo oneroso em termos de manipulação, processamento e apresentação dos mesmos. Observou-se, através da revisão de estudos correlatos, que a aplicação da transformada *wavelet* está se expandindo intensamente, tanto no Brasil quanto internacionalmente, com usos diversificados, sendo notavelmente prevalente no tratamento de dados meteorológicos.

Para este estudo, delimitou-se a análise a dados de previsão de precipitação cobrindo um período de cinco anos. No entanto, destaca-se a possibilidade de extensão da análise para abranger um leque mais amplo de variáveis meteorológicas, disponíveis em bases de dados renomadas como a do ECMWF. A inclusão de variáveis inter-relacionadas pode potencializar a análise, proporcionando uma visão mais detalhada e conclusões mais robustas acerca dos fenômenos climáticos estudados.

REFERÊNCIAS

- ALMEIDA, E. C. S. d. **Agroindústrias: a indústria motriz que desenvolve a região Oeste do Paraná**. 2020. Dissertação (Mestrado), 2020.
- BOLZAN, M. J. A. Análise da transformada em ondeletas aplicada em sinal geofísico. **Revista Brasileira de Ensino de Física**, SciELO Brasil, v. 26, p. 37–41, 2004.
- BOLZAN, M. J. A. Transformada em ondeleta: Uma necessidade. **Revista Brasileira de Ensino de Física**, SciELO Brasil, v. 28, p. 563–567, 2006.
- LEPIK, Ü.; HEIN, H. Haar wavelets. *In: Haar Wavelets: With Applications*. [S.l.]: Springer, 2014. p. 7–20.
- LINA, J.-M.; MAYRAND, M. Complex daubechies wavelets. **Applied and Computational Harmonic Analysis**, Elsevier, v. 2, n. 3, p. 219–229, 1995.
- LOPES, G. R.; DELBEM, A. C.; SOUSA, J. B. de. Introdução à análise de dados geoespaciais com python. **Sociedade Brasileira de Computação**, 2021.
- MALHI, G. S.; KAUR, M.; KAUSHIK, P. Impact of climate change on agriculture and its mitigation strategies: A review. **Sustainability**, MDPI, v. 13, n. 3, p. 1318, 2021.
- MORETTIN, P. A.; TOLOI, C. M. **Análise de séries temporais: modelos lineares univariados**. [S.l.]: Editora Blucher, 2018.
- MORLET, J. *et al.* Wave propagation and sampling theory; Part I, Complex signal and scattering in multilayered media. **Geophysics**, v. 47, n. 2, p. 203–221, 02 1982. ISSN 0016-8033. Disponível em: <https://doi.org/10.1190/1.1441328>.
- NORDEMANN, D. J. R. Periodicidades, tendências e previsão a partir da análise espectral dinâmica da série dos níveis do rio paraguai, em ladário (1900/1995). **Pesquisa Agropecuária Brasileira**, v. 33, n. 13, p. 1787–1790, 1998.
- ROESCH, A.; SCHMIDBAUER, H. **WaveletComp: Computational Wavelet Analysis**. [S.l.], 2018. R package version 1.1. Disponível em: <https://CRAN.R-project.org/package=WaveletComp>.
- VETTERLI, M.; HERLEY, C. Wavelets and filter banks: Theory and design. **IEEE transactions on signal processing**, v. 40, n. ARTICLE, p. 2207–2232, 1992.
- WEI, D. **Coiflet-type wavelets: theory, design, and applications**. [S.l.]: The University of Texas at Austin, 1998.