



Gabriel Ribeiro Gomes Cifuentes Gonçalves

**Série de Manchas Solares: Evidências do Ciclo
Estocástico**

Relatório de Projeto Final de Graduação

Projeto apresentado ao Programa de Graduação , do Departamento de Engenharia Mecânica da PUC-Rio como requisito parcial para finalização do curso de Engenharia Mecânica.

Orientador: Cristiano Augusto Coelho Fernandes

Rio de Janeiro
February de 2025

Agradecimentos

A realização deste trabalho só foi possível graças ao apoio e orientação de diversas pessoas, a quem expresso minha gratidão.

Primeiramente, agradeço ao meu orientador, Prof. Dr. Cristiano Augusto Coelho, por sua dedicação, paciência e orientação ao longo deste percurso acadêmico. Seu conhecimento e incentivo foram fundamentais para a concretização deste trabalho.

Aos meus familiares, pelo apoio incondicional, compreensão e motivação em todos os momentos desta jornada. Sem o suporte e carinho de vocês, esta conquista não teria sido possível.

Por fim, agradeço a todos os professores e pesquisadores que, direta ou indiretamente, contribuíram para minha formação acadêmica e para o desenvolvimento deste trabalho.

Resumo

Ribeiro Gomes Cifuentes Gonçalves, Gabriel; . **Série de Manchas Solares: Evidências do Ciclo Estocástico**. Rio de Janeiro, 2025. 80p. Relatório de Projeto Final de Graduação – Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

O principal objetivo dessa monografia é investigar a estocasticidade do ciclo de aproximadamente 11 anos da série de manchas solares (1749-2024), conhecido como ciclo de Schwabe. Para isso, foram utilizadas as técnicas de Análise de Fourier (AF), Análise de Ondas (AO) e modelos estruturais para séries temporais.

Inicialmente, realizamos uma análise espectral via AF na série de manchas solares em 3 segmentos distintos (1749-1840, 1841-1932 e 1933-2024), confirmando a variabilidade temporal tanto da amplitude quanto do período do ciclo ao longo desses segmentos.

Em seguida através do escalograma da série de manchas solares, obtido via AO, pudemos observar visualmente a evolução das frequências e amplitudes da série de manchas solares ao longo do tempo.

Finalmente modelamos a série de manchas solares utilizando o modelo estrutural de Harvey [10], onde a série é decomposta nas componentes de tendência e ciclo estocásticos. Esse modelo foi utilizado para prever a série no período de 2013 a 2024. A sua acurácia preditiva foi comparada com um outro modelo que possuía a mesma tendência estocástica, mas um ciclo determinístico. As medidas de acurácia utilizadas (RMSE e MAE) indicaram a superioridade preditiva do modelo com ciclo estocástico. Esse resultado reforça as evidências sobre a natureza estocástica do ciclo de 11 anos da série de manchas solares.

Palavras-chave

Manchas solares; Ondas; Ciclo estocástico; Filtro de Kalman;.

Abstract

Ribeiro Gomes Cifuentes Gonçalves, Gabriel; (Advisor). **Sunspot Series: Evidence of the Stochastic Cycle**. Rio de Janeiro, 2025. 80p. Tese de Doutorado – Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

The main objective of this monograph is to investigate the stochasticity of the approximately 11-year cycle of the sunspot series (1749-2024), known as the Schwabe cycle. To achieve this, we employed Fourier Analysis (FA), Wavelet Analysis (WA), and structural models for time series.

Initially, we performed a spectral analysis using FA on the sunspot series divided into three distinct segments (1749-1840, 1841-1932, and 1933-2024), confirming the temporal variability of both the amplitude and period of the cycle across these segments.

Next, through the scalogram of the sunspot series obtained via WA, we visually observed the evolution of the series' frequencies and amplitudes over time.

Finally, we modeled the sunspot series using Harvey's structural model [10], where the series is decomposed into stochastic trend and cycle components. This model was then used to forecast the series for the period from 2013 to 2024. Its predictive accuracy was compared with another model that had the same stochastic trend but a deterministic cycle. The accuracy metrics used (RMSE and MAE) indicated the superior predictive performance of the model with a stochastic cycle. This result reinforces the evidence of the stochastic nature of the 11-year cycle in the sunspot series.

Keywords

Sunspots; Wavelets; Stochastic Cycle; Kalman Filte.

Sumário

1	Introdução	13
1.1	Introdução	13
1.2	Objetivos	14
2	Série Temporal de Manchas Solares	16
2.0.1	Fonte da Série de Manchas Solares	18
2.1	Ciclos da série de manchas solares	19
3	Metodologia	21
3.1	Análise de Fourier	21
3.1.1	Revisão teórica	21
3.1.2	Método dos Mínimos Quadrados Ordinários (MQO)	24
3.1.3	Cálculo dos Coeficientes	24
3.1.4	Estimativa de função espectral	25
3.1.5	Espectro de Fourier	25
3.1.6	Periodograma	26
3.1.7	Método de Welch	27
3.1.8	Exemplo	29
3.1.9	Série Temporal de Ventos	29
3.2	Análise de Ondaletas	32
3.2.1	Introdução à Análise de Ondaletas	33
3.2.2	Revisão teórica	34
3.2.3	Funções Ondaleta Mãe e Ondaleta Pai	34
3.2.4	Ondaleta de Haar	36
3.2.5	Ondaleta de Morlet	37
3.2.6	Decomposição e Reconstrução de Ondaletas	38
3.2.7	Transformada Discreta de Ondaletas (DWT)	38
3.2.8	Cálculo dos Filtros h_k e g_k	39
3.2.9	Exemplo Numérico da DWT para Ondaleta Haar	40
3.2.10	Transformada Contínua de Ondaletas (CWT)	42
3.2.11	Interpretação dos Parâmetros de Escala	43
3.2.12	Representação Visual: Escalograma	43
3.2.13	Especificação de Parâmetros em Ondaletas Contínuas	44
3.2.14	Reconstrução de Ondaletas	46
3.2.15	Exemplo Numérico da Reconstrução a partir dos Coeficientes	47
3.2.16	Filtragem e Redução de Ruído com Ondaletas	48
3.2.17	Exemplo	50
3.3	Modelos de Séries Temporais com Ciclos	51
3.3.1	Tendência	51
3.3.2	Ciclo	51
3.3.3	Irregular (Erro)	52
3.3.4	Importância da Decomposição	52
3.3.5	Modelo com Tendência e Ciclo	53
3.3.6	Modelos para Ciclo	53

3.3.7	a) Modelo de Ciclo Determinístico	53
3.3.8	b) Modelo de Ciclo Estocástico	54
3.3.9	Modelos para Tendência	55
3.3.10	a) Tendência Estocástica Estacionária	55
3.3.11	b) Tendência Estocástica Não Estacionária	55
3.3.12	c) Modelo de Tendência Constante	55
3.3.13	Modelo com Ciclo e Tendência Estocásticas	55
3.3.14	Estimativa dos Parâmetros	56
3.3.15	Espaço de Estado e Filtro de Kalman	56
3.3.16	Estimativa via Filtro de Kalman	57
3.3.17	Previsão	58
3.3.18	Aplicabilidade	59
3.4	Outras Ferramentas para Análise de Séries Temporais	59
3.4.1	Correlograma	59
3.4.2	Teste de Normalidade	60
3.4.3	Métricas para Quantificação da Acurácia Preditiva	61
3.5	Implementação	61
3.5.1	Análise de Fourier	62
3.5.2	Análise de Ondaletas	62
3.5.3	Modelos Estruturais para Ciclo Estocástico	63
4	Resultados	64
4.1	O ciclo da série de manchas solares	64
4.2	Evidências da estocasticidade do ciclo da série de manchas solares	66
4.2.1	Análise exploratória	66
4.2.2	Evidências suplementares	69
4.2.3	Aplicação da Análise de Ondaletas	69
4.3	Modelagem da Série de Manchas Solares via Modelo de Tendência e Ciclo Estocásticos	73
4.3.1	Ajuste do Modelo com Ciclo Determinístico Monocromático	74
5	Conclusão	77
	Referências bibliográficas	79

Lista de figuras

Figura 2.1	Imagem de manchas solares observadas pelo satélite SOHO.	16
Figura 2.2	Série temporal anual do número médio de manchas solares (1749-2024).	19
Figura 2.3	Espectro de amplitude da série temporal anual de manchas solares (1700-2000).	20
Figura 3.1	Exemplo da representação gráfica do espectro de Fourier.	26
Figura 3.2	Exemplo de periodograma.	26
Figura 3.3	Exemplo do periodograma da Fig(3.2) suavizado pelo método de Welch.	28
Figura 3.4	Série de ventos em São Paulo.	29
Figura 3.5	Correlograma da série de ventos em São Paulo.	30
Figura 3.6	Espectro de linha de Fourier da série de ventos.	30
Figura 3.7	Periodograma da série de ventos.	32
Figura 3.8	Ondaleta mãe de Haar.	36
Figura 3.9	Ondaleta pai de Haar.	37
Figura 3.10	Comparação entre um impulso mecânico (a) e a ondaleta de Morlet (b).	37
Figura 3.11	Escalograma ilustrativo de uma série com ciclo determinística.	45
Figura 3.12	Escalograma ilustrativo de uma série com amplitude e frequência variáveis.	46
Figura 3.13	Exemplo de decomposição de uma série temporal ruidosa utilizando a ondaleta <i>Daubechies-4</i> . Cada nível de decomposição separa componentes do sinal em diferentes escalas, evidenciando tendências e detalhes.	49
Figura 3.14	Série temporal de ventos após aplicação do filtro com ondaletas <i>Daubechies-4</i> (<i>db4</i>). As oscilações cíclicas são mais evidentes após a remoção do ruído.	50
Figura 3.15	Exemplo de correlograma de uma série temporal mensal.	60
Figura 4.1	Série temporal do número médio de manchas solares (1749-2024).	65
Figura 4.2	Correlograma da série temporal de manchas solares.	66
Figura 4.3	Periodograma suavizado da série de manchas solares.	67
Figura 4.4	Série temporal dos máximos consecutivos das manchas solares.	67
Figura 4.5	Série temporal dos mínimos consecutivos das manchas solares.	68
Figura 4.6	Série da variação do período dos ciclos solares.	68
Figura 4.7	Decomposição wavelet da série de manchas solares.	70
Figura 4.8	Correlograma da componente de ciclo obtida por DWT.	71
Figura 4.9	Períodos dos 24 ciclos solares calculados a partir da componente de ciclo extraída por ondaletas.	72
Figura 4.10	Escalograma da série de manchas solares.	73

Figura 4.11	Previsão do modelo com ciclo determinístico (2012-2024).	76
Figura 4.12	Previsão do modelo com ciclo estocástico (2012-2024).	76

Lista de tabelas

Tabela 2.1	Ciclos solares observáveis.	20
Tabela 3.1	Resultado para alguns coeficientes da FFT da série de Ventos.	30
Tabela 3.2	Resultado da regressão OLS na série de ventos.	32
Tabela 3.3	Comparação entre Decomposição e Reconstrução utilizando Ondaletas	47
Tabela 4.1	Valores da correlação da série de manchas solares.	65
Tabela 4.2	Coefficientes da FFT para a série de manchas solares.	66
Tabela 4.3	Estatísticas descritivas da série dos períodos do ciclo solar	69
Tabela 4.4	Frequências dominantes e períodos estimados para cada segmento.	69
Tabela 4.5	Autocorrelação para a componente de ciclo obtida por DWT.	71
Tabela 4.6	Tabela de parâmetros para modelo ajustado à série de manchas solares.	73
Tabela 4.7	Parâmetros do segmento 1 para série de manchas solares.	74
Tabela 4.8	Parâmetros do segmento 2 para série de manchas solares.	74
Tabela 4.9	Parâmetros do segmento 3 para série de manchas solares.	74
Tabela 4.10	Comparação dos testes de aderência para os modelos com ciclo determinístico e ciclo estocástico.	76

1

Introdução

1.1

Introdução

A série de manchas solares, que registra a atividade magnética solar, é um exemplo icônico de série temporal com grande relevância científica. Suas primeiras medições datam em torno de 1600, apresentando um comportamento cíclico com período aproximado de 11 anos, também conhecido como Ciclo de Schwabe. As manchas dizem respeito a áreas escurecidas na superfície visível solar, onde observações astronômicas ao longo de vários séculos mostram que o número de manchas aumenta de um mínimo para um máximo e, em seguida, retorna ao mínimo.

Estudos indicam que essa série apresenta ciclos estocásticos de aproximadamente 11 anos, cuja periodicidade e amplitude variam ao longo do tempo. De acordo com Charbonneau (2005) [14]:

“At any rate, the notion of a nicely regular 11/22-year cycle does not hold long upon even cursory scrutiny, as the amplitude of successive cycles is clearly not constant, and their overall shape often differs significantly from one cycle to another. ... Even the cycle's duration is not uniform, spanning in fact a range going from 9 yr (cycle 2) to nearly 14 yr (cycle 4).”

A compreensão desses ciclos é crucial para a climatologia e a engenharia espacial, em especial, uma vez que a atividade solar pode impactar diretamente o clima terrestre e sistemas tecnológicos, como satélites e redes elétricas.

Na parte inicial desse trabalho utilizaremos as técnicas de Análise de Fourier e de Análise de Ondaletas com intuito de melhor compreender a natureza do período do ciclo da série de manchas solares. Essas ferramentas matemáticas são amplamente reconhecidas por sua capacidade de identificar padrões cíclicos em séries temporais (AF) e de decompor sinais em componentes de alta e baixa frequência (AO). Mais especificamente essas técnicas serão utilizadas com 2 objetivos: i) utilizar a AF para investigar a periodicidade da série de manchas solares em diversos segmentos temporais, de forma a ter uma melhor idéia da estocasticidade do seu ciclo; ii) através da AO dividir a série de manchas solares

em componentes de interesse como tendência, ciclo e erro para poder fazer uma investigação focada na componente de ciclo. A transformada contínua de ondaleta também foi usada para nos dar uma representação visual da evolução das frequências através do escalograma, e mostrar como a AO é uma técnica versátil para análise de séries temporais. Na segunda etapa do nosso trabalho iremos utilizar o modelo estrutural de Harvey [10], que decompõe uma série temporal nas componentes de ciclo, tendência e erro. Esse modelo será aplicado com objetivo de prever a série alguns anos à frente em dois modelos diferentes: tendência e ciclo estocástico e tendência estocástica com ciclo determinístico. Assim, comparando a acurácia preditiva para os dois modelos de estrutura, contribuindo para o entendimento da dinâmica estocástica do ciclo da série de manchas solares.

1.2

Objetivos

Este trabalho tem como objetivo principal investigar a estocacidade do ciclo de aproximadamente 11 anos presente na série de manchas solares. A abordagem adotada combina técnicas de Análise de Fourier (AF), Análise de Ondaletas (AO) e o modelo estrutural de ciclo e tendência estocásticos (ver Harvey [10]), permitindo uma compreensão mais aprofundada das variações observadas na periodicidade e amplitude desse ciclo. Os objetivos específicos do trabalho são:

- Desenvolver uma base teórica sólida sobre as técnicas de AF e AO, destacando suas vantagens, limitações e aplicações em séries temporais discretas, estacionárias e não estacionárias.
- Analisar a periodicidade e a amplitude do ciclo das manchas solares em diferentes segmentos da série temporal, utilizando AF para identificar as principais frequências harmônicas e estimar variações no ciclo ao longo do tempo.
- Aplicar a decomposição e a reconstrução de ondaletas (DWT) para realizar uma análise multirresolução e explorar a influência de diferentes configurações de reconstrução na dinâmica do ciclo.
- Desenvolver dois modelos estruturais com tendência estocástica: um com ciclo estocástico e outro com ciclo determinístico, comparando suas acurácias preditivas para avaliar qual modelo melhor previu a série no período de teste de 12 anos (2012-2024).

Com esta abordagem, espera-se contribuir para o entendimento da dinâmica do ciclo de manchas solares, demonstrando como as técnicas de AF e AO podem complementar-se na análise de fenômenos cíclicos complexos.

A estrutura deste trabalho está organizada da seguinte forma:

- **Capítulo 1 - Introdução:** apresenta o contexto e a relevância científica do estudo das manchas solares, destacando a natureza estocástica de seu ciclo de aproximadamente 11 anos e suas implicações para diversas áreas, como a climatologia e a engenharia espacial.
- **Capítulo 2 - Série de Manchas Solares:** apresenta uma revisão histórica e técnica sobre a origem e a importância da série de manchas solares, descrevendo os métodos de coleta e suas propriedades estatísticas.
- **Capítulo 3 - Metodologia:** apresenta, de forma sucinta e prática, os procedimentos estatísticos utilizados na nossa investigação da série de manchas solares: a Análise de Fourier, a Análise de Ondas e o modelo estrutural de ciclo e tendência estocásticos. Este capítulo também apresenta algumas ferramentas utilizadas e as bibliotecas utilizadas na implementação em Python.
- **Capítulo 4 - Aplicação:** discute a aplicação prática das técnicas à série temporal de manchas solares. Além disso, serão apresentadas evidências que sustentam a hipótese de estocasticidade do ciclo de Schwabe. Neste capítulo, também serão propostos dois modelos estruturais para previsão da série temporal, permitindo avaliar qual abordagem melhor se ajusta aos dados e apresenta maior acurácia preditiva.
- **Capítulo 5 - Conclusão:** apresenta um resumo dos principais achados, destacando as contribuições do trabalho e propondo direções para estudos futuros.

Neste trabalho, as técnicas de AF e AO serão implementadas em Python, utilizando ferramentas como *NumPy*, *Matplotlib* e *PyWavelets*. O modelo estrutural foi estimado a partir do software *Time Series Lab*[15]. O ciclo de Schwabe de aproximadamente 11 anos da série de manchas solares será usado como estudo de caso principal, buscando evidenciar e confirmar a estocasticidade desse ciclo.

2

Série Temporal de Manchas Solares

Introdução

A contagem de manchas solares é uma das mais antigas práticas de observação astronômica, desempenhando um papel fundamental na compreensão da atividade solar e suas influências na Terra. Desde o século XVII, astrônomos como Galileo Galilei e Christoph Scheiner registraram manchas solares, marcando o início do estudo sistemático desses fenômenos. No século XIX, Rudolf Wolf padronizou a contagem de manchas solares, introduzindo o chamado Números de Wolf, que consideram tanto o número de manchas individuais quanto os grupos de manchas em uma fórmula padronizada.

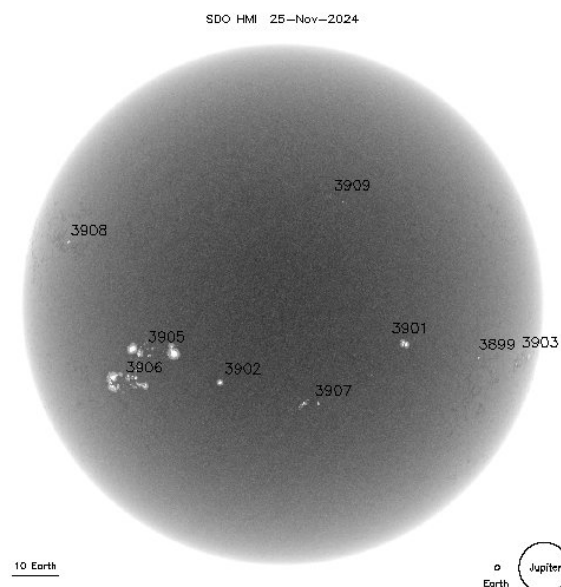


Figura 2.1: Imagem de manchas solares observadas pelo satélite SOHO.

Fonte: SOHO (Solar and Heliospheric Observatory), NASA/ESA [23].

A análise das manchas solares é essencial para compreender o comportamento da atividade solar e suas implicações tanto no clima terrestre quanto em sistemas tecnológicos que sejam suscetíveis a radiação solar.

“A atividade solar aumenta e diminui com um ciclo de 11 anos que afeta a vida moderna de muitas maneiras. O aumento da atividade

solar inclui incrementos nas emissões de raios ultravioleta extremo e raios-X provenientes do Sol, que produzem efeitos dramáticos na alta atmosfera terrestre. O aquecimento atmosférico associado aumenta tanto a temperatura quanto a densidade da atmosfera em muitas altitudes de satélites. O aumento no arrasto atmosférico em satélites em órbitas baixas pode encurtar drasticamente a vida útil orbital desses ativos valiosos” [16].

As manchas solares, originadas por intensos campos magnéticos na superfície do Sol, seguem um ciclo periódico aproximado de 11 anos. Durante o máximo desse ciclo, a quantidade de manchas solares aumenta substancialmente, resultando em maior atividade magnética solar, como por exemplo, erupções solares e ejeções de massa coronal. Uma descrição resumida e precisa dada por [12] desse fenômeno é dada a seguir.

“As manchas solares são regiões de campos magnéticos extremamente fortes na superfície do Sol. Esses campos magnéticos podem se torcer, armazenando energia de forma semelhante a um elástico tensionado. Elas aparecem escuras na superfície do Sol porque são mais frias do que as regiões ao seu redor. Os campos magnéticos que criam as manchas solares impedem que parte do calor do Sol alcance sua superfície. Quando as manchas solares interagem entre si, elas causam explosões de energia.”

A identificação e previsão do ciclo solar são de grande importância para diversas áreas. No contexto científico, esses estudos auxiliam na compreensão da dinâmica solar e das interações entre o Sol e a Terra. Por exemplo, há evidências de que os ciclos solares podem influenciar padrões de seca e precipitação em regiões tropicais, além de afetar o balanço energético do planeta.

Além disso, a previsão do ciclo solar tem aplicações práticas e tecnológicas. Durante os períodos de alta atividade solar, fenômenos como as erupções solares podem causar impactos significativos em sistemas de comunicação por rádio, redes elétricas e satélites. A radiação emitida em eventos solares extremos também representa um risco para astronautas em missões espaciais. Assim, previsões precisas permitem implementar medidas preventivas, como colocar satélites em modo de segurança e adiar caminhadas espaciais. Ver por exemplo [8].

O estudo estatístico das séries temporais de manchas solares fornece uma base científica robusta para monitorar e prever os ciclos solares. Essa informação não apenas contribui para o avanço do conhecimento astronômico, mas também auxilia na proteção de tecnologias essenciais e na mitigação de riscos associados à atividade solar.

2.0.1

Fonte da Série de Manchas Solares

A série temporal de manchas solares é um registro do número observado de manchas na superfície visível do Sol ao longo do tempo.

“Atualmente, este índice de atividade está disponível em duas formas principais: o Número Internacional de Manchas Solares, iniciado por R. Wolf em 1849, e o Número de Grupos, construído mais recentemente por Hoyt e Schatten (1998a, b). Infelizmente, essas duas séries não coincidem em diversos aspectos, o que causa confusões e contradições quando usadas em estudos contemporâneos cruciais sobre o dínamo solar ou sobre a influência solar no clima da Terra.”

Frédéric Clette, 2014[7]

Em 1849, R. Wolf do Observatório de Zurique propôs a fórmula que estima o número de manchas solares, ainda hoje utilizada, que é apresentada a seguir (detalhes sobre o método para calcular o número total de manchas solares é definido em [22]).

$$R = k \times (10g + s), \quad (2-1)$$

onde:

- R é o número de manchas solares de Wolf (ou número relativo de manchas solares).
- g é o número de grupos de manchas solares observados.
- s é o número total de manchas solares individuais observadas em todos os grupos.
- k é o fator de correção para diferenças nas técnicas de observação.

Explicação da fórmula (eq. (2-1)):

1. **Ponderação de Grupos:** Cada grupo de manchas é multiplicado por 10, pois grupos geralmente representam regiões maiores e mais ativas no Sol, refletindo sua maior influência na atividade solar.
2. **Contagem de Manchas Individuais:** As manchas individuais dentro dos grupos são contadas e adicionadas ao total.
3. **Fator de Correção (k):** Este fator ajusta variações entre diferentes observadores, telescópios ou condições ambientais. Cada observador determina seu próprio fator k com base em comparações com observações padrão.

A média mensal do número de manchas solares é calculada como a média dos números diários de manchas solares de Wolf ao longo de um mês. De forma semelhante, a média anual é obtida pela média dos valores mensais ao longo de um ano. Esse processo de consolidação é importante para suavizar variações de curto prazo, como flutuações diárias e ruídos instrumentais, fornecendo uma visão mais clara dos ciclos solares de longo prazo.

A Fig. (2.2) apresenta a série temporal anual das manchas solares com início em 1749 e fim em 2024, perfazendo um total de $N = 275$ observações.

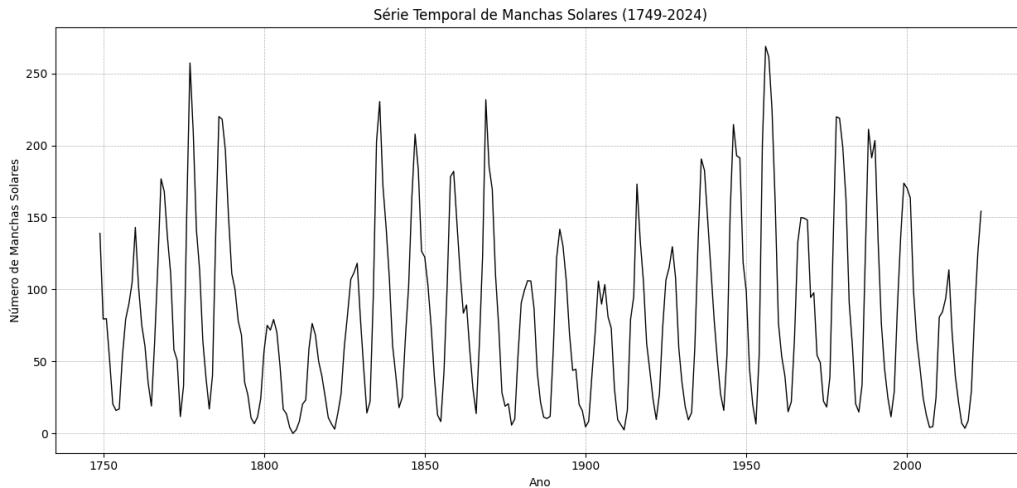


Figura 2.2: Série temporal anual do número médio de manchas solares (1749-2024).

Fonte: Solar Influences Data Analysis Center (SIDC).[9]

A precisão das medições é garantida por procedimentos rigorosos de calibração e validação, que incluem a comparação entre diferentes observatórios e o uso de técnicas modernas de processamento de dados. Esses cuidados são essenciais para assegurar a consistência histórica da série temporal, que remonta ao ano de 1749, permitindo uma análise detalhada dos ciclos solares e suas possíveis correlações com fenômenos climáticos e outros eventos naturais.

2.1

Ciclos da série de manchas solares

A Fig 2.3 apresenta o periodograma, obtido para a série de manchas solares a partir de um estudo realizado por [25]. A informação contida nessa figura (vide o Capítulo 3, seção 3.1) permite inferir a existência de 5 ciclos principais nessa série, sendo o ciclo de (aproximadamente) 11 anos, o ciclo de Schawbe, o de maior relevância, e o que será objeto de nossa investigação.

A partir da informação da figura anterior construímos a tabela a seguir, com os principais ciclos da série de manchas solares.

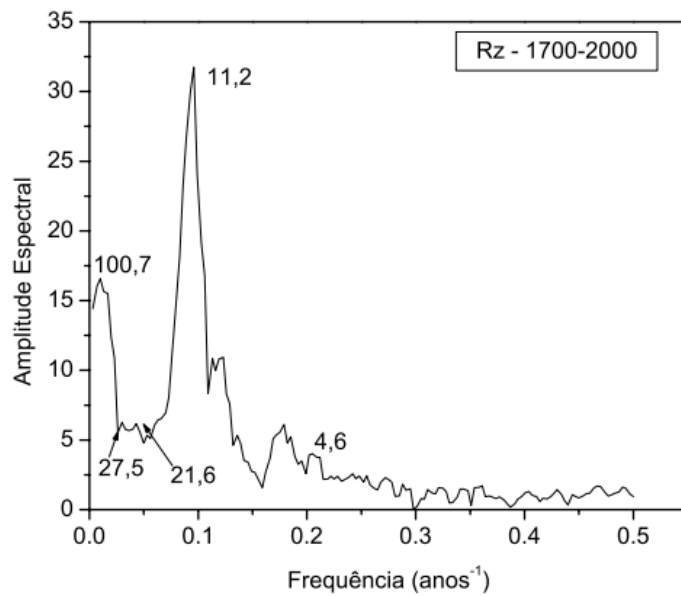


Figura 2.3: Espectro de amplitude da série temporal anual de manchas solares (1700-2000).

Fonte: Echer et al. (2003).[25]

Nome do ciclo	Período (anos)
Segundo harmônico de Schwabe	4,6
Ciclo de Schwabe	11
Ciclo magnético de 22 anos de Hale	21,6
Ciclo de Gleissberg	100,7

Tabela 2.1: Ciclos solares observáveis.

Fonte: Echer et al. (2003).[25]

Assim sendo, embora a série de manchas solares apresente 3 ciclos, o nosso trabalho estará focado no ciclo de 11 anos, ou ciclo de Schwabe.

3

Metodologia

Neste capítulo, apresentaremos de forma sucinta os métodos utilizados para investigar a natureza estocástica do ciclo da série de manchas solares. Abordaremos as técnicas de Análise de Fourier (AF) e Análise de Ondas (AO) para decomposição e caracterização dos padrões cíclicos na série. Além disso, nesse capítulo apresentaremos um modelo para série temporal com ciclos, denominado de modelo estrutural. A partir desse modelo, podemos melhor investigar a natureza estocástica do ciclo da série e realizar previsões para essa série.

3.1

Análise de Fourier

Neste subcapítulo, iremos descrever as técnicas de **Análise de Fourier**, que serão aplicadas na análise da série temporal de ventos. A seguir, detalharemos a teoria e o processo de aplicação dessa técnica, incluindo os procedimentos de processamento dos sinais e as ferramentas computacionais utilizadas.

“A razão pela qual a Análise de Fourier é tão importante na física é que muitas (embora certamente não todas) das equações diferenciais que governam sistemas físicos são lineares, o que implica que a soma de duas soluções também é uma solução. Portanto, como a Análise de Fourier nos mostra que qualquer função pode ser escrita em termos de funções seno e cosseno, podemos limitar nossa atenção a essas funções ao resolver as equações diferenciais. Em seguida, podemos construir qualquer outra função a partir dessas funções específicas. Esta é uma estratégia muito útil, pois é invariavelmente mais fácil lidar com funções seno e cosseno do que com funções gerais.” Morin (2009) [24]

3.1.1

Revisão teórica

Uma função periódica é uma função que repete seus valores em intervalos regulares. Matematicamente, uma função $f(t)$ é considerada periódica se, para um determinado período $T > 0$, a seguinte condição for satisfeita:

$$f(t + T) = f(t) \quad (3-1)$$

para todo $t \in \mathbb{N}$.

O menor valor positivo de T que satisfaz essa condição é chamado de período da função. A forma mais simples de representar uma função periódica é através de harmônicos, que são componentes senoidais ou cossenoidais com diferentes frequências, amplitudes e fases. Um harmônico básico pode ser expresso como:

$$f(t) = A \cos(\lambda t + \phi) \quad (3-2)$$

onde A é a amplitude, λ é a frequência angular, que está relacionada à frequência da função por $\lambda = 2\pi f$, f é a frequência em hertz (Hz), ϕ é a fase e o período T da função será dado por $T = \frac{2\pi}{\lambda}$. Essa combinação de amplitude, frequência e fase descreve completamente o comportamento de um harmônico em uma função periódica.

A **Análise de Fourier** é uma técnica para decompor uma função ou sinal em suas componentes harmônicas. Este método permite que qualquer sinal periódico seja representado como uma soma de funções senoidais de diferentes frequências, com cada frequência associada a um coeficiente específico.

“O objetivo básico é de aproximar uma função $f(t)$ por uma combinação linear de componentes senoidais, cada uma com dada frequência. O conjunto $\{w_n(t) = e^{int}, n = 0, \pm 1, \dots\}$ de funções ortogonais, de período 2π , forma a base para a análise de Fourier. Na realidade, esse conjunto é gerado por dilatações de uma única função, $w(t) = e^{it}$, ou seja, $w_n(t) = w(nt)$ para qualquer n inteiro” (MORETTIN, 2014).[3]

A dilatação da função é dada pelo acréscimo de n , que representa o número do harmônico. Toda função periódica, de período 2π , e de quadrado integrável, pode ser gerada por uma superposição de dilatações inteiras da função $w(t)$. Isso significa que qualquer função periódica $f(t)$ pode ser expressa como uma combinação linear de funções harmônicas elementares.

Para uma função periódica $f(t)$, $t \in \mathbb{R}$ com período T , sua representação em série de Fourier pode ser escrita da seguinte forma:

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos(\lambda_n t) + b_n \sin(\lambda_n t)) \quad (3-3)$$

onde $\lambda_n = \frac{2\pi n}{T}$ são as frequências de Fourier e a_n e b_n são os coeficientes de Fourier, calculados como:

$$a_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \cos(\lambda_n t) dt, \quad n \geq 0 \quad (3-4)$$

$$b_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \sin(\lambda_n t) dt, \quad n \geq 1 \quad (3-5)$$

Supondo que o sinal contínuo está sendo amostrado em intervalos iguais no tempo, então o tempo pode ser tratado de forma discreta. Vamos chamar de f_j a série $f(t)$ na forma discreta. Para o caso específico de uma função com o tempo discreto e frequência discreta, podemos obter os coeficientes a_n e b_n por:

$$a_n^N = \frac{2}{N} \sum_{j=0}^{N-1} f_j \cos(\lambda_n j), \quad (3-6)$$

$$b_n^N = \frac{2}{N} \sum_{j=0}^{N-1} f_j \sin(\lambda_n j), \quad (3-7)$$

onde $n = 0, 1, 2, \dots, N$ é o tamanho do sinal após discretização.

Para uma série real $\{y_t = 1, \dots, N\}$, com N par, podemos escrever:

$$y_t = a_0^{(N)} + \sum_{j=1}^{N/2} \{a_j^{(N)} \cos(\lambda_j t) + b_j^{(N)} \sin(\lambda_j t)\}, \quad (3-8)$$

lembrando que $a_0^{(N)} = \bar{y}$ e $\lambda_j = \frac{2\pi j}{N}$ são as frequências de Fourier. A energia média/variação da série é dada por:

$$s_N^2 = \frac{1}{N} \sum_{t=0}^{N-1} y_t^2 = [a_0^{(N)}]^2 + \frac{1}{2} \sum_{j=1}^{N/2} [r_j^{(N)}]^2. \quad (3-9)$$

Também usaremos os valores a_n e b_n para calcular r_n :

$$r_n^{(N)} = \sqrt{[a_n^{(N)}]^2 + [b_n^{(N)}]^2}, \quad (3-10)$$

Na forma complexa, a transformada discreta de Fourier (DFT) da sequência f_j é dada por:

$$F_n = \sum_{j=0}^{N-1} f_j e^{-i\lambda_n j} \quad (3-11)$$

ou então, pela identidade de Euler:

$$F_n = \sum_{j=0}^{N-1} f_j (\cos(\lambda_n j) - i \sin(\lambda_n j)) \quad (3-12)$$

Para otimizar os cálculos da Transformada Discreta de Fourier (DFT) em sinais com muitos dados, utiliza-se a *Fast Fourier Transform* (FFT), um algoritmo computacionalmente eficiente desenvolvido por Cooley e Tukey em 1965 [6]. A FFT reduz o número de cálculos necessários para obter a DFT, passando de N^2 operações em uma DFT direta, para aproximadamente $N \log_2 N$ operações na FFT. Essa eficiência torna a FFT essencial para aplicações que envolvem grandes conjuntos de dados, como processamento de sinais e análise de séries temporais de altíssimas frequências.

A FFT requer que o número de pontos N da série seja uma potência de dois, ou seja, que N seja par e preferencialmente uma potência de 2 (por exemplo,

$N = 2^k$ para algum inteiro k). Essa restrição simplifica o algoritmo, permitindo a divisão recursiva dos cálculos em etapas menores, o que contribui para a redução drástica de operações.

Segundo Meyer (1993),[5] “a análise de Fourier pode ser vista como o algoritmo a ser usado quando tratamos de processos estocásticos estacionários periódicos”, visto que as funções seno e cosseno representam bem o comportamento estacionário da série.

3.1.2

Método dos Mínimos Quadrados Ordinários (MQO)

O Método dos Mínimos Quadrados Ordinários (MQO) é uma técnica estatística amplamente utilizada para estimar os coeficientes de uma função linear que melhor se ajusta a um conjunto de dados observados. No contexto de séries temporais, o MQO permite ajustar uma função de componentes senoidais e cosenoidais a uma série temporal, de modo a decompor o sinal em suas componentes harmônicas, semelhante ao que ocorre na Análise de Fourier.

3.1.3

Cálculo dos Coeficientes

Assumimos que a série temporal, estacionária e periódica, é representada por uma soma de senos e cossenos, de acordo com a Eq. (3-3). Ou seja:

$$y_t = a_0^{(N)} + \sum_{j=1}^{N/2} \left\{ a_j^{(N)} \cos(\lambda_j t) + b_j^{(N)} \sin(\lambda_j t) \right\} + \varepsilon_t, \quad (3-13)$$

onde ε_t é o termo aleatório. Nessa equação os termos a_n e b_n serão estimados pelo princípio de MQO, ou seja, minimizando a soma do quadrado dos resíduos $S(\hat{\beta})$ dado por:

$$S(\hat{\beta}) = \sum_{t=1}^N \varepsilon_t^2 = \sum_{t=1}^N (y_t - \hat{y}_t)^2 \quad (3-14)$$

onde:

- $\hat{\beta} = (\hat{a}_0, \hat{a}_1, \hat{b}_1, \hat{a}_2, \hat{b}_2, \hat{a}_3, \hat{b}_3, \dots, \hat{a}_{N/2}, \hat{b}_{N/2})^T$ é o vetor de estimativas dos coeficientes de Fourier;
- y_t é a observação da série em t ;
- \hat{y}_t é o ajuste dado pela expansão de Fourier:

$$\hat{y}_t = \hat{a}_0 + \sum_{j=1}^{N/2} \left(\hat{a}_j^{(N)} \cos(\lambda_j t) + \hat{b}_j^{(N)} \sin(\lambda_j t) \right),$$

E assim a solução de MQO será dada por:

$$\hat{\beta} = \min S(\hat{\beta}) \quad (3-15)$$

É fato conhecido que esse problema de minimização possui solução analítica, dada por:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (3-16)$$

de dimensão $(N+1) \times 1$, onde :

- $y = (y_1, y_2, \dots, y_N)^T$ de dimensão $N \times 1$ é o vetor dos dados;
- X é a matriz de “regressores” de dimensão $N \times (N + 1)$, que inclui as funções trigonométricas nas suas diferentes frequências, sendo dada por:

$$X = \begin{bmatrix} 1 & \cos(\lambda_1) & \sin(\lambda_1) & \cos(\lambda_2) & \sin(\lambda_2) & \dots & \sin(\lambda_{N/2}) & \cos(\lambda_{N/2}) \\ 1 & \cos(2\lambda_1) & \sin(2\lambda_1) & \cos(2\lambda_2) & \sin(2\lambda_2) & \dots & \sin(2\lambda_{N/2}) & \cos(2\lambda_{N/2}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 1 & \cos(N\lambda_1) & \sin(N\lambda_1) & \cos(N\lambda_2) & \sin(N\lambda_2) & \dots & \sin(N\lambda_{N/2}) & \cos(N\lambda_{N/2}) \end{bmatrix}$$

Pode-se também mostrar que as estimativas para os coeficientes de Fourier obtidos via MQO coincidirão com aqueles especificados nas equações (3-6) e (3-7). Esse resultado baseia-se na equivalência entre os estimadores de MQO e os estimadores do Método dos Momentos, que são exatamente aqueles dados em (3-6) e (3-7).

3.1.4

Estimativa de função espectral

3.1.5

Espectro de Fourier

O espectro de linhas de Fourier é uma representação visual que destaca a contribuição de cada harmônico específico para a variância total de uma série temporal, vide Eq. (3-9) e (3-10), mais especificamente, trata-se de um gráfico de $(\frac{(r_n^N)^2}{2} \times n)$, conforme Fig. (3.1).

Para calcular o espectro de linhas de Fourier, utilizamos os coeficientes a_n e b_n para cada harmônico n , que representam as amplitudes das componentes cosenoidais e senoidais da série, respectivamente. A contribuição de cada harmônico para a variância total pode ser expressa pela magnitude r_n , definida na equação (3-10). O valor r_n indica a intensidade de cada frequência harmônica no sinal, permitindo que o espectro de linhas mostre quais frequências possuem maior impacto na série temporal.

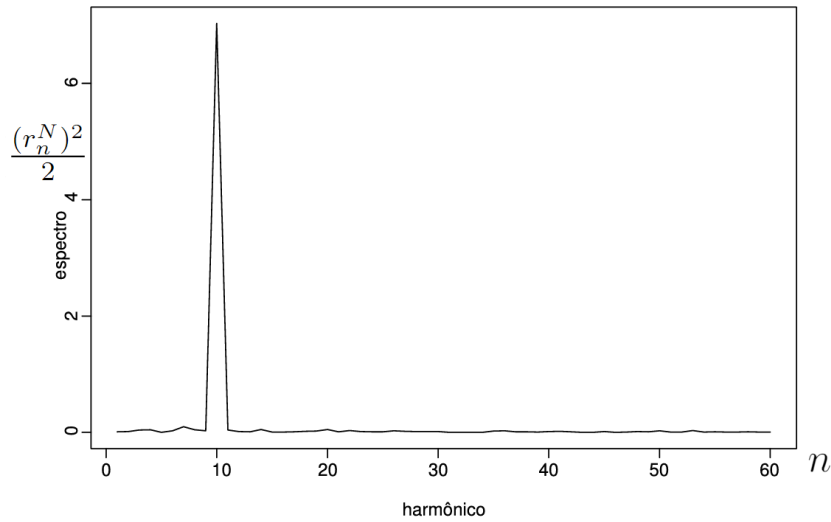


Figura 3.1: Exemplo da representação gráfica do espectro de Fourier.

3.1.6 Periodograma

O periodograma é uma estimativa da densidade espectral de potência que representa a distribuição de energia do sinal ao longo das frequências.

Na prática, o sinal é obtido aplicando-se a transformada de Fourier à série temporal e, em seguida, calculando o quadrado da magnitude dessa transformada para cada frequência. Para uma série temporal $y(t)$ com N pontos de dados, o periodograma $P(\lambda_n)$ é dado por:

$$P(\lambda_n) = \frac{1}{N} |F(\lambda_n)|^2 \quad (3-17)$$

onde $F(\lambda_n)$ é a Transformada de Fourier Discreta do sinal $y(t)$ dado em (3-11), e λ_n representa as frequências discretas.

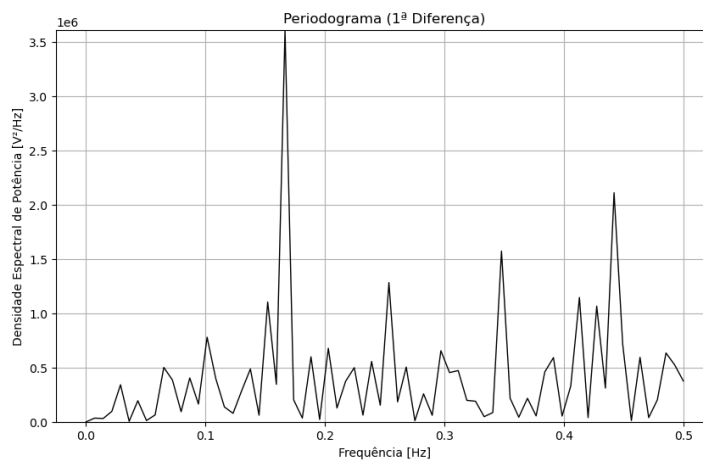


Figura 3.2: Exemplo de periodograma.

Observamos que na Fig (3.2) o periodograma apresenta no eixo horizontal as frequências em Hz, que representam o número de ciclos por segundo de cada componente do sinal. No eixo vertical, temos a potência espectral, medida em unidades quadráticas da amplitude original do sinal, que indica a intensidade de cada frequência presente. Os picos no periodograma identificam as frequências dominantes, ou seja, aquelas que mais contribuem para a composição do sinal analisado.

3.1.7

Método de Welch

O método de Welch é uma técnica de estimativa espectral que reduz a variância do periodograma através da segmentação e média dos espectros. Similar ao periodograma tradicional, o eixo x representa as frequências (Hz), enquanto o eixo y mostra a densidade espectral de potência, porém com uma estimativa mais suave e estável.

A estimativa de função *densidade espectral de potência* pelo método de Welch pode ser representada pela seguinte equação:

$$P_{Welch}(\lambda_n) = \frac{1}{K} \sum_{k=1}^K P_k(\lambda_n) \quad (3-18)$$

Onde:

- $P_{Welch}(\lambda_n)$ é a estimativa da densidade espectral de potência para uma dada frequência λ_n ;
- K é o número total de segmentos em que o sinal foi dividido;
- $P_k(\lambda_n)$ é a densidade espectral de potência calculada para o segmento k ;

Cada segmento do sinal é suavizado com uma janela de ponderação (por exemplo, a janela de Hanning) para reduzir a distorção causada pela descontinuidade nos limites dos segmentos.

A densidade espectral de potência para cada segmento $P_k(\lambda_n)$ é dada por:

$$P_k(\lambda_n) = \frac{1}{N} \left| \sum_{n=0}^{N-1} y_k(n) e^{-j\lambda_n n} \right|^2 \quad (3-19)$$

Onde:

- N é o número de pontos em cada segmento;
- $y_k(n)$ é o valor do sinal no segmento k ;

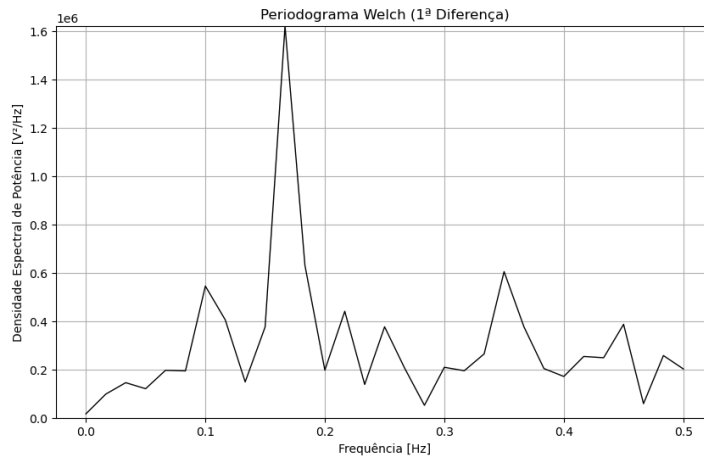


Figura 3.3: Exemplo do periodograma da Fig(3.2) suavizado pelo método de Welch.

– λ_n é a frequência analisada;

Em suma, o espectro de linha de Fourier fornece uma decomposição direta dos harmônicos do sinal, apresentando a amplitude de cada componente de frequência com base nos coeficientes de Fourier a_n e b_n . Esse método é particularmente eficaz para sinais periódicos, oferecendo alta resolução em frequências discretas.

Em contraste, o periodograma calcula a densidade espectral de potência utilizando a FFT, possibilitando uma visão rápida da distribuição de energia em frequências discretas. Porém, ele é sensível a variações e ruídos, especialmente em sinais de curta duração, resultando em estimativas menos estáveis. O periodograma suavizado de Welch aplica o método de média em segmentos sobrepostos, reduzindo a variância na estimativa da densidade espectral e tornando-se uma escolha robusta para sinais ruidosos e não estacionários.

Em síntese, enquanto o espectro de linha de Fourier é mais preciso para frequências discretas, o periodograma suavizado de Welch oferece uma estimativa mais estável e suave da potência espectral para sinais complexos.

3.1.8

Exemplo

Na sequência, iremos ilustrar a aplicação da FFT especificamente na série temporal de ventos. Para essa série, também iremos estimar os coeficientes da FFT e compará-los com aqueles obtidos através do método de MQO, destacando os padrões cíclicos presentes nos dados.

3.1.9

Série Temporal de Ventos

A Figura (3.4) apresenta a série de velocidade dos ventos (em m/s) em São Paulo, coletada com frequência mensal, onde cada medição é a média aritmética de todas as observações horárias obtidas ao longo do mês. A série teve início em janeiro de 1981 e final em abril de 2012, com um total de 376 observações (N). Podemos observar as oscilações periódicas ao longo do ano devido ao efeito das estações no ciclo dos ventos.

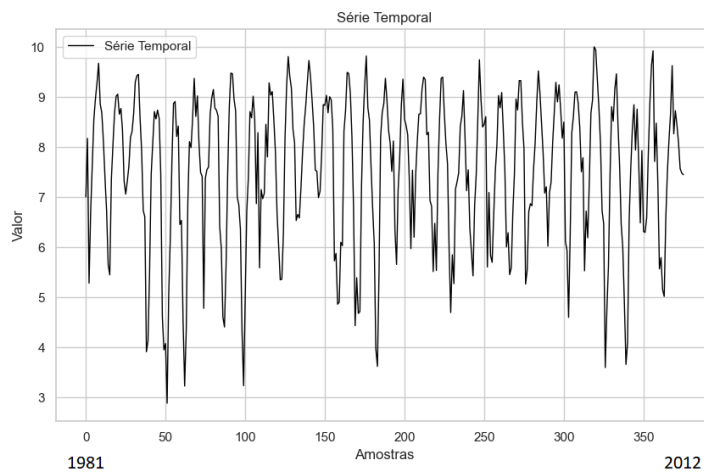


Figura 3.4: Série de ventos em São Paulo.

Fonte: (MORETTIN, 2014).

O correlograma em Fig. (3.5) indica que observações distanciadas de 12 meses e seus múltiplos (12, 24, 36) são fortemente positivamente correlacionadas, enquanto que observações distanciadas de 6 meses e seus múltiplos (18, 30) são negativamente correlacionadas. Este padrão evidencia a periodicidade de 12 meses dessa série, que será confirmada via análise de Fourier, apresentada a seguir (vide Eqs 3-6, 3-7 e 3-10).

Na Tabela (3.1) podemos observar que o harmônico de número 31 tem o maior módulo r_n , e assim este é o harmônico de maior influência nesse sinal.

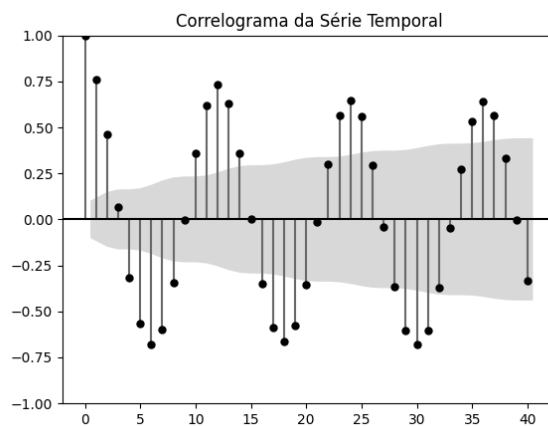


Figura 3.5: Correlograma da série de ventos em São Paulo.

n	a_n	b_n	r_n
1	15.071	0	15.071
3	0.160988	0.100341	0.189699
4	0.099084	0.244006	0.263356
19	-0.10131	-0.14337	0.175553
23	-0.15198	-0.13132	0.200855
31	-1.49912	-0.12856	1.504618

Tabela 3.1: Resultado para alguns coeficientes da FFT da série de Ventos.

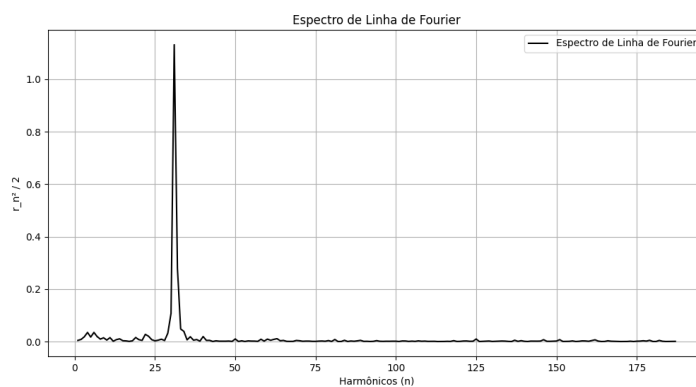


Figura 3.6: Espectro de linha de Fourier da série de ventos.

Utilizando os valores de a_1 , a_{31} e b_{31} , um modelo conveniente para representar o sinal então seria

$$y_t = 15.071 - 1.49912 \cos\left(\frac{\pi t}{6}\right) - 0.12856 \sin\left(\frac{\pi t}{6}\right) + \varepsilon_t \quad (3-20)$$

Como esse é o harmônico de maior contribuição para a “energia” da série, ele irá definir a periodicidade da série, que será obtida por:

$$T = \frac{2\pi}{\lambda_n},$$

onde T representa o período em unidades de tempo. Como $\lambda_n = \frac{2\pi n}{N}$, então segue que:

$$T = \frac{N}{n},$$

e assim sendo, temos que

$$T_{\text{ventos}} = \frac{376}{31} = 12.13 \approx 12.$$

A ideia agora é obter os coeficientes da série de Fourier listados na Tabela (3.1) através do método estatístico de MQO. Ao invés de usarmos diretamente as equações (3-6) e (3-7) para obter esses coeficientes via cálculo de Fourier, eles serão obtidos através do problema de otimização da equação (3-15), onde escolhemos os coeficientes que minimizam o erro:

$$\hat{\beta} = \min S(\hat{\beta})$$

Como as funções $\cos(\lambda_j t)$ e $\sin(\lambda_j t)$ são conhecidas para todo λ_j e t , o problema de minimização em (3-15) é facilmente resolvível através de uma solução analítica. Ademais, a partir de testes estatísticos de significância, podemos eliminar os coeficientes que contribuem de forma irrelevante para o sinal.

Formalmente, os testes de significância são estabelecidos através das seguintes hipóteses:

$$H_0 : \beta_j = 0, \quad (\text{coeficiente estatisticamente insignificante})$$

$$H_1 : \beta_j \neq 0 \quad j = 0, 1, \dots, \frac{N}{2} \quad (\text{coeficiente estatisticamente significativo})$$

onde $\beta_j = a_j$ ou b_j conforme a expansão de Fourier na equação (3-8).

Observe que, ao utilizar o método de MQO, teremos $\frac{N}{2} = 188$ coeficientes estimados. Assim como foi feito com os coeficientes obtidos via FFT (ver tabela 3.1), por uma questão de espaço, reportaremos apenas aqueles estatisticamente significantes que satisfaçam $|t_j| > 3$ ($|t_j| > 2$ já garante significância ao nível de 5% ou menos, mas queremos evitar excesso de parâmetros escolhidos).

Sendo que $t_j = \left| \frac{\hat{\beta}_j}{s_t(\hat{\beta}_j)} \right|$, se $|t_j| > 3$, então o coeficiente a_j ou b_j é dito estatisticamente significativo e deve ser mantido na série de Fourier. Caso contrário,

o coeficiente a_j ou b_j deve ser descartado da soma de Fourier.

Termo	Coeficiente	Erro Padrão	Estatística t
b_3	0.100341	0.047804	2.099
a_3	0.160988	0.047804	3.370
b_4	0.244006	0.047804	5.104
a_4	0.099084	0.047804	2.072
b_{19}	-0.14337	0.047804	-3.000
a_{19}	-0.10131	0.047804	-2.119
b_{23}	-0.13132	0.047804	-2.747
a_{23}	-0.15198	0.047804	-3.179
b_{31}	-0.12856	0.047804	-2.690
a_{31}	-1.49912	0.047804	-31.371

Tabela 3.2: Resultado da regressão OLS na série de ventos.

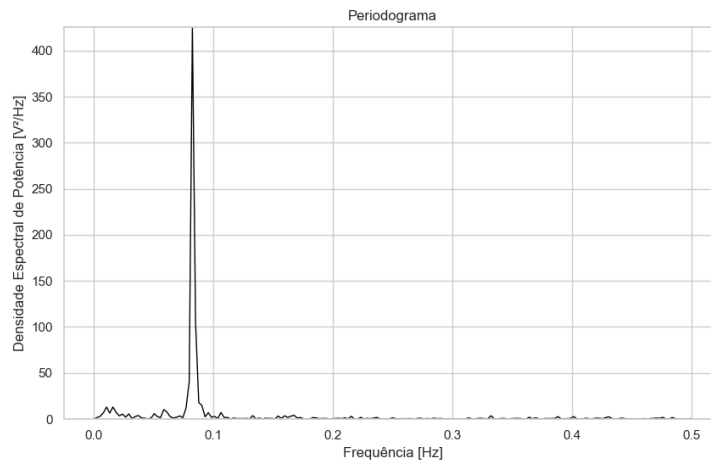


Figura 3.7: Periodograma da série de ventos.

Observamos que na Fig. (3.7) o periodograma da série apresenta um pico claro na frequência ω próxima à 0.082 Hz que equivale a um período de $T = \frac{1}{\omega} = \frac{1}{0.082} \approx 12$ meses caracterizando a periodicidade anual da série cujo harmônico 31 descreve a maioria do seu comportamento.

3.2

Análise de Ondaletas

Neste subcapítulo, iremos analisar a mesma série temporal de ventos apresentada no subcapítulo 3.1, utilizando técnicas baseadas na Análise de Ondaletas (AO) para decompor e filtrar o sinal.

3.2.1

Introdução à Análise de Ondaletas

A **Análise de Ondaletas** (AO) é uma técnica matemática poderosa que, assim como a **Análise de Fourier** (AF), permite decompor um sinal em suas componentes elementares. No entanto, a grande vantagem da AO está na sua capacidade de realizar uma análise simultânea no domínio do tempo e no domínio da frequência, oferecendo uma representação local de eventos transientes e padrões de curta duração. Essa característica torna a AO especialmente útil para o estudo de sinais não estacionários, como aqueles que apresentam variações abruptas, transientes ou comportamentos frequenciais que mudam ao longo do tempo.

“Um sinal é estacionário se suas propriedades forem estatisticamente invariantes ao longo do tempo. A ferramenta ideal para estudar sinais estacionários é a transformada de Fourier. Em outras palavras, sinais estacionários se decompõem canonicamente em combinações lineares de ondas (senos e cossenos). Da mesma forma, sinais que não são estacionários se decompõem em combinações lineares de ondaletas.”

Meyer, 1993 [?]

Enquanto a AF é amplamente utilizada para sinais estacionários, ela apresenta limitações significativas na identificação de eventos localizados no tempo, pois depende de componentes senoidais e cossenoidais globais. A AO, por outro lado, utiliza funções base denominadas *ondaletas*, que são localizadas no tempo e permitem representar com precisão tanto as frequências quanto o momento exato em que ocorrem alterações no sinal.

Outra vantagem importante da AO está na sua capacidade de filtrar ou remover ruído do sinal. Graças à sua decomposição em diferentes níveis de frequência, ela permite separar componentes indesejadas, como ruídos de alta frequência, preservando as informações relevantes do sinal original.

Ademais, a AO oferece flexibilidade ao permitir o uso de diferentes *ondaletas-mãe*, como a *ondaleta de Haar*, *Daubechies* ou *Morlet*, cada uma projetada para capturar padrões específicos de sinais. Isso possibilita a adaptação da análise às características do problema em questão, garantindo resultados mais robustos e interpretáveis.

Por fim, a AO não apenas complementa a Análise de Fourier, mas também supera suas limitações em cenários onde a temporalidade do sinal desempenha um papel crucial. Dessa forma, ela se destaca como uma ferramenta indispensável em diversas áreas de pesquisa e aplicações industriais.

3.2.2

Revisão teórica

3.2.3

Funções Ondaleta Mãe e Ondaleta Pai

No centro da AO está a **ondaleta mãe**, uma função de base que serve para decompor o sinal. Ao contrário dos senos e cossenos usados na Análise de Fourier, que são ilimitados no tempo, a ondaleta mãe é uma função localizada no tempo, o que significa que tem duração finita. Essa localização no tempo permite capturar eventos transientes e abruptos em sinais.

Matematicamente, uma ondaleta mãe $\psi(t)$ pode ser escalada e transladada para adaptar-se ao sinal que está sendo analisado. A função de ondaleta escalada e transladada, também conhecida como **ondaleta filha**, é dada por:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (3-21)$$

Onde:

- a é o **fator de escala**, que controla a frequência da ondaleta. Para $a > 1$, a ondaleta se expande, analisando componentes de baixa frequência; para $a < 1$, a ondaleta se comprime, analisando componentes de alta frequência. Na prática $a = 2^j$, $j \in \mathbb{Z}$.
- b é o **fator de translação**, que controla a localização temporal da ondaleta. Conforme b varia, a ondaleta desliza ao longo do tempo, permitindo que diferentes partes do sinal sejam analisadas. Na prática $b = k \cdot 2^{-j}$, $k \in \mathbb{Z}$.
- t é o tempo, $t \in \mathbb{R}$.

A função $\psi(t)$ para ser considerada ondaleta deve satisfazer as seguintes propriedades:

1. Admissibilidade

$$\int_{-\infty}^{\infty} \psi(t) dt = 0$$

2. Energia não infinita

$$\int_{-\infty}^{\infty} |\psi(t)| dt < \infty$$

- 3.

$$c_\psi = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty$$

onde $\Psi(\omega)$ é a transformada de Fourier de $\psi(t)$. Para que essa condição seja satisfeita, é necessário que $\Psi(0) = 0$, o que é equivalente à condição (1).

4.

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1$$

ou

$$\int_{-\infty}^{\infty} |\Psi(\omega)|^2 d\omega = 2\pi$$

5. Os primeiros $r - 1$ momentos de $\psi(t)$ se anulam, isto é:

$$\int_{-\infty}^{\infty} t^j \psi(t) dt = 0, \quad j = 0, 1, \dots, r - 1$$

para algum $r \geq 1$ e

$$\int_{-\infty}^{\infty} |t^r \psi(t)| dt < \infty$$

As funções $\psi_{j,k}(t)$ não necessariamente formam uma base ortogonal embora a ortogonalidade facilite o uso prático de ondaletas. Essa transformação possibilita a análise multiescala de um sinal, onde cada escala revela detalhes de diferentes frequências e localizações temporais. A grande vantagem da AO é essa flexibilidade de analisar o sinal em diferentes escalas e tempos, o que torna possível capturar tanto tendências de baixa frequência quanto detalhes de alta frequência.

Além da ondaleta mãe, a **ondaleta pai** $\phi(t)$ desempenha um papel fundamental na decomposição do sinal.

A ondaleta pai, também conhecida como *função de escala*, é usada para representar as componentes de baixa frequência e as tendências globais do sinal, capturando a sua aproximação.

$$\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k), \quad (3-22)$$

Para formar a ondaleta pai $\phi(t)$, é comum utilizar uma série de convoluções com um filtro de passa-baixa (chamado aqui de h_k), o qual permite a construção de aproximações sucessivas de baixa frequência. Esse filtro é aplicado iterativamente a uma função inicial de alta resolução para suavizar o sinal e reter as componentes de baixa frequência. Queremos que, ao final, a ondaleta pai tenha uma média não nula para capturar a tendência global do sinal.

A função de escala $\phi(t)$ deve satisfazer as seguintes propriedades matemáticas:

1. **Normalização:**

$$\int_{-\infty}^{\infty} \phi(t) dt = 1, \quad (3-23)$$

garantindo que a energia total capturada por $\phi(t)$ seja igual a 1.

2. **Ortogonalidade:**

$$\int_{-\infty}^{\infty} \phi(t) \phi(t - k) dt = \delta_k, \quad (3-24)$$

onde δ_k é o delta de Kronecker, que assegura a separação entre as escalas.

3. Admissibilidade:

$$\int_{-\infty}^{\infty} |\hat{\phi}(\omega)|^2 d\omega < \infty, \quad (3-25)$$

garantindo que $\phi(t)$ possua uma transformação de Fourier finita.

Essas propriedades permitem que $\phi(t)$ atue como uma base adequada para a decomposição multirresolução e a reconstrução perfeita do sinal.

De maneira prática, a função de escala é resolvida através de uma relação recursiva de convolução, onde o filtro h_k é aplicado repetidamente para “suavizar” uma função base até que se obtenha uma forma estável de baixa frequência. A cada aplicação do filtro, a função resultante representa uma versão de menor resolução do sinal original, e os coeficientes h_k controlam essa redução da resolução.

A solução analítica exata para $\phi(t)$ geralmente não será possível de se obter, especialmente quando não há uma fórmula explícita para a função de escala. Em muitos casos, é usada uma abordagem numérica para calcular a função de escala a partir da ondaleta mãe.

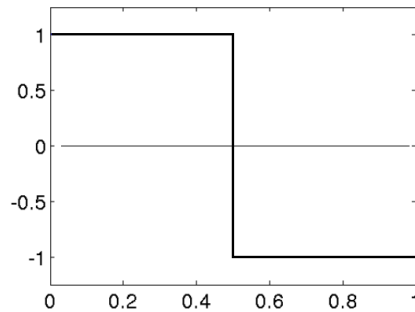
Esse processo iterativo e de verificação garante que as duas funções possam de fato se complementar para decompor um sinal, compondo a base da transformada discreta de ondaleta.

3.2.4

Ondaleta de Haar

A **ondaleta de Haar** é uma das ondaletas mais simples e é frequentemente usada para introduzir a teoria de ondaletas devido à sua simplicidade e propriedades discretas. A ondaleta de Haar é definida como:

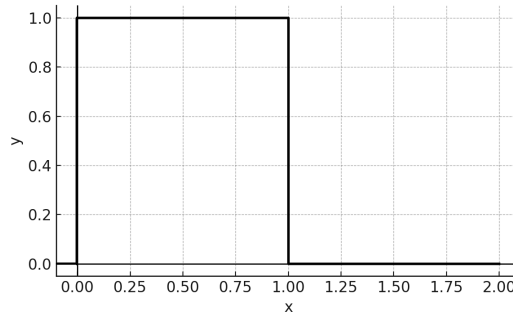
$$\psi(t) = \begin{cases} 1, & \text{se } 0 \leq t < \frac{1}{2} \\ -1, & \text{se } \frac{1}{2} \leq t < 1 \\ 0, & \text{caso contrario} \end{cases}$$



Fonte: (Autor).

Figura 3.8: Ondaleta mãe de Haar.

Ela é útil para detectar mudanças bruscas em sinais, sendo amplamente empregada em compressão de imagem e em análises de sinais discretos.



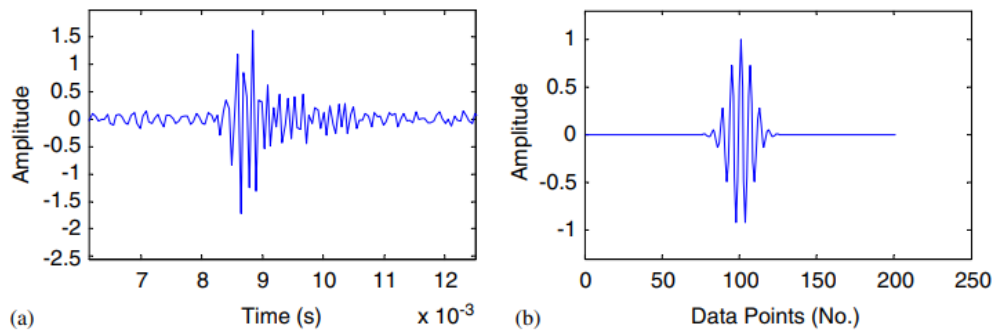
Fonte: (Autor).

Figura 3.9: Ondaleta pai de Haar.

3.2.5

Ondaleta de Morlet

Entre as diversas ondaletas-mães que podem ser utilizadas, a **ondaleta de Morlet** é amplamente utilizada para análise de sinais mecânicos e vibrações, devido à sua capacidade de detectar eventos impulsivos e periódicos em sinais de alta frequência.



Fonte: (H. Qiu al. / Journal of Sound and Vibration 289 (2006) 1066-1090).

Figura 3.10: Comparação entre um impulso mecânico (a) e a ondaleta de Morlet (b).

A função de ondaleta de Morlet combina uma senóide com uma envoltória Gaussiana, sendo definida como:

$$\psi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \cos(5t), \quad t \in \mathbb{R}. \quad (3-26)$$

Essa ondaleta tem uma boa localização tanto no tempo quanto na frequência, o que a torna ideal para a detecção de eventos transientes. O fator 5 na função cosseno define a frequência central da ondaleta, enquanto o termo gaussiano $e^{-t^2/2}$ garante que a função seja localizada no tempo.

A ondaleta de Morlet é escalada e transladada da mesma maneira que outras ondaletas-mães, permitindo que se ajuste a diferentes partes do sinal e diferentes escalas de frequência.

3.2.6

Decomposição e Reconstrução de Ondaletas

Considere a aproximação de uma série temporal $y(t)$ discreta em t , $t \in \mathbb{N}$, pela seguinte combinação linear de ondaletas, conhecida como **decomposição de ondaletas** (ou **expansão de multirresolução**):

$$y(t) = \sum_{k=1}^n c_{j_0,k} \phi_{j_0,k}(t) + \sum_{j=j_0}^{J-1} \sum_{k=1}^n d_{j,k} \psi_{j,k}(t), \quad (3-27)$$

onde $\phi_{j,k}(t)$ e $\psi_{j,k}(t)$ são as ondaletas **filho** e **filha**, respectivamente, e são obtidas a partir de dilatações e translações, como mostrado a seguir:

$$\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k), \quad (3-28)$$

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \quad (3-29)$$

com $j, k \in \mathbb{Z}$. Essa decomposição forma a base para muitas aplicações práticas de ondaletas, como a redução de ruído (*denoise*) e a extração de componentes (tendências e sazonalidades), que serão abordadas oportunamente.

A ondaleta mãe $\psi(t)$ fornece os **detalhes** do sinal, associados às frequências mais altas, enquanto a ondaleta pai $\phi(t)$ captura as frequências mais baixas, representando as **aproximações** do sinal (smoothing).

Os coeficientes $c_{j_0,k}$ são os **coeficientes de aproximação** na escala j_0 e localização k , enquanto os coeficientes $d_{j,k}$ são conhecidos como **coeficientes de detalhe** na escala j e localização k . Aqui, J indica o nível superior da decomposição. A forma de obter esses coeficientes será vista a seguir.

3.2.7

Transformada Discreta de Ondaletas (DWT)

A implementação da decomposição de ondaletas para um sinal dado $y(t)$, conforme mostrado na Eq.(3-27), requer a escolha dos parâmetros n e J , que dependem da aplicação. Por sua vez, os coeficientes $c_{j_0,k}$ e $d_{j,k}$ são calculados pelas seguintes relações de convolução entre a ondaleta escolhida e o sinal de interesse, assumido observado em tempo discreto.

$$c_{j_0,k} = \sum_n y(t) \phi_{j_0,k}(t), \quad d_{j,k} = \sum_n y(t) \psi_{j,k}(t), \quad (3-30)$$

Essas equações definem a **Transformada Discreta de Ondaletas** (DWT), que é implementada computacionalmente utilizando os coeficientes de filtros passa-baixa (h_k) e passa-alta (g_k), associados à ondaleta escolhida.

Pode-se mostrar que a DWT também pode ser representada pelas seguintes equações, envolvendo os coeficientes dos filtros passa-alto (h_k) e passa-baixo (g_k), associados às funções de escala $\phi(t)$ e das ondaletas $\psi(t)$. Formalmente, temos:

$$c_{j,k} = \sum_t y(t) h_k(2k - t), \quad d_{j,k} = \sum_t y(t) g_k(2k - t) \quad (3-31)$$

Se os filtros forem ortonormais, então pode-se mostrar, adicionalmente, que a partir desses coeficientes obtém-se uma reconstrução perfeita do sinal $y(t)$. Na prática a escolha dos níveis inferior (j_0) e superior (J) de resolução dependerá da aplicação. Por exemplo, em análises de redução de ruído (*denoise*), tendem-se a usar níveis mais baixos para preservar a estrutura global do sinal, enquanto em extração de componentes específicas são usados níveis mais altos para isolar variações específicas.

3.2.8

Cálculo dos Filtros h_k e g_k

Os filtros passa-baixa (h_k) e passa-alta (g_k) utilizados na DWT são derivados diretamente das propriedades da **ondaleta mãe** ($\psi(t)$) e da **função de escala** ($\phi(t)$). Esses filtros permitem que o sinal seja decomposto em suas componentes de baixa e alta frequência, respectivamente.

Pode-se mostrar que os coeficientes dos filtros passa-baixa (h_k) e passa-alta (g_k) podem ser obtidos a partir das seguintes integrais, utilizando as funções de escala $\phi(t)$ e a ondaleta-mãe $\psi(t)$:

$$h_k = \sqrt{2} \int \phi(t) \phi(2t - k) dt, \quad (3-32)$$

$$g_k = \sqrt{2} \int \psi(t) \phi(2t - k) dt. \quad (3-33)$$

A equação (3-32) descreve como a função escala $\phi(t)$ interage consigo mesma em diferentes escalas e translações, resultando nos coeficientes h_k , que formam o filtro passa-baixa. Por sua vez, a equação (3-33) descreve a interação entre a ondaleta-mãe $\psi(t)$ e a função de escala $\phi(t)$, fornecendo os coeficientes g_k , que formam o filtro passa-alta.

Os coeficientes h_k são calculados a partir da **equação de refinamento** da função de escala $\phi(t)$, definida como:

$$\phi(t) = \sum_k h_k \phi(2t - k), \quad (3-34)$$

onde $\phi(t)$ captura as componentes de baixa frequência do sinal e os coeficientes h_k correspondem aos pesos que descrevem como $\phi(t)$ se refina em diferentes escalas.

Os coeficientes do filtro passa-alta g_k são obtidos a partir de uma relação com h_k chamada *quadrature mirror filter relation*, sendo definidos como:

$$g_k = (-1)^k h_{1-k}, \quad (3-35)$$

onde a combinação de h_k e g_k garante que o filtro passa-alta extraia as componentes de alta frequência do sinal.

Os coeficientes h_k e g_k são obtidos a partir da análise da função de escala e da ondaleta-mãe escolhidas. Essas relações definem os filtros que permitem a separação do sinal em componentes de alta e baixa frequência, formando a base para a DWT.

Com essa abordagem, a DWT oferece uma forma eficiente e flexível de decompor sinais complexos em suas componentes essenciais, proporcionando tanto análises detalhadas quanto tendências globais.

3.2.9

Exemplo Numérico da DWT para Ondaleta Haar

Os coeficientes dos filtros passa-baixa (h_k) e passa-alta (g_k) da ondaleta Haar podem ser calculados a partir das equações 3-32 e 3-33.

Os filtros são calculados considerando os suportes das funções $\phi(t)$ e $\psi(t)$, e avaliando as integrais em seus intervalos de definição.

Cálculo de h_k Para $k = 0$:

$$h_0 = \sqrt{2} \int_{-\infty}^{\infty} \phi(t) \phi(2t) dt.$$

Sabendo que $\phi(t)$ tem suporte no intervalo $[0, 1]$, e $\phi(2t)$ tem suporte no intervalo $[0, \frac{1}{2}]$, o intervalo de integração será $[0, \frac{1}{2}]$. Assim:

$$h_0 = \sqrt{2} \int_0^{\frac{1}{2}} 1 \cdot 1 dt = \sqrt{2} \cdot \frac{1}{2} = \frac{1}{\sqrt{2}}.$$

Para $k = 1$:

$$h_1 = \sqrt{2} \int_{-\infty}^{\infty} \phi(t) \phi(2t - 1) dt.$$

Aqui, $\phi(2t - 1)$ tem suporte no intervalo $[\frac{1}{2}, 1]$, e a interseção com $\phi(t)$ ocorre no mesmo intervalo $[\frac{1}{2}, 1]$. Assim:

$$h_1 = \sqrt{2} \int_{\frac{1}{2}}^1 1 \cdot 1 dt = \sqrt{2} \cdot \frac{1}{2} = \frac{1}{\sqrt{2}}.$$

Cálculo de g_k Para $k = 0$:

$$g_0 = \sqrt{2} \int_{-\infty}^{\infty} \psi(t) \phi(2t) dt.$$

O suporte de $\phi(2t)$ é $[0, \frac{1}{2}]$, e o suporte de $\psi(t)$ é $[0, 1]$. No intervalo $[0, \frac{1}{2}]$, $\psi(t) = 1$, então:

$$g_0 = \sqrt{2} \int_0^{\frac{1}{2}} 1 \cdot 1 dt = \sqrt{2} \cdot \frac{1}{2} = \frac{1}{\sqrt{2}}.$$

Para $k = 1$:

$$g_1 = \sqrt{2} \int_{-\infty}^{\infty} \psi(t)\phi(2t-1) dt.$$

O suporte de $\phi(2t-1)$ é $[\frac{1}{2}, 1]$. Nesse intervalo, $\psi(t) = -1$. Assim:

$$g_1 = \sqrt{2} \int_{\frac{1}{2}}^1 (-1) \cdot 1 dt = -\sqrt{2} \cdot \frac{1}{2} = -\frac{1}{\sqrt{2}}.$$

Os coeficientes dos filtros para a ondaleta Haar são:

$$h_0 = \frac{1}{\sqrt{2}}, \quad h_1 = \frac{1}{\sqrt{2}}, \quad g_0 = \frac{1}{\sqrt{2}}, \quad g_1 = -\frac{1}{\sqrt{2}}.$$

Para obter esses coeficientes iremos utilizar a seguinte série temporal discreta $y = [4, 6, 10, 12, 14, 18, 20, 22]$ com $N = 8$. Vamos realizar os cálculos da DWT para calcular os coeficientes de **aproximação** ($c_{j,k}$) e **detalhe** ($d_{j,k}$) usando os filtros h_k e g_k da ondaleta Haar:

Os coeficientes de aproximação $c_{j,k}$ e detalhe $d_{j,k}$ são obtidos pela aplicação dos filtros a pares consecutivos da série:

$$c_{j,k} = \sum_{n=0}^1 h_n \cdot y_{2k+n}, \quad d_{j,k} = \sum_{n=0}^1 g_n \cdot y_{2k+n}.$$

Para $k = 0, 1, 2, 3$.

Cálculo passo a passo

– Para $k = 0$:

$$c_{0,0} = \frac{1}{\sqrt{2}} \cdot 4 + \frac{1}{\sqrt{2}} \cdot 6 = \frac{4+6}{\sqrt{2}} = \frac{10}{\sqrt{2}} \approx 7.07,$$

$$d_{0,0} = \frac{1}{\sqrt{2}} \cdot 4 - \frac{1}{\sqrt{2}} \cdot 6 = \frac{4-6}{\sqrt{2}} = \frac{-2}{\sqrt{2}} \approx -1.41.$$

– Para $k = 1$:

$$c_{0,1} = \frac{1}{\sqrt{2}} \cdot 10 + \frac{1}{\sqrt{2}} \cdot 12 = \frac{10+12}{\sqrt{2}} = \frac{22}{\sqrt{2}} \approx 15.56,$$

$$d_{0,1} = \frac{1}{\sqrt{2}} \cdot 10 - \frac{1}{\sqrt{2}} \cdot 12 = \frac{10-12}{\sqrt{2}} = \frac{-2}{\sqrt{2}} \approx -1.41.$$

– Para $k = 2$:

$$c_{0,2} = \frac{1}{\sqrt{2}} \cdot 14 + \frac{1}{\sqrt{2}} \cdot 18 = \frac{14 + 18}{\sqrt{2}} = \frac{32}{\sqrt{2}} \approx 22.63,$$

$$d_{0,2} = \frac{1}{\sqrt{2}} \cdot 14 - \frac{1}{\sqrt{2}} \cdot 18 = \frac{14 - 18}{\sqrt{2}} = \frac{-4}{\sqrt{2}} \approx -2.83.$$

– Para $k = 3$:

$$c_{0,3} = \frac{1}{\sqrt{2}} \cdot 20 + \frac{1}{\sqrt{2}} \cdot 22 = \frac{20 + 22}{\sqrt{2}} = \frac{42}{\sqrt{2}} \approx 29.70,$$

$$d_{0,3} = \frac{1}{\sqrt{2}} \cdot 20 - \frac{1}{\sqrt{2}} \cdot 22 = \frac{20 - 22}{\sqrt{2}} = \frac{-2}{\sqrt{2}} \approx -1.41.$$

Os coeficientes calculados são:

$$c_{j,k} = [7.07, 15.56, 22.63, 29.70], \quad d_{j,k} = [-1.41, -1.41, -2.83, -1.41].$$

Através dos filtros h_k e g_k da ondaleta Haar, obtivemos os coeficientes de aproximação $c_{j,k}$, que capturam as componentes de baixa frequência, e os coeficientes de detalhe $d_{j,k}$, que representam as variações rápidas no sinal. Este processo ilustra a primeira etapa da decomposição multi-escala por DWT.

3.2.10

Transformada Contínua de Ondaletas (CWT)

A **Transformada Contínua de Ondaletas** (CWT, do inglês *Continuous Wavelet Transform*), assim como a DWT, permite a decomposição de um sinal $y(t)$ em diferentes escalas e localizações temporais utilizando uma **ondaleta-mãe** $\psi(t)$. Ao contrário da DWT, a CWT não requer a discretização rígida dos parâmetros de escala e translação, oferecendo uma representação contínua do sinal no plano tempo-frequência. A CWT é definida matematicamente como:

$$W(a, b) = \int_{-\infty}^{\infty} y(t) \psi_{a,b}^*(t) dt, \quad (3-36)$$

onde $\psi_{a,b}(t)$ é a ondaleta-mãe escalada e transladada, dada por:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), \quad (3-37)$$

sendo a o parâmetro de escala (associado à frequência) e b o parâmetro de translação (associado ao tempo). O símbolo $*$ denota o conjugado complexo, necessário quando a ondaleta é uma função complexa.

3.2.11

Interpretação dos Parâmetros de Escala

O parâmetro de escala a controla a dilatação ou compressão da ondaleta-mãe. Escalas pequenas ($a \ll 1$) correspondem a ondaletas mais comprimidas, sensíveis a altas frequências, enquanto escalas grandes ($a \gg 1$) correspondem a ondaletas dilatadas, associadas a frequências mais baixas.

Relação entre Escala e Frequência O eixo das escalas (a) no escalograma está relacionado de forma inversa à frequência física do sinal analisado. Escalas menores ($a \ll 1$) correspondem a ondaletas comprimidas, que são sensíveis a altas frequências, enquanto escalas maiores ($a \gg 1$) correspondem a ondaletas dilatadas, associadas a frequências mais baixas. A conversão entre escala e frequência é fundamental para interpretar o escalograma em termos físicos e pode ser feita por meio da relação:

$$f = \frac{\text{scale2frequency}(\psi, a)}{\Delta t}, \quad (3-38)$$

onde Δt é o intervalo de amostragem do sinal e ψ é a ondaleta utilizada. Essa conversão é implementada na biblioteca PyWavelets pela função `scale2frequency`, que permite mapear diretamente escalas para frequências, facilitando a análise.

3.2.12

Representação Visual: Escalograma

Os coeficientes $W(a, b)$ da CWT podem ser visualizados como um **escalograma**, um gráfico bidimensional que apresenta a intensidade dos coeficientes em função do tempo (b , no eixo horizontal) e da escala (a , no eixo vertical). O escalograma fornece uma visão intuitiva sobre como as frequências de um sinal evoluem ao longo do tempo, facilitando a detecção de fenômenos localizados e padrões periódicos. A intensidade no escalograma, indicado pelo mapa de cores, indica a magnitude do coeficiente de ondaleta naquela frequência e tempo.

Como Interpretar o Escalograma A interpretação do escalograma requer atenção aos seguintes aspectos:

- **Mapa de cor:** indica a magnitude dos coeficientes da ondaleta, sendo representada por um gradiente cromático no escalograma: as cores mais quentes (vermelho) indicam mais energia naquela escala/tempo, enquanto que cores mais frias (azuis) representam menores energias. Quanto maior o conteúdo de energia mais dominante é a frequências/escala no sinal.

- **Eixo das escalas (vertical):** reflete as diferentes frequências presentes no sinal. Escalas altas com cores quentes indicam presença de estruturas de baixa frequência (como tendências ou ciclos longos), enquanto escalas baixas com cores quentes identificam componentes de alta frequência (como picos ou ruídos).
- **Eixo do tempo (horizontal):** indica a progressão da frequência no tempo. Se um determinado valor de escala apresenta cor quente, e esta evolução é uniforme no tempo, então essa frequência não muda no tempo. Se o mapa de cor mudar para cor fria em alguma faixa temporal, então nessa faixa a frequência atua de forma mais branda ou é inexistente.

3.2.13

Especificação de Parâmetros em Ondaletas Contínuas

Ao utilizar ondaletas contínuas para gerar escalogramas, é essencial especificar os parâmetros que controlam suas propriedades, como largura de banda e frequência central. Esses parâmetros influenciam diretamente a resolução temporal e espectral, afetando a precisão da representação no escalograma.

- **Largura de banda:** Controla a concentração da ondaleta no domínio da frequência. Maior largura de banda melhora a resolução espectral, enquanto uma largura menor prioriza a localização no tempo.
- **Frequência central:** Define a região espectral de interesse. Um ajuste adequado foca a análise nas frequências relevantes do sinal.

A escolha inadequada desses parâmetros pode causar artefatos ou dificultar a identificação de padrões e eventos transientes no sinal. Por isso, recomenda-se ajustá-los de forma coerente com as características do sinal e os objetivos da investigação.

A capacidade do escalograma de representar simultaneamente o comportamento do sinal no tempo e na frequência é uma de suas principais vantagens, especialmente em comparação com técnicas lineares como a Transformada de Fourier. Ele é amplamente utilizado para identificar padrões de frequência variáveis, como pulsos em sinais biomédicos, picos em vibrações mecânicas ou padrões climáticos.

Na prática, a CWT é implementada computacionalmente utilizando funções de ondaleta pré-definidas, como a Morlet, Mexican Hat e Gaussian Derivative. Essas ondaletas são selecionadas de acordo com a natureza do sinal em análise, considerando sua resolução no tempo e na frequência.

A título de ilustração, apresentamos nas figuras (3.11) e (3.12) os escalogramas para duas séries temporais geradas artificialmente. A primeira série foi gerada a partir do seguinte modelo:

$$y_1(t) = \sin(2\pi 10t) + \epsilon_t, \quad t = 0, \dots, 2$$

$$\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2), \quad \sigma_\epsilon^2 = 0.2$$

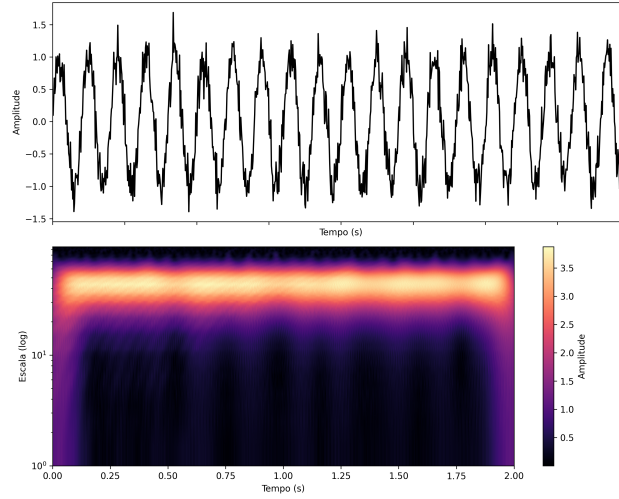


Figura 3.11: Escalograma ilustrativo de uma série com ciclo determinística.

Portanto trata-se de um um modelo com ciclo determinístico monocromático (ver Seção 3.3.1 eqs (3-43) e (3-44). Mais especificamente a frequência de 10Hz (ou período de $T=1/10$), observada no eixo da Escala, é constante ao longo do tempo e a amplitude varia deterministicamente, a menos da superposição do choque. O mapa de calor indica uma faixa nítida e uniforme em torno da frequência de 10 Hz, indicando que essa é a frequência dominante e ela permanece a mesma ao longo do tempo. A segunda série, por sua vez, foi gerada pela seguinte equação:

$$y_2(t) = (1 + 0.5 \sin(2\pi t)) \sin \left(2\pi \left(10t + 10t^2 \right) \right) + \epsilon_t, \quad t = 0, \dots, 2,$$

$$\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2), \quad \sigma_\epsilon^2 = 0.2$$

onde a frequência e a amplitude variam deterministicamente com o tempo segundo as equações $10(1+t)$ e $(1 + 0.5 \sin(2\pi t))$, respectivamente. Nesse caso, o escalograma evidencia a variação temporal tanto do período quanto da amplitude do sinal, representados pelas mudanças na distribuição das cores do mapa de calor e nas escalas ao longo do tempo.

Esses exemplos ilustram como o escalograma pode ser utilizado para distinguir séries com comportamento determinístico de séries mais complexas, nas quais a amplitude e a frequência variam, destacando a flexibilidade da análise por ondaletas na representação de diferentes dinâmicas.

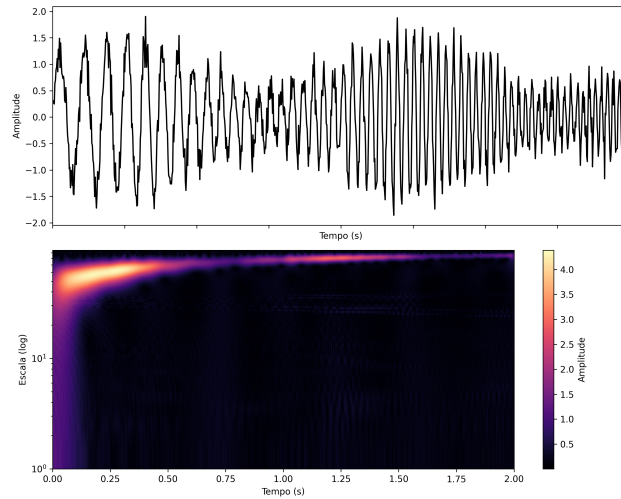


Figura 3.12: Escalograma ilustrativo de uma série com amplitude e frequência variáveis.

3.2.14

Reconstrução de Ondaletas

A **reconstrução de ondaletas**, também conhecida como **recomposição**, é o processo inverso da decomposição. Ela reconstitui o sinal original ou uma versão modificada a partir dos coeficientes $c_{j_0,k}$ e $d_{j,k}$ utilizando as funções base $\phi(t)$ e $\psi(t)$:

$$y(t) = \sum_{k=1}^n c_{j_0,k} \phi_{j_0,k}(t) + \sum_{j=j_0}^{J-1} \sum_{k=1}^n d_{j,k} \psi_{j,k}(t). \quad (3-39)$$

Durante a recomposição, as frequências altas (detalhes) e baixas (aproximações) são combinadas para restaurar o sinal. Os coeficientes de alta frequência ($d_{j,k}$) capturam variações rápidas (como ruídos), enquanto os coeficientes de baixa frequência ($c_{j_0,k}$) representam a tendência ou estruturas globais do sinal, que são mais suaves no tempo.

Entre a decomposição e a reconstrução, os coeficientes $c_{j_0,k}$ e $d_{j,k}$ podem ser modificados ou não, dependendo da aplicação. Por exemplo, podem ser aplicadas técnicas como **limiarização** (*thresholding*) para remover ruídos ou comprimir o sinal.

A reconstrução do sinal com esses coeficientes filtrados resulta em um sinal mais suave, preservando a tendência original e eliminando ruídos indesejados. A Tabela (3.3) a seguir apresenta as principais diferenças entre os procedimentos de Decomposição e Reconstrução de um sinal obtidos via DWT.

Essas técnicas permitem que a DWT não seja apenas uma ferramenta de decomposição, mas também um método versátil para modificação e aprimoramento de sinais antes de sua recomposição.

Aspecto	Decomposição	Reconstrução
Direção do Processo	Quebra o sinal em componentes de diferentes frequências	Combina os componentes para recriar o sinal
Filtros	Aplica filtros passa-baixa e passa-alta	Usa filtros inversos para recombinar as frequências
Subamostragem	Reduz a amostragem após a filtragem	Expande os coeficientes antes da filtragem
Objetivo	Analisar ou processar o sinal em diferentes escalas	Recuperar o sinal original ou uma versão simplificada
Resultado	Sequência de coeficientes de aproximação e detalhe	Um sinal contínuo (reconstruído)

Tabela 3.3: Comparação entre Decomposição e Reconstrução utilizando Ondaletas

3.2.15

Exemplo Numérico da Reconstrução a partir dos Coeficientes

Dado o sinal $y = (4, 6, 10, 12, 14, 18, 20, 22)$, que foi decomposto em coeficientes de aproximação ($c_{j,k}$) e detalhe ($d_{j,k}$), reconstruiremos o sinal utilizando a **Transformada Inversa de Ondaletas Discreta (IDWT)**.

Os coeficientes calculados anteriormente foram:

$$c_{j,k} = [7.07, 15.56, 22.63, 29.70], \quad d_{j,k} = [-1.41, -1.41, -2.83, -1.41].$$

A reconstrução do sinal baseia-se nas fórmulas para os valores originais:

$$y_{2k} = h_0 \cdot c_{j,k} + g_0 \cdot d_{j,k},$$

$$y_{2k+1} = h_1 \cdot c_{j,k} + g_1 \cdot d_{j,k}.$$

Para ondaleta Haar já sabemos que:

$$h_0 = \frac{1}{\sqrt{2}}, \quad h_1 = \frac{1}{\sqrt{2}}, \quad g_0 = \frac{1}{\sqrt{2}}, \quad g_1 = -\frac{1}{\sqrt{2}}.$$

Cálculo passo a passo

Par $k = 0$: Para $c_{j,0} = 7.07$ e $d_{j,0} = -1.41$:

$$y_0 = \frac{1}{\sqrt{2}} \cdot 7.07 + \frac{1}{\sqrt{2}} \cdot (-1.41) = 4,$$

$$y_1 = \frac{1}{\sqrt{2}} \cdot 7.07 - \frac{1}{\sqrt{2}} \cdot (-1.41) = 6.$$

Par $k = 1$: Para $c_{j,1} = 15.56$ e $d_{j,1} = -1.41$:

$$y_2 = \frac{1}{\sqrt{2}} \cdot 15.56 + \frac{1}{\sqrt{2}} \cdot (-1.41) = 10,$$

$$y_3 = \frac{1}{\sqrt{2}} \cdot 15.56 - \frac{1}{\sqrt{2}} \cdot (-1.41) = 12.$$

Par $k = 2$: Para $c_{j,2} = 22.63$ e $d_{j,2} = -2.83$:

$$y_4 = \frac{1}{\sqrt{2}} \cdot 22.63 + \frac{1}{\sqrt{2}} \cdot (-2.83) = 14,$$

$$y_5 = \frac{1}{\sqrt{2}} \cdot 22.63 - \frac{1}{\sqrt{2}} \cdot (-2.83) = 18.$$

Par $k = 3$: Para $c_{j,3} = 29.70$ e $d_{j,3} = -1.41$:

$$y_6 = \frac{1}{\sqrt{2}} \cdot 29.70 + \frac{1}{\sqrt{2}} \cdot (-1.41) = 20,$$

$$y_7 = \frac{1}{\sqrt{2}} \cdot 29.70 - \frac{1}{\sqrt{2}} \cdot (-1.41) = 22.$$

Portanto, o sinal reconstruído é:

$$y = [4, 6, 10, 12, 14, 18, 20, 22].$$

Isso confirma que a IDWT utilizando a ondaleta de Haar consegue reconstruir perfeitamente o sinal original.

3.2.16

Filtragem e Redução de Ruído com Ondaletas

A filtragem e redução de ruído com ondaletas consiste na decomposição do sinal $y(t)$ em componentes de diferentes escalas, permitindo eliminar ruídos indesejados preservando as frequências de interesse. Para isso, os coeficientes de detalhe $d_{j,k}$ são processados por um **limiar** λ , de forma que coeficientes abaixo do limiar sejam zerados, resultando na atenuação do ruído e na suavização do sinal:

$$d_{j,k} = \begin{cases} 0, & \text{se } |d_{j,k}| < \lambda \\ d_{j,k}, & \text{caso contrário} \end{cases} \quad (3-40)$$

O limiar λ pode ser determinado por critérios como o limiar universal de Donoho e Johnstone:

$$\lambda = \sigma \sqrt{2 \log N},$$

onde σ representa o desvio padrão dos coeficientes de detalhe e N o número total de amostras do sinal.

A filtragem e a remoção de ruído (*denoising*) são abordagens relacionadas, mas distintas: a filtragem pode eliminar tanto ruídos quanto pequenas variações do sinal, enquanto o *denoising* busca preservar suas características essenciais. A escolha entre essas técnicas depende do objetivo da análise.

A Figura (3.13) nos mostra um exemplo de uma decomposição por ondaletas em uma série genérica com ruído, denominada “Sinal Original” nessa Figura.

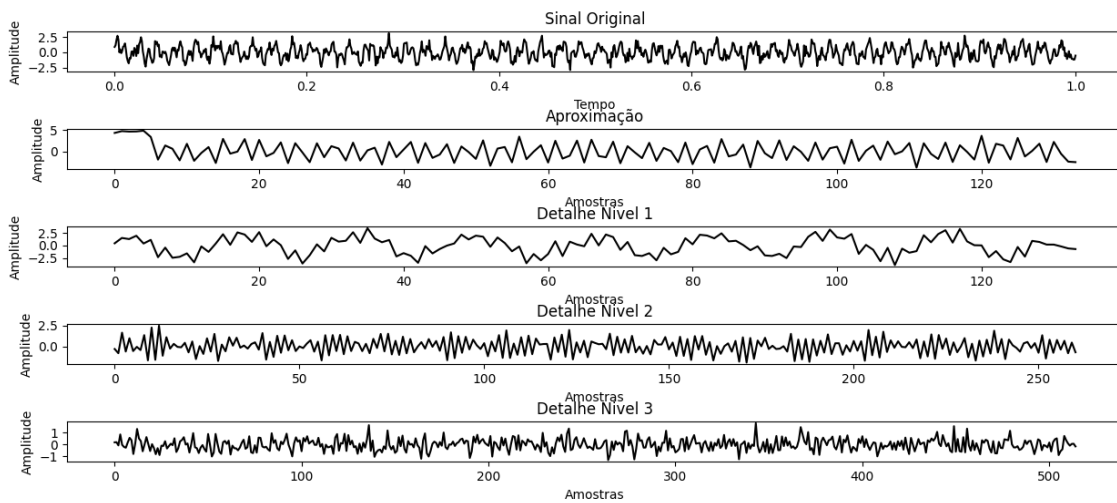


Figura 3.13: Exemplo de decomposição de uma série temporal ruidosa utilizando a ondaleta *Daubechies-4*. Cada nível de decomposição separa componentes do sinal em diferentes escalas, evidenciando tendências e detalhes.

3.2.17

Exemplo

Neste subcapítulo, apresentamos a aplicação da análise de ondaletas à mesma série temporal de ventos apresentada anteriormente na Fig (3.4), utilizando a Transformada Discreta de Ondaletas (DWT) com a ondaleta Daubechies-4 (*db4*). O objetivo principal foi reduzir o ruído da série original e evidenciar sua estrutura cíclica.

O programa desenvolvido em Python realizou a decomposição da série temporal em níveis hierárquicos de detalhe e aproximação. Durante o processamento, os coeficientes de detalhe de maior frequência foram zerados, preservando apenas as informações de menor frequência, relacionadas às oscilações de longo prazo. Em seguida, a série foi reconstruída a partir dos coeficientes ajustados.

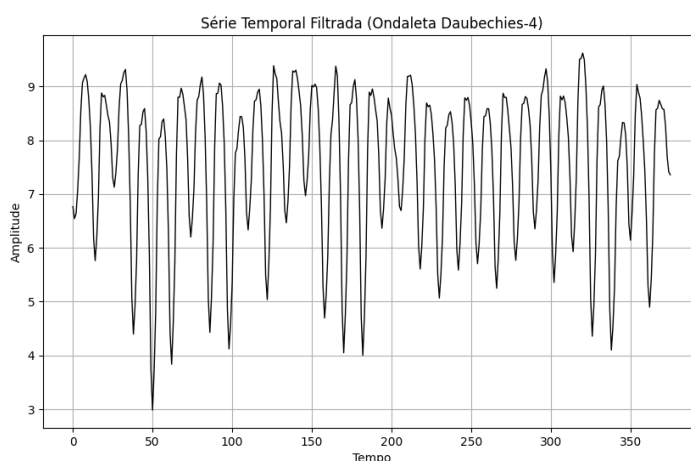


Figura 3.14: Série temporal de ventos após aplicação do filtro com ondaletas Daubechies-4 (*db4*). As oscilações cíclicas são mais evidentes após a remoção do ruído.

A Figura (3.14) apresenta o resultado da série filtrada, onde é possível observar com maior clareza os ciclos predominantes na dinâmica do vento. Esse procedimento é especialmente útil para identificar padrões cíclicos em séries temporais ruidosas, pois as flutuações de alta frequência, muitas vezes associadas a ruídos, podem mascarar as tendências e ciclos de interesse.

3.3

Modelos de Séries Temporais com Ciclos

A modelagem de ciclos em séries temporais desempenha um papel crucial na compreensão de fenômenos dinâmicos em diversas áreas da ciência. Ciclos representam padrões recorrentes em séries temporais, e sua modelagem pode variar entre abordagens puramente determinísticas e modelos estocásticos mais sofisticados. Este capítulo apresenta a teoria dos modelos de ciclo, desde a formulação determinística mais simples até a incorporação de componentes estocásticos, de acordo com o modelo estrutural de Harvey [10].

Nos modelos estruturais de Harvey para séries não sazonais [10], uma série temporal (ST) pode ser decomposta em **componentes não observáveis (CNO)** de interesse, tais como tendência, ciclo e irregular (erro), sendo representadas esquematicamente pela seguinte expressão:

$$ST = \text{Tendência} + \text{Ciclo} + \text{Irregular}. \quad (3-41)$$

Cada uma dessas componentes apresenta características específicas que são essenciais para a modelagem e compreensão de uma determinada série temporal.

3.3.1

Tendência

A **tendência** representa a componente da série temporal que **muda pouco ao longo do tempo**, exibindo suavidade em sua estimativa. Ela reflete a evolução de longo prazo da série e está associada a fenômenos de crescimento ou mudança gradual. Exemplos incluem:

- **Demografia**: crescimento populacional;
- **Economia**: mudanças no poder aquisitivo, expansão econômica (turismo, fábricas, fronteira agrícola, comércio);
- **Inflação**: variações de longo prazo no nível de preços;
- **Mudanças Climáticas**: alterações nos padrões climáticos ao longo de décadas.

3.3.2

Ciclo

A **componente cíclica** é caracterizada por **flutuações recorrentes com periodicidade superior a um ano**. Diferentemente da sazonalidade, os ciclos possuem duração mais longa e nem sempre apresentam uma periodicidade perfeitamente regular. Os ciclos estão geralmente associados a:

- **Atividade Econômica:** ciclos de crescimento e recessão econômica;
- **Ciclos Climáticos:** variações de longo prazo em chuvas e temperaturas;
- **Fenômenos Naturais:** ciclos solares, como a variação no número de manchas solares.

A modelagem dos ciclos é essencial para prever comportamentos de médio e longo prazo em fenômenos cíclicos.

3.3.3

Irregular (Erro)

A componente **irregular**, também chamada de **erro**, é composta por **flutuações aleatórias** que não seguem padrões determinísticos ou recorrentes. Assim como nos modelos econométricos, assume-se que os erros possuem as seguintes propriedades:

1. **Distribuição Normal:** $\epsilon_t \sim N(0, \sigma^2)$.
2. **Ausência de correlação:** os erros são descorrelacionados no tempo, $\rho(k) = 0, \quad \forall \quad k \geq 1$.
3. **Homoscedasticidade:** a variância dos erros é constante ao longo do tempo, $\text{Var}(\epsilon_t) = \sigma^2$.

A verificação dessas hipóteses é realizada através de diagnósticos do modelo, tais como análise dos resíduos, testes de autocorrelação e heterocedasticidade.

3.3.4

Importância da Decomposição

A compreensão e modelagem de cada uma dessas componentes permitem um maior entendimento dos fenômenos cíclicos e são essenciais na análise de séries temporais periódicas.

Na prática, a especificação de um modelo de ciclo para uma série temporal inclui também um modelo para tendência, além da componente de erro. Na especificação mais simples desse modelo, ambos a tendência e o ciclo podem ser considerados determinísticos. Na especificação mais complexa, que será adotada na nossa investigação da série de manchas solares, ambas as componentes de tendência e ciclo serão consideradas estocásticas. Essas diferentes abordagens serão consideradas a seguir.

3.3.5

Modelo com Tendência e Ciclo

O formato básico de um modelo de série temporal com tendência e ciclo pode ser representado como:

$$y_t = \mu_t + \psi_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \quad (3-42)$$

onde:

- y_t : representa o valor observado da série temporal no instante t ;
- μ_t : é a componente de tendência;
- ψ_t : é a componente de ciclo;
- ϵ_t : representa o **termo de erro**, um ruído aleatório normalmente distribuído com média zero e variância constante σ^2 .

Esse formato separa a série temporal em uma parte sistemática μ_t e ψ_t e uma parte não sistemática, ϵ_t . As componentes μ_t e ψ_t capturam a estrutura principal do comportamento da série, enquanto o termo ϵ_t representa flutuações aleatórias que não podem ser explicadas pelo modelo.

3.3.6

Modelos para Ciclo

3.3.7

a) Modelo de Ciclo Determinístico

Inicialmente, vamos considerar a versão determinística do modelo para ciclo, que servirá de base para o modelo estrutural de ciclo estocástico:

$$y_t = \psi_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \zeta^2), \quad (3-43)$$

$$\psi_t = \alpha \cos(\lambda_c t) + \beta \sin(\lambda_c t), \quad (3-44)$$

onde

$$\lambda_c = \frac{2\pi}{T}$$

é a frequência em radianos do ciclo.

Aqui, λ_c representa a frequência desconhecida do ciclo(em radianos) e T o período do ciclo medido na unidade de tempo da série (meses, anos etc). Os parâmetros α e β , associados a amplitude do ciclo, também são constantes desconhecidas. Todas essas quantidades podem ser estimadas via Análise de Fourier (AF), conforme demonstrado no Capítulo de Metodologia (Capítulo 3). Caso λ_c seja estimado previamente via AF, então pode-se demonstrar que α e β podem ser estimados via Mínimos Quadrados Ordinários (MQO).

Tipicamente, os modelos são ajustados no período de treinamento e projetados no período de teste, onde as previsões (projeções) podem ser comparadas com os valores observados. Se T é o último ano do período de treinamento, então a previsão para esse modelo será dada por:

$$\hat{y}_{T+k|T} = \hat{\psi}_{T+k|T} \quad (3-45)$$

onde

$$\hat{\psi}_{T+k|T} = \hat{\alpha} \cos(\lambda_c(T+k)) + \hat{\beta} \sin(\lambda_c(T+k)) \quad (3-46)$$

com $\hat{\alpha}$, $\hat{\beta}$ e λ_c estimados no período de treinamento, sendo

$$\lambda_c = \frac{2\pi}{T}.$$

3.3.8

b) Modelo de Ciclo Estocástico

De acordo com Harvey[10], o modelo estrutural de ciclo estocástico pode ser representado como:

$$y_t = \psi_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \quad (3-47)$$

$$\begin{pmatrix} \psi_t \\ \psi_t^* \end{pmatrix} = \rho \begin{pmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{pmatrix} \begin{pmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{pmatrix} + \begin{pmatrix} \kappa_t \\ \kappa_t^* \end{pmatrix} \quad (3-48)$$

$$\kappa_t \sim \kappa_t^* \sim N(0, \sigma_k^2),$$

donde segue que:

$$\psi_t = \rho \psi_{t-1} \cos \lambda_c + \rho \psi_{t-1}^* \sin \lambda_c + \kappa_t, \quad \kappa_t \sim N(0, \sigma_k^2), \quad (3-49)$$

onde ρ é o coeficiente de amortecimento ($0 < \rho < 1$), κ_t e κ_t^* são dois choques aleatórios (ruídos brancos descorrelatados) que tornam o ciclo estocástico, $\lambda_c = 2\pi/T$ é a frequência em radianos e T o período do ciclo. Nessa equação ψ_t é a componente de ciclo, enquanto ψ_t^* surge por construção na definição do ciclo estocástico, não possuindo interpretação física. A amplitude do ciclo está associada aos termos $\rho \psi_{t-1}$ e $\rho \psi_{t-1}^*$. Observe que se $\sigma_k^2 = 0$, então o ciclo estocástico torna-se determinístico, conforme eq (3-43). É importante notar que nesse modelo o ciclo é “conjuntamente” estocástico, no sentido de que não é possível separar a natureza estocástica do ciclo como um todo em termos das estocasticidades individuais do período T e da amplitude.

3.3.9

Modelos para Tendência

3.3.10

a) Tendência Estocástica Estacionária

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \quad (3-50)$$

$$\mu_t = c + \phi\mu_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \quad (3-51)$$

Onde c é o nível do modelo. O este modelo possui um comportamento estacionário quando $-1 < \phi < 1$.

3.3.11

b) Tendência Estocástica Não Estacionária

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \quad (3-52)$$

Fazendo $\phi = 1$ na equação (3-51), obtemos o modelo de passeio aleatório com drift:

$$\mu_{t+1} = c + \mu_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \quad (3-53)$$

Se $c = 0$, então obtem-se o modelo de passeio aleatório puro, também conhecido como o modelo de nível local:

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2), \quad (3-54)$$

Observe que se $\sigma_\eta^2 = 0$, então $\mu_t = \mu_{t-1}$, ou seja, a tendência torna-se constante, que coincide com o próximo modelo.

3.3.12

c) Modelo de Tendência Constante

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \quad (3-55)$$

$$\mu_t = \mu_{t-1} = \mu, \quad \forall t \quad (3-56)$$

3.3.13

Modelo com Ciclo e Tendência Estocásticas

Podemos combinar os dois casos de estrutura apresentados anteriormente com estocacidade tanto na tendência quanto no ciclo para construir um modelo estocástico mais complexo. Esse será o modelo a ser utilizado na série de manchas solares.

$$y_t = \mu_t + \psi_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \quad (3-57)$$

$$\begin{pmatrix} \psi_t \\ \psi_t^* \end{pmatrix} = \rho \begin{pmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{pmatrix} \begin{pmatrix} \psi_{t-1} \\ \psi_{t-1}^* \end{pmatrix} + \begin{pmatrix} k_t \\ k_t^* \end{pmatrix} \quad (3-58)$$

$$\mu_t = \mu_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2) \quad (3-59)$$

3.3.14

Estimativa dos Parâmetros

Os parâmetros fixos dos modelos apresentados anteriormente podem ser estimados utilizando o método da máxima verossimilhança.

3.3.15

Espaço de Estado e Filtro de Kalman

Os modelos com tendência estocástica, ciclo estocástico e ciclo e tendência estocásticos, todos eles, para serem estimados, necessitam ser representados na forma de espaço de estado. Uma vez nessa forma, a estimação do modelo é realizada através de um algoritmo chamado Filtro de Kalman.

Na forma espaço de estado uma série temporal y_t é expressa em termos de um conjunto de **variáveis de estado** que descrevem a sua dinâmica ao longo do tempo. A forma geral do espaço de estado é dada por:

$$\text{Equação das observações: } y_t = \mathbf{H}_t \mathbf{x}_t + v_t, \quad (3-60)$$

$$\text{Equação de transição: } \mathbf{x}_{t+1} = \mathbf{F}_t \mathbf{x}_t + \mathbf{G}_t \mathbf{w}_t. \quad (3-61)$$

onde: - \mathbf{x}_t representa o vetor de variáveis de estado no instante t ; - \mathbf{F}_t é a matriz de transição de estado, que descreve a evolução temporal das variáveis de estado; - \mathbf{w}_t é o vetor de ruídos do sistema, com $\mathbf{w}_t \sim N(0, \mathbf{Q}_t)$; - \mathbf{G}_t é uma matriz de seleção com valores 0 ou 1; - y_t é a observação da série temporal no instante t ; - \mathbf{H}_t é a matriz que relaciona as variáveis de estado com a observação y_t ; - $v_t \sim N(0, \mathbf{R}_t)$ representa o ruído da observação. Adicionalmente, assume-se que os erros w_t e v_t são descorrelacionados entre si, e também descorrelacionados da condição inicial x_1 .

O uso dessa formulação é necessária para especificar e estimar modelos de séries temporais onde as componentes variam ao longo do tempo, como no caso **tendência, sazonalidade e ciclo** são estocásticas. Nesse caso as componentes estarão contidas no vetor de estado.

O modelo com tendência e ciclo estocástico, especificado na Seção 3.3.13 pode ser escrito na forma de espaço de estado como apresentado a seguir.

$$y_t = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mu_t \\ \psi_t \\ \psi_t^* \end{pmatrix} + \varepsilon_t$$

$$\begin{pmatrix} \mu_t \\ \psi_t \\ \psi_t^* \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho \cos \lambda_c & \rho \sin \lambda_c \\ 0 & -\rho \sin \lambda_c & \rho \cos \lambda_c \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \psi_{t-1} \\ \psi_{t-1}^* \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_t \\ \kappa_t \\ \kappa_t^* \end{pmatrix}.$$

As matrizes especificadas representam diretamente a **equação de observação** e a **equação de transição** (Eqs. 3-60 e 3-61) do modelo de espaço de estado. Assim sendo, temos as seguintes definições das matrizes e vetores que definem a formulação em espaço de estado:

$$\mathbf{x}_t = \begin{pmatrix} \mu_t \\ \psi_t \\ \psi_t^* \end{pmatrix}, \quad \mathbf{H}_t = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix},$$

$$\mathbf{F}_t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \rho \cos \lambda_c & \rho \sin \lambda_c \\ 0 & -\rho \sin \lambda_c & \rho \cos \lambda_c \end{pmatrix}, \quad \mathbf{G}_t = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$v_t = \varepsilon_t, \quad \mathbf{w}_t = \begin{pmatrix} \eta_t \\ \kappa_t \\ \kappa_t^* \end{pmatrix}.$$

3.3.16

Estimativa via Filtro de Kalman

O **Filtro de Kalman** é o algoritmo utilizado para estimar as variáveis de estado em um modelo colocado na forma espaço de estado, como no caso do modelo de tendência e ciclo estocásticos utilizados no nosso trabalho. A partir desse algoritmo obtém-se a **previsão e atualização recursiva** dos estados a cada observação da série temporal. O filtro de Kalman é dado por duas equações:

- **Eq. da previsão** (passo de projeção 1 passo à frente): A partir das variáveis de estado no tempo t , o filtro calcula as previsões para o tempo $t + 1$:

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{F}_t \hat{\mathbf{x}}_{t|t}, \quad (3-62)$$

$$\mathbf{P}_{t+1|t} = \mathbf{F}_t \mathbf{P}_{t|t} \mathbf{F}_t' + \mathbf{G}_t \mathbf{Q}_t \mathbf{G}_t', \quad (3-63)$$

onde $\mathbf{P}_{t|t}$ é a matriz de covariância do estado estimado no tempo t .

- **Eq. de atualização** (passo de correção): Com a observação y_t no tempo $t+1$, o filtro ajusta as previsões usando o **ganho de Kalman** \mathbf{K}_{t+1} , calculado como:

$$\mathbf{K}_{t+1} = \mathbf{P}_{t+1|t} \mathbf{H}_t' (\mathbf{H}_t \mathbf{P}_{t+1|t} \mathbf{H}_t' + \mathbf{R}_t)^{-1}. \quad (3-64)$$

As estimativas ajustadas são então dadas por:

$$\mathbf{x}_{t+1|t+1} = \mathbf{x}_{t+1|t} + \mathbf{K}_{t+1} (y_{t+1} - \mathbf{H}_t \mathbf{x}_{t+1|t}), \quad (3-65)$$

$$\mathbf{P}_{t+1|t+1} = (\mathbf{I} - \mathbf{K}_{t+1} \mathbf{H}_t) \mathbf{P}_{t+1|t}. \quad (3-66)$$

onde

$$\hat{\mathbf{x}}_{t|t} = \mathbb{E}(\mathbf{x}_t | \mathbf{y}_t) \quad (3-67)$$

$$\mathbf{P}_{t|t} = \text{Var}(\mathbf{x}_t | \mathbf{y}_t) \quad (3-68)$$

$$\hat{\mathbf{x}}_{t+1|t} = \mathbb{E}(\mathbf{x}_{t+1} | \mathbf{y}_t) \quad (3-69)$$

$$\mathbf{P}_{t+1|t} = \text{Var}(\mathbf{x}_{t+1} | \mathbf{y}_t) \quad (3-70)$$

$$\mathbf{y}_t = (y_t, y_{t-1}, \dots, y_2, y_1)$$

3.3.17

Previsão

Pode-se mostrar que a previsão desse modelo será dada por:

$$\hat{y}_{T+k|T} = \hat{\mu}_{T+k|T} + \hat{\psi}_{T+k|T}, \quad (3-71)$$

onde

$$\hat{\mu}_{T+k|T} = \hat{\mu}_{T|T}, \quad k = 1, 2, \dots \quad (3-72)$$

com $\hat{\mu}_{T|T}$ sendo estimado pelo filtro de Kalman. A componente de ciclo da previsão é dada por:

$$\hat{\psi}_{T+k|T} = \rho^k \cos(k\lambda_c) \hat{\psi}_{T|T} + \rho^k \sin(k\lambda_c) \hat{\psi}_{T|T}^*, \quad (3-73)$$

onde $\hat{\psi}_{T|T}$ e $\hat{\psi}_{T|T}^*$ são valores obtidos pelo filtro de Kalman.

3.3.18

Aplicabilidade

Os modelos apresentados neste capítulo são amplamente aplicáveis em diferentes áreas, como economia, meteorologia e engenharias. No caso específico de séries cíclicas como a de manchas solares, esses modelos oferecem insights valiosos sobre a dinâmica dos ciclos e permitem prever comportamentos futuros com base em padrões passados.

3.4

Outras Ferramentas para Análise de Séries Temporais

Além das técnicas clássicas como a Análise de Fourier e a Análise de Ondas, existem outros métodos que complementam e ampliam a capacidade de interpretação das séries temporais, oferecendo perspectivas no domínio do tempo e da frequência.

Nesta seção, apresentamos algumas dessas ferramentas, incluindo o correlograma e testes estatísticos como o de normalidade e os testes aplicados aos resíduos dos modelos.

3.4.1

Correlograma

O **correlograma** é uma ferramenta utilizada para medir e visualizar o grau de dependência entre os valores de uma série temporal em diferentes intervalos de tempo. Ele representa a *autocorrelação* da série em função do *lag* (ou defasagem) k , permitindo identificar padrões de repetição ao longo do tempo e o comportamento cíclico do sinal.

Para uma série temporal estacionária y_t , a autocorrelação para um dado lag k é definida como:

$$r_k = \frac{\sum_{t=1}^{N-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^N (y_t - \bar{y})^2}, k = 1, 2, \dots \quad (3-74)$$

onde N é o número total de observações, y_t e y_{t+k} são valores da série temporal separados por k unidades de tempo, e \bar{y} é a média aritmética da série. O correlograma plota os valores de r_k para diversos valores de k , permitindo identificar as defasagens em que existe uma correlação significativa entre os pontos da série.

A análise do correlograma pode revelar diversas características da série temporal, a saber:

- Se as autocorrelações decaem rapidamente para valores próximos de zero conforme k aumenta, isso indica que a série deve ser *estacionária*.
- Se as autocorrelações oscilam ou decaem lentamente, isso sugere que a série possui uma estrutura sazonal ou cíclica.

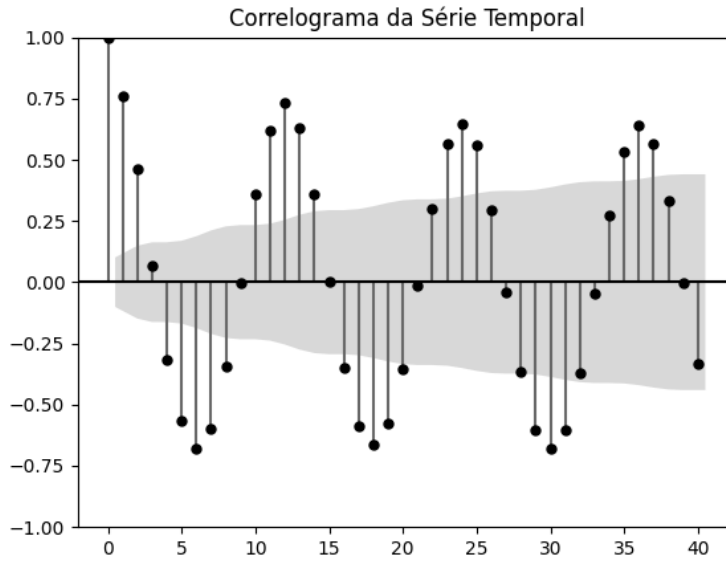


Figura 3.15: Exemplo de correlograma de uma série temporal mensal.

- A presença de picos em determinados lags “sazonais” (12 para série mensal) no correlograma pode indicar sazonalidade na série.

3.4.2

Teste de Normalidade

A normalidade de uma série temporal pode ser investigada através do teste de Jarque-Bera, especificado a seguir:

$$\begin{cases} H_0 : a \text{ série é normal} \iff S = 0 \text{ e } K = 3 \\ H_1 : a \text{ série não é normal} \iff c.c. \end{cases} \quad (3-75)$$

Onde a estatística de teste é dada por:

$$JB = \frac{N}{6}(\hat{S} - 0)^2 + \frac{N}{24}(\hat{K} - 3)^2,$$

sendo \hat{S} o coeficiente de assimetria amostral da série e \hat{K} o coeficiente de curtose amostral, estimados por:

$$\hat{S} = \frac{\sum_{t=1}^N (y_t - \bar{y})^3 / N}{\left(\sum_{t=1}^N (y_t - \bar{y})^2 / N \right)^{3/2}},$$

$$\hat{K} = \frac{\sum_{t=1}^N (y_t - \bar{y})^4 / N}{\left(\sum_{t=1}^N (y_t - \bar{y})^2 / N \right)^2}.$$

Sob H_0 , $JB \sim \chi^2(2)$.

3.4.3

Métricas para Quantificação da Acurácia Preditiva

A avaliação da qualidade das previsões em séries temporais pode ser feita por meio de diferentes métricas estatísticas, que quantificam a discrepância entre os valores previstos ($\hat{y}_{T+k|T}$) e os valores observados (y_{T+k}) no período de teste. Nessa notação T é a última observação do período de treinamento, e k é o horizonte de previsão no período de teste, $k = 1, 2, \dots, L$; o condicionante $|T$ no sub-escrito $T + k|T$ indica que a previsão será obtida no período de teste $T + k$ usando toda a informação da série contida no período de treinamento.

Neste trabalho, utilizamos três das principais medidas de erro preditivo: o Erro Quadrático Médio (RMSE), o Erro Absoluto Médio (MAE) e o Erro Absoluto Médio Padronizado (MASE), que serão descritas a seguir.

Erro Quadrático Médio (RMSE) - O RMSE (*Root Mean Squared Error*) mede a raiz quadrada da média dos erros quadráticos, sendo definido por:

$$RMSE = \sqrt{\frac{1}{L} \sum_{k=1}^L (y_{T+k} - \hat{y}_{T+k|T})^2} \quad (3-76)$$

O RMSE penaliza erros maiores devido à elevação ao quadrado, sendo sensível a valores discrepantes (outliers). Quanto menor o RMSE, melhor a qualidade da previsão.

Erro Absoluto Médio (MAE) O MAE (*Mean Absolute Error*) quantifica o erro médio em termos absolutos, sendo definido por:

$$MAE = \frac{1}{L} \sum_{k=1}^L |y_{T+k} - \hat{y}_{T+k|T}| \quad (3-77)$$

Diferentemente do RMSE, o MAE trata erros positivos e negativos da mesma forma, tornando-o mais intuitivo para interpretação.

Essas métricas são essenciais para comparar a qualidade dos modelos ajustados, auxiliando na escolha da abordagem mais adequada para previsões futuras.

3.5

Implementação

Nesta seção, detalhamos os procedimentos para a implementação das técnicas de Análise de Fourier e Análise de Ondaletas, empregando bibliotecas e ferramentas computacionais da linguagem Python.

3.5.1

Análise de Fourier

Para a implementação da Análise de Fourier e o cálculo da densidade espectral de potência, utilizamos as bibliotecas *numpy* (<https://numpy.org/doc/stable/>) e *scipy* (<https://docs.scipy.org/doc/scipy/>). A função *fft* da biblioteca *numpy* foi empregada para calcular os coeficientes da Transformada Rápida de Fourier (FFT) (<https://numpy.org/doc/stable/reference/routines.fft.html>), possibilitando a obtenção das frequências e amplitudes das componentes harmônicas de um sinal.

Adicionalmente, para estimar a densidade espectral de potência, utilizamos a função *periodogram* da biblioteca SciPy (*scipy.signal.periodogram*) (<https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.periodogram.html>). Esta ferramenta calcula a distribuição das frequências no sinal e a potência associada a cada frequência. Para melhorar a estabilidade das estimativas espectrais, implementamos também o método de Welch, disponível como *scipy.signal.welch* (<https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.welch.html>). Este método divide o sinal em segmentos sobrepostos e calcula a média dos espectros individuais, reduzindo a variância e oferecendo uma visualização mais clara das frequências predominantes.

Com o propósito de comparar resultados, utilizamos o método de mínimos quadrados, implementado pelo comando *OLS* da biblioteca *statsmodels.api* (*statsmodels.regression.linear_model.OLS*) (https://www.statsmodels.org/stable/generated/statsmodels.regression.linear_model.OLS.html). Esta abordagem permitiu ajustar modelos lineares aos dados e calcular os coeficientes de Fourier a_n e b_n com alta precisão, facilitando a interpretação das componentes frequenciais.

3.5.2

Análise de Ondaletas

Para a Análise de Ondaletas, utilizamos a biblioteca *PyWavelets* do Python (<https://pywavelets.readthedocs.io/>), que oferece suporte para realizar a decomposição do sinal em diferentes escalas e obter a Transformada de Ondaletas. Essa ferramenta permite analisar tanto as frequências quanto a localização temporal das variações no sinal, característica essencial para sinais não estacionários.

A *PyWavelets* possibilita a aplicação de diferentes tipos de ondaletas, como a ondaleta de Daubechies e a ondaleta de Haar. A função *pywt.wavedec* foi utilizada para realizar a decomposição do sinal em componentes de alta e baixa frequência, enquanto a função *pywt.waverec* foi empregada para a reconstrução do sinal.

Com essa abordagem, conseguimos identificar eventos transientes e variações abruptas no sinal, que são características importantes em séries temporais complexas. Além disso, utilizamos a função de limiarização `pywt.threshold` para realizar o encolhimento de coeficientes, reduzindo o ruído sem comprometer as informações relevantes.

3.5.3

Modelos Estruturais para Ciclo Estocástico

O programa *Time Series Lab* (TSLab)[15] é uma ferramenta especializada para a modelagem e análise de séries temporais. O TSLab oferece suporte abrangente para a implementação de modelos estruturais, permitindo a decomposição de séries temporais em componentes como tendência, sazonalidade e ciclos estocásticos.

Com uma interface intuitiva e funcionalidades robustas, o programa facilita a estimação de parâmetros por métodos de mínimos quadrados generalizados e filtros de Kalman. Além disso, sua capacidade de gerar visualizações detalhadas dos resultados auxilia na interpretação e na validação dos modelos ajustados. No presente trabalho, o TSLab foi utilizado para ajustar modelos estruturais que descrevem a natureza estocástica dos ciclos observados na série temporal das manchas solares, permitindo a investigação da variação de amplitude e período ao longo do tempo.

A escolha pelo *Time Series Lab* deve-se à sua capacidade de lidar com modelos complexos de forma eficiente, proporcionando resultados confiáveis e interpretações significativas no contexto dos objetivos deste trabalho.

4

Resultados

Neste capítulo iremos apresentar os resultados referentes a nossa investigação sobre a estocacidade do ciclo da série de manchas solares. Iniciamos com a estimativa do período do ciclo utilizando a Análise de Fourier (AF) conforme descrita no Capítulo 3. Observe que na aplicação dessa técnica assume-se que a série possui ciclo com período fixo ao longo do tempo. Posteriormente, através de um estudo exploratório da série de manchas solares apresentamos, de forma simples e direta, evidências sobre a natureza estocástica desse ciclo. Mais precisamente apresentamos evidências estatísticas sobre a variabilidade temporal da amplitude e do período da série. A Análise de ondaleta (AO), apresentada na Seção 3.2, foi utilizada para decompor o sinal em suas componentes de ciclo, tendência e erro aleatório por meio da DWT para uma investigação isolada em cima da componente de ciclo. Utilizamos a CWT para gerar escalogramas, que fornecem uma representação visual da variação do período e da amplitude em função do tempo.

Finalmente, a partir da utilização do modelo estrutural de ciclo e tendência estocásticos de Harvey [10], confirmamos as evidências trazidas pelo estudo exploratório sobre a natureza estocástica desse ciclo.

4.1

O ciclo da série de manchas solares

A Fig. (4.1) apresenta a série temporal das manchas solares com início em 1749 e fim em 2024, perfazendo um total de $N = 275$ observações.

Para estimar o período da série iremos utilizar dois procedimentos apresentados no Capítulo de Metodologia, o correlograma (ou função de autocorrelação (FAC)) e o periodograma, obtido via AF. Como é sabido que essa série possui período de aproximadamente 11 anos (Ciclo de Schwabe), então se espera que a sua FAC apresente autocorrelação positiva e relevante em torno do lag 11. Este padrão pode ser confirmado observando a Fig. (4.2) e a Tabela 4.1, evidenciando o que a literatura de manchas solares afirma.

A mesma informação sobre a periodicidade de 11 anos pode também ser obtida utilizando a Análise de Fourier (AF). Como vimos no Capítulo 3.1, a AF decompõe uma série temporal em suas frequências harmônicas, possibilitando

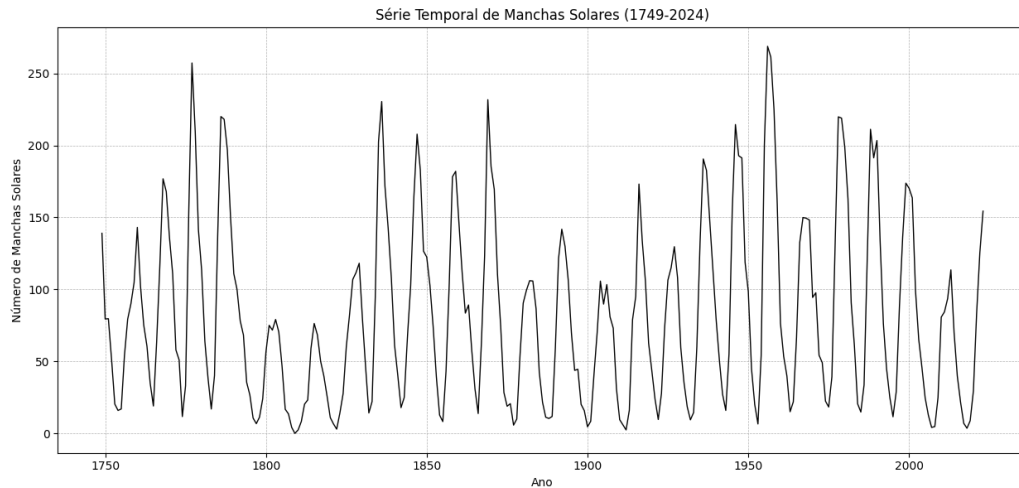


Figura 4.1: Série temporal do número médio de manchas solares (1749-2024).

Fonte: Solar Influences Data Analysis Center (SIDC).[9]

k	correlação
9	0.4375
10	0.6377
11	0.6386
12	0.4513

Tabela 4.1: Valores da correlação da série de manchas solares.

identificar a contribuição de cada frequência para a variação da série. As frequências de maior amplitudes serão aquelas responsáveis pelos ciclos na série.

Na Tabela (4.2) apresentamos os resultados da aplicação da Transformada de Fourier (FFT) à série de manchas solares. Observe que o harmônico 29 possui o maior valor de r_n (a amplitude), indicando que esta é a frequência de maior influência nessa série, estando diretamente associada ao ciclo solar de 11 anos, o que é facilmente verificado pelo cálculo $T = \frac{N}{n}$, logo $T = \frac{323}{29} \approx 11,1$.

Estimação mais confiável do período da série de manchas solares pode ser obtida através do periodograma suavizado apresentado na Figura (4.3). Observe que esse procedimento permite obter evidência de algumas frequências de ciclos além do ciclo de Schwabe de 11 anos, conforme apresentadas na Tabela 2.1 no Capítulo 2.

Observe que as estimativas obtidas do período da séries de manchas solares via FAC e periodograma não permitem inferir a provável natureza estocástica do ciclo, uma vez que estes procedimentos assumem, por construção, valores fixos para o período da série e sua amplitude ao longo do tempo.

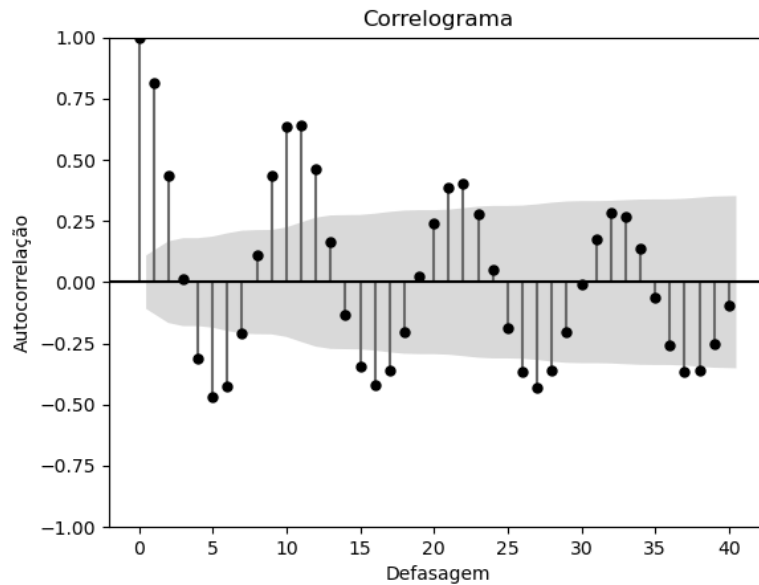


Figura 4.2: Correlograma da série temporal de manchas solares.
Fonte: Solar Influences Data Analysis Center (SIDC).

n	a_n	b_n	r_n
3	-24.7769	0.772613	24.78899
6	-15.7174	-4.21562	16.2729
27	-3.99542	23.47145	23.80908
29	6.026598	37.49471	37.97595
30	17.47783	-16.9584	24.35286
31	-14.8027	-21.1266	25.79635
32	-22.0152	-25.0791	33.3711
38	12.351	6.373306	13.89842
40	-2.24683	-12.0147	12.22298

Tabela 4.2: Coeficientes da FFT para a série de manchas solares.
Fonte: Solar Influences Data Analysis Center (SIDC).[9]

4.2

Evidências da estocasticidade do ciclo da série de manchas solares

A análise da série de manchas solares revela variações na amplitude e na periodicidade do ciclo de Schwabe ao longo do tempo, sugerindo a presença de componentes estocásticas. Essas flutuações são evidenciadas através de algumas análises apresentadas a seguir.

4.2.1

Análise exploratória

A natureza estocástica do ciclo das manchas solares será inicialmente estabelecida através de algumas análises exploratórias. Uma simples inspeção visual do gráfico da série apresentado na Fig. (4.1) mostra facilmente que a amplitude da

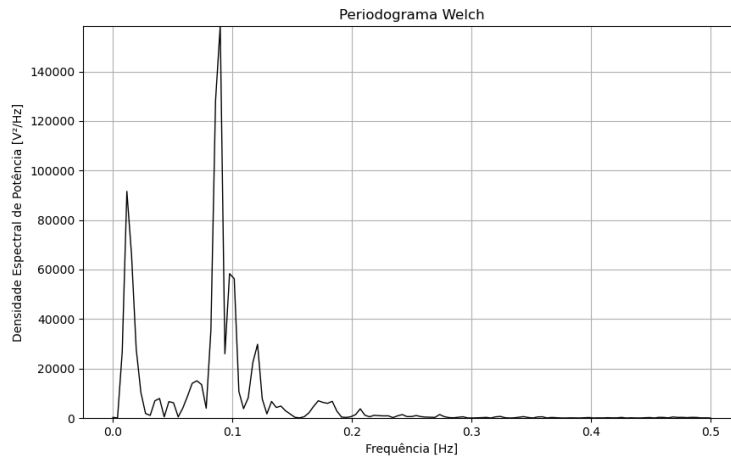


Figura 4.3: Periodograma suavizado da série de manchas solares.

Fonte: Solar Influences Data Analysis Center (SIDC).[9]

série não permanece a mesma ao longo do tempo, alterando-se entre períodos de baixa e alta amplitude. Esse comportamento pode ser melhor ilustrado observando-se o padrão das séries de mínimos e de máximos consecutivos obtidos a partir da série de manchas solares. Se a amplitude fosse determinística cada uma dessas séries apresentaria um valor único ao longo do tempo, mas não é isso que ocorre. Conforme ilustrado nos gráficos das Figuras (4.4) e (4.5), as séries de máximos e de mínimos consecutivos exibem uma variabilidade significativa ao longo do tempo, evidenciando o caráter estocástico da amplitude do ciclo das manchas solares.

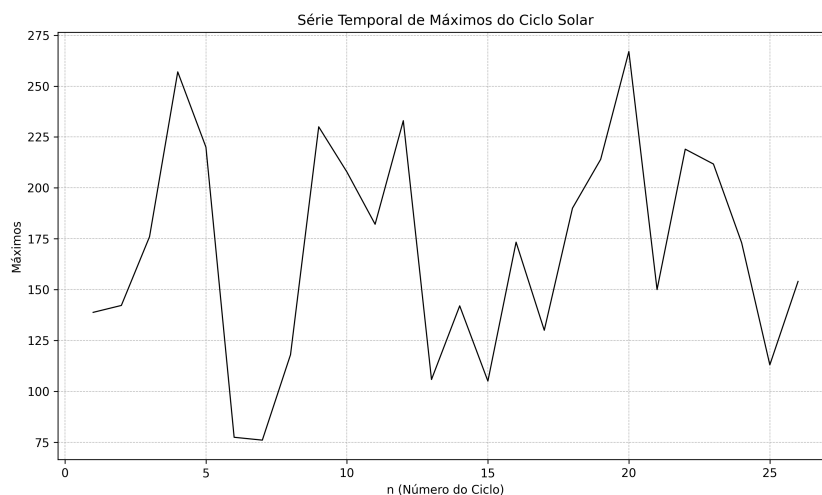


Figura 4.4: Série temporal dos máximos consecutivos das manchas solares.

Agora iremos investigar, de forma exploratória, a natureza estocástica do período da série de manchas solares.

A Figura (4.6) apresenta a série temporal dos períodos das manchas solares obtida a partir da mensuração do tempo entre dois mínimos solares consecutivos,

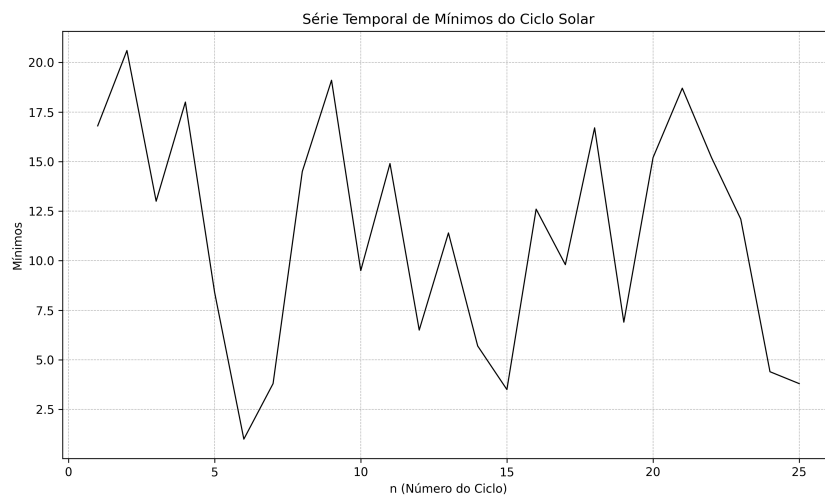


Figura 4.5: Série temporal dos mínimos consecutivos das manchas solares.

entre 1755 e 2014, resultando em 24 valores distintos para o período. Embora o período médio do ciclo solar seja de aproximadamente 11 anos, as flutuações observadas nos valores dessa série indicam a natureza estocástica do período dessa série.

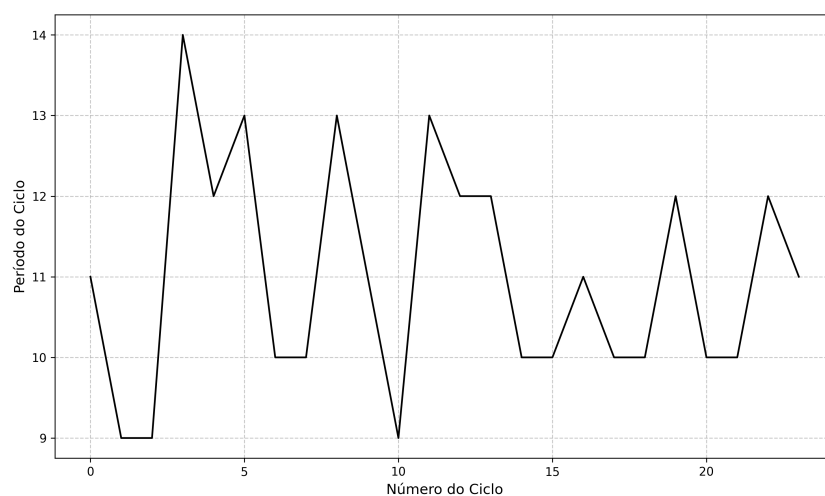


Figura 4.6: Série da variação do período dos ciclos solares.

A variabilidade do período do ciclo solar poderá melhor ser evidenciada através das estatísticas descritivas dessa série, conforme ilustradas na Tabela (4.2.1) a seguir. Assim, vemos que o ciclo das manchas solares tem variado entre 9 e 14 anos, com valor médio de 11 anos.

A variabilidade observada na amplitude e no período da série de manchas solares ao longo dos anos sugere a série de manchas solare não pode ser descrita adequadamente por um modelo de ciclo determinístico. Em particular observa-se que há maior variabilidade na amplitude do que no ciclo. Essa variabilidade reflete a complexidade dos fenômenos físicos subjacentes ao ciclo solar, reforçando

estatística	valor
N	24
mínimo	9
máximo	14
média	11
mediana	11
desvio padrão	1.41

Tabela 4.3: Estatísticas descritivas da série dos períodos do ciclo solar

a necessidade de modelos que incorporem componentes estocásticas para capturar adequadamente essas dinâmicas.

4.2.2

Evidências suplementares

A natureza estocástica do ciclo da série de manchas solares evidenciada, de forma exploratória, na sessão anterior, será aqui aprofundada. Para isso iremos dividir a série em três segmentos consecutivos de mesmo tamanho ($N=92$ obs), descritos a seguir: 1749-1840, 1841-1932 e 1933-2024. Para cada um desses segmentos foi estimado o periodograma suavizado utilizando o método de Welch, com o objetivo de identificar o período dos ciclos. O resultado é apresentado na Tabela 4.2.2.

Segmento	Frequência (Hz)	Período (anos)
1749-1840	0.111	9.01
1841-1932	0.0889	11.25
1933-2024	0.087	11.49

Tabela 4.4: Frequências dominantes e períodos estimados para cada segmento.

A análise realizada permite observar que o ciclo de Schwabe está presente de forma consistente nos três segmentos da série temporal, com pequenas variações dos períodos estimados, corroborando a evidência anterior de que a variabilidade do período do ciclo é bem menos acentuada do que a da sua amplitude. Essa caracterização é fundamental para compreender as dinâmicas cíclicas da atividade solar.

4.2.3

Aplicação da Análise de Ondaletas

Nessa subseção iremos utilizar técnicas baseadas em ondaletas, vistas no Capítulo 3, seção 3.2, para melhor evidenciar a natureza estocástica do ciclo de manchas solares.

Nosso primeiro resultado utiliza a decomposição a série de manchas solares nas

componentes de tendência, ciclo e erro aleatório via Transformada Discreta de Ondas (DWT). Nessa aplicação a DWT foi estimada considerando a ondaleta Daubechies-4 (*db4*), permitindo uma análise hierárquica das frequências do sinal. É importante recordar que quanto maior for o nível de resolução adotado na decomposição, mais informação é removida do coeficiente de aproximação (baixa frequência), sendo realocada aos coeficientes de detalhe (alta frequência). Desta forma é interessante que o nível seja escolhido de forma que a informação do ciclo seja devidamente separada da tendência, mas sem que se perca informação relevante da tendência. Assim sendo, nessa aplicação usamos 5 níveis de decomposição, resultando em 5 grupos de coeficientes. A Fig. (4.7) mostra, além do gráfico da

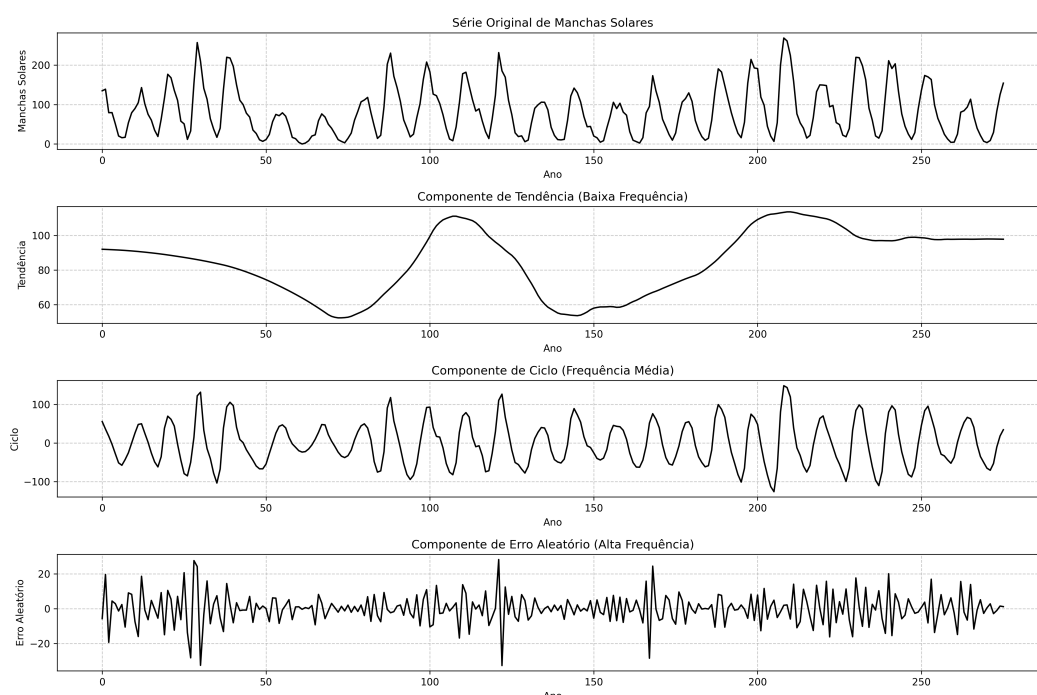


Figura 4.7: Decomposição wavelet da série de manchas solares.

série de manchas solares, os gráficos obtidos dessa decomposição. A passagem dos coeficientes de aproximação e de detalhe para separar os componentes da série em diferentes escalas de frequência está detalhada na Seção 3.2. Estes coeficientes foram separados de maneira oportuna em 3 séries diferentes(ver Fig4.7), por nós nomeadas como componentes de tendência, ciclo e erro aleatório, como descritos a seguir.

- A primeira (figura de cima para baixo) ilustra a série original das manchas solares, que é constituída por **todos** os coeficientes resultantes da decomposição por ondaleta.
- A segunda figura representa a tendência (*trend*) extraída da série original, sendo formada pelos coeficientes de aproximação (nível 1). Note que a

tendência parece capturar um ciclo de baixa frequência da série de manchas solares (aprox. 110 anos), correspondente ao ciclo de Gleissberg, conforme visto na Tabela (2.1).

- A terceira figura representa o ciclo estocástico de 11 anos (*cycle*) formado pelos coeficientes de detalhe do nível 2, 3 e 4.
- A quarta figura representa o erro ou resíduo (*error*), representando as diferenças não explicadas pela tendência e pelo ciclo, formado pelos coeficientes de detalhe de nível 5 (os de maiores frequências).

Para melhor evidenciar a nossa decisão de associar à terceira figura o ciclo de Schwabe de aproximadamente 11 anos, apresentamos a seguir, o correlograma dessa série na Fig (4.8), complementado pela Tabela 4.5, que detalha os valores numéricos dessa autocorrelação para os lags $k = 9, 10, 11$ e 12 anos.

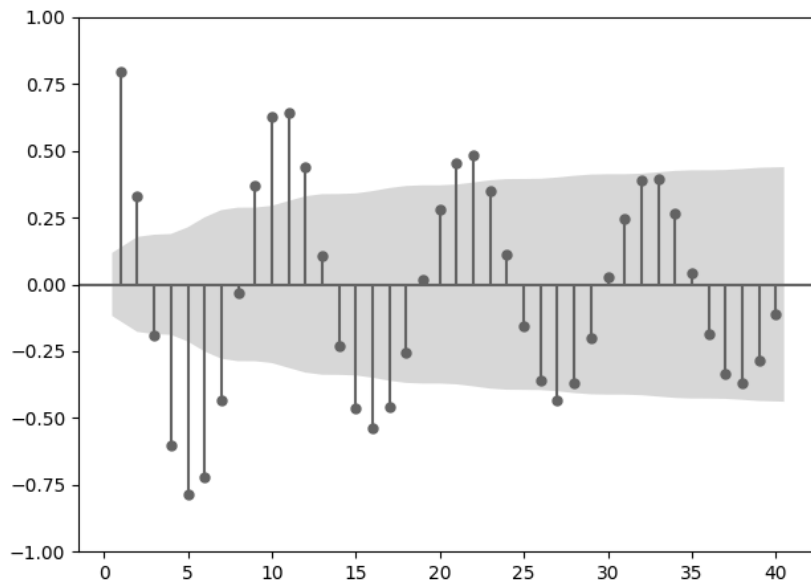


Figura 4.8: Correlograma da componente de ciclo obtida por DWT.

lag k	autocorrelação
9	0.3785
10	0.6397
11	0.6544
12	0.4483

Tabela 4.5: Autocorrelação para a componente de ciclo obtida por DWT.

Como pode ser observado na Tabela 4.5, a autocorrelação da componente de ciclo obtida por DWT é alta para $k = 10, 11$ e 12 anos, indicando que essa componente captura o ciclo de Schwabe.

A natureza estocástica do ciclo obtido via DWT pode ser inferida observando-se os padrões apresentados por essa série conforme Fig. 4.7. É facilmente observável a variação da amplitude do ciclo ao longo do tempo.

Em relação a estocacidade do período, de forma análoga ao que foi feito com a série de manchas solares, iremos obter a série entre os mínimos consecutivos do ciclo extraído por DWT. Essa série, mostrada na Fig. 4.9 evidencia de forma clara a natureza estocástica desse ciclo. O seus 24 valores distintos, variando entre 9 e 14 anos, indicam a natureza estocástica do período do ciclo de Schwabe, e corroboram os resultados reportado por [14]. Essa análise fornece uma base mais confiável para investigar a estocasticidade e possíveis variações do ciclo solar de Schwabe.

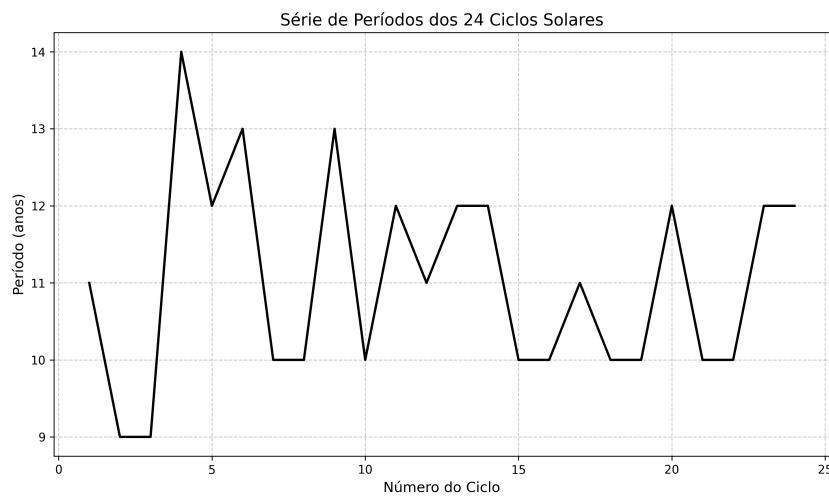


Figura 4.9: Períodos dos 24 ciclos solares calculados a partir da componente de ciclo extraída por ondaletas.

Para finalizar nossa análise iremos obter um intervalo de confiança (IC) de 95% para o valor fixo do período desse ciclo. Utilizando a normalidade desses valores (p -valor do teste Jarque-Bera = 0.57), o IC de 95% é dado por (8.5, 13.6) anos, o que resulta em evidências favoráveis a estocacidade do ciclo. Essa análise reforça a importância de trabalhar diretamente com a componente de ciclo extraída por ondaletas para uma caracterização mais robusta do período do ciclo de Schwabe.

Escalograma A Fig. (4.10) apresenta o escalograma da série temporal de manchas solares, que evidencia a variação do período e da amplitude do ciclo de Schwabe ao longo do tempo. O eixo de escala, inversamente proporcional à frequência e diretamente proporcional ao período T , permite identificar as diferentes escalas temporais presentes no sinal.

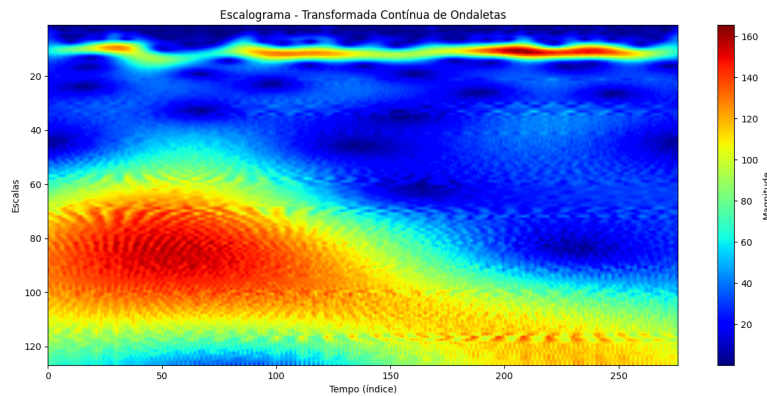


Figura 4.10: Escalograma da série de manchas solares.

As cores no escalograma representam a magnitude dos coeficientes de ondaleta: regiões mais quentes (próximas ao vermelho) indicam maior magnitude do sinal, enquanto regiões mais frias (próximas ao azul) correspondem a baixas magnitudes. O ciclo de Schwabe pode ser facilmente identificado pelas áreas de maior magnitude, localizadas próximas à escala de 10. Essa representação visual permite observar a dinâmica estocástica do ciclo de Schwabe ao longo do tempo.

4.3

Modelagem da Série de Manchas Solares via Modelo de Tendência e Ciclo Estocásticos

Nessa seção, iremos estimar a série de manchas solares através do modelo estrutural com tendência e ciclo estocásticos, conforme apresentado no Capítulo de Metodologia, Seção 3.3. A estocacidade do ciclo será evidenciada pela estimativa da variância do choque aleatório que impacta essa componente. Se essa estimativa for não nula, e com magnitude não “desprezível”, então pode-se concluir que o ciclo é estocástico.

A seguir, apresentamos os resultados do ajuste desse modelo à série de manchas solares, obtido a partir da utilização do software Time Series Lab [15]. Primeiro iremos ajustar um modelo para a série completa, de 1749 até 2024 e observar as estimativas de cada parâmetro ajustadas.

Parâmetro	Descrição	Estimativa	IC 95%
σ_ϵ^2	variância do choque das observações	0	0
σ_η^2	variância do choque da tendência	66,91	(31.3524, 142.792)
σ_K^2	variância do choque do ciclo	318,91	(252.999, 402.0)
T	período do ciclo	10,53	(10.009, 11.086)
ρ	fator de amortecimento	0,95	(0.9154, 0.9714)

Tabela 4.6: Tabela de parâmetros para modelo ajustado à série de manchas solares.

Agora iremos ajustar o mesmo modelo para cada segmento da série de manchas solares, conforme a descrição na Tab.(4.2.2). Os resultados são reportados a seguir.

Segmento 1 (1749-1840)			
Parâmetro	Descrição	Estimativa	IC 95%
σ_K^2	variância do choque do ciclo	299.4283	(191.1680, 468.9974)
T	período do ciclo	9.7730	(8.8121, 10.8695)
σ_η^2	variância do choque da tendência	127.2814	(49.0808, 330.0791)
ρ	fator de amortecimento	0.9372	(0.8587, 0.9734)

Tabela 4.7: Parâmetros do segmento 1 para série de manchas solares.

Segmento 2 (1841-1932)			
Parâmetro	Descrição	Estimativa	IC 95%
σ_K^2	variância do choque do ciclo	315.8972	(228.1012, 437.4860)
T	período do ciclo	11.1090	(10.05, 12.3072)
σ_η^2	variância do choque da tendência	14.4417	(2.2830, 91.3559)
ρ	fator de amortecimento	0.9397	(0.8693, 0.9733)

Tabela 4.8: Parâmetros do segmento 2 para série de manchas solares.

Segmento 3 (1933-2024)			
Parâmetro	Descrição	Estimativa	IC 95%
σ_K^2	variância do choque do ciclo	351.6755	(237.0101, 521.8159)
T	período do ciclo	10.4745	(9.7061, 11.3195)
σ_η^2	variância do choque da tendência	43.4238	(6.5436, 288.1646)
ρ	fator de amortecimento	0.9604	(0.8987, 0.9852)

Tabela 4.9: Parâmetros do segmento 3 para série de manchas solares.

Assim, podemos concluir que, considerando a série como um todo e os seus 3 segmentos, o ajuste do modelo estrutural evidencia a natureza estocástica do ciclo de 11 anos, uma vez que, em todas essas situações, a variância do choque do ciclo (σ_K^2) apresenta valor estimado bem acima de zero.

4.3.1

Ajuste do Modelo com Ciclo Determinístico Monocromático

A idéia nessa sessão é apresentar evidências suplementares sobre a natureza estocástica do ciclo das manchas solares comparando a acurácia preditiva do modelo com ciclo estocástico com a do modelo com ciclo determinístico apresentado na Seção 3.3.1. Se a natureza do ciclo da série de manchas solares é realmente estocástica, então espera-se que a acurácia preditiva do modelo estocástico seja melhor do que aquela do modelo determinístico. Para tornar a comparação justa, iremos considerar a mesma tendência estocástica em ambos os modelos.

Inicialmente ajustamos o modelo estrutural com tendência e ciclo estocásticos (ver modelo 3.3.3) à série de manchas solares no período de treinamento, entre 1840 e 2012. Esse modelo foi então utilizado para prever a série no período de teste, entre 2013 e 2024, perfazendo 12 anos. A partir desse modelo obtivemos a estimativa da componente de tendência estocástica no período de treinamento $\hat{\mu}_t$, com a qual foi obtida a série sem tendência, $Z_t = y_t - \hat{\mu}_t$. O modelo de ciclo determinístico (ver modelo 3.3.1) foi ajustado utilizando o método de MQO a essa série sem tendência. Entretanto, para que possamos usar MQO é necessário, antes do ajuste, identificarmos o período da série Z_t . Através da **Análise de Fourier** obtivemos a seguinte estimativa do período $T_c = 10.81$. Em seguida, utilizamos o **Método dos Mínimos Quadrados (MQO)** para estimar os coeficientes α e β no modelo de ciclo determinístico à série Z_t :

$$Z_t = \alpha \cos(\lambda_c t) + \beta \sin(\lambda_c t) + \varepsilon_t.$$

Onde $\lambda_c = \frac{2\pi}{10.81}$. Com isso, construímos uma representação determinística do ciclo baseada nos coeficientes ajustados. Por fim, para obter a previsão para a série original no período de teste a partir desse modelo, utilizamos que:

$$\begin{aligned} Z_t &= y_t - \hat{\mu}_t \\ y_t &= Z_t + \hat{\mu}_t \end{aligned}$$

Fazendo $t = T + k$ nessa relação, onde T é o último ano do período de treinamento, portanto $T = 2012$, e k é o horizonte de previsão, $k = 1, 2, \dots, 12$; o condicionante $|T$ no sub-escrito $T + k|T$ indica que a previsão será obtida no período de treinamento $T + k$ usando toda a informação da série contida no período de treinamento. Portanto segue que a previsão para a série original será dada por

$$\hat{y}_{T+k|T} = \hat{Z}_{T+k|T} + \hat{\mu}_{T+k|T}$$

onde $\hat{Z}_{T+k|T}$ é a previsão do modelo com ciclo determinístico e $\hat{\mu}_{T+k|T}$ a previsão da tendência estocástica. A previsão realizada por ambos os modelos para o período de 2012 a 2024 pode ser visualizada nas Figuras 4.11 e 4.12 a seguir.

Os gráficos mostram que o modelo com ciclo estocástico (eq 3-47) capturou melhor as flutuações da série temporal, refletindo sua dinâmica real com maior precisão. Em contrapartida, o modelo com ciclo determinístico monocromático (eq 3-43) apresentou previsões mais rígidas, sem conseguir representar adequadamente as variações de amplitude e período ao longo do tempo.

Esses resultados podem ser melhor apresentados comparando a acurácia preditiva dos dois modelos pelas métricas de RMSE, MAE e MASE, no período treinamento, conforme a Tabela 4.3.1 a seguir.

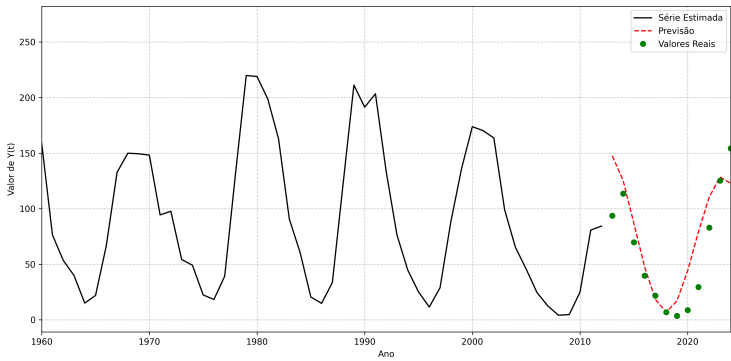


Figura 4.11: Previsão do modelo com ciclo determinístico (2012-2024).

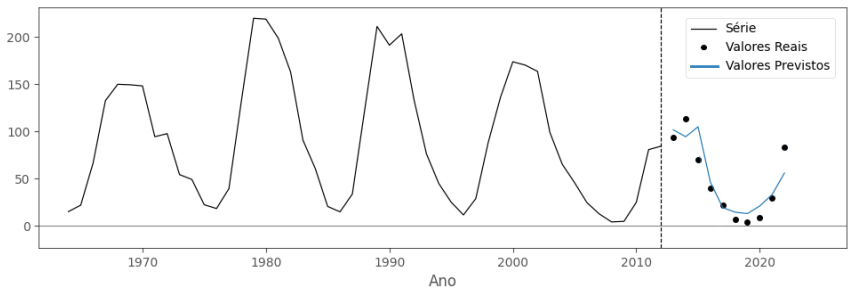


Figura 4.12: Previsão do modelo com ciclo estocástico (2012-2024).

Métrica	Ciclo Determinístico	Ciclo Estocástico
RMSE	27.4743	16.6100
MAE	21.2418	13.1200

Tabela 4.10: Comparação dos testes de aderência para os modelos com ciclo determinístico e ciclo estocástico.

Os resultados evidenciam que o modelo com ciclo estocástico apresentou melhor desempenho, uma vez que todos os indicadores de acurácia preditiva foram significativamente menores. Esse comportamento sugere que a dinâmica do ciclo solar não é rigidamente determinística, mas sim sujeita a variações aleatórias que afetam sua amplitude e periodicidade ao longo do tempo.

5

Conclusão

Este trabalho teve como objetivo principal investigar a natureza estocástica do ciclo de Schwabe da série temporal de manchas solares, empregando uma abordagem metodológica estruturada em técnicas estatísticas exploratórias de dados, análise de Fourier (AF), análise por ondaletas (AO) e modelagem estrutural. Os resultados obtidos evidenciam, de forma convincente, a natureza estocástica da série de manchas solares.

Inicialmente, foram empregadas técnicas exploratórias para investigar o comportamento do ciclo de Schwabe. A **Análise de Fourier** revelou como principal harmônico aquele associado à periodicidade de aproximadamente 11 anos. De forma complementar, o periodograma suavizado pelo *método de Welch* confirmou a periodicidade de 11 anos desse ciclo.

Estatísticas descritivas calculadas para diferentes segmentos da série de manchas solares corroboraram a hipótese de estocasticidade do ciclo, evidenciando flutuações na sua amplitude e na periodicidade ao longo dos anos. Essas variações reforçam a ideia de que o ciclo não é puramente determinístico, mas sim sujeito a incertezas e fatores externos.

Por outro lado, a análise por ondaletas foi utilizada para decompor a série de manchas solares nas suas componentes cíclicas, de tendência e erro aleatório. A **Transformada Discreta de Ondaletas (DWT)** foi utilizada para extrair a componente de ciclo da série, permitindo uma análise mais detalhada das características do ciclo de Schwabe. Esta técnica possibilitou a separação de componentes de alta frequência, como ruídos e perturbações, preservando a tendência cíclica subjacente. A análise da componente cíclica extraída por esse procedimento confirmou sua natureza estocástica.

Adicionalmente, a **Transformada Contínua de Ondaletas (CWT)** foi empregada para construir o escalograma dessa série, oferecendo uma representação detalhada do comportamento da escala/frequência do ciclo em diferentes tempos. A análise do escalograma evidenciou a variabilidade da amplitude e da periodicidade ao longo do tempo, destacando anos específicos em que a atividade solar apresentou comportamento não determinístico, confirmando a natureza estocástica do ciclo.

Finalmente uma investigação mais formal da natureza estocástica do ciclo

das manchas solares foi efetuada utilizando o modelo estrutural de ciclo estocástico proposto por Harvey [10]. A implementação desse modelo foi realizada através do software *Time Series Lab* [15]. O modelo foi ajustado a série de manchas solares como um todo e aos três segmentos conformes especificados na Tabela 4.4. Os ciclos extraídos para a série de manchas solares e os seus segmentos, evidenciaram, mais uma vez, a sua natureza estocástica, devido a magnitude da variância do choque dessa componente, conforme pode ser observado nas Tabelas 4.6 , 4.7, 4.8 e 4.9.

Finalmente, foram ajustados dois modelos distintos à série de manchas solares no período de 1840-2012 (período de teste): um modelo com ciclo determinístico e tendência estocástica (eq 3-43) e outro com ciclo estocástico e tendência estocástica (eq 3-47). Para comparar a acurácia preditiva dos modelos, foram realizados previsões no período de teste entre 2012 a 2024 (12 anos). As medidas de acurácia preditiva evidenciaram que o modelo com ciclo e tendência estocásticos, apresentou melhor desempenho do que o modelo com ciclo determinístico. Essa evidência reforça a hipótese de que o ciclo de manchas solares possui características estocásticas, uma vez que o modelo que incorpora um ciclo estocástico foi mais eficaz em capturar a dinâmica subjacente dos dados no período de treinamento, e prever melhor a série no período de teste.

Como trabalhos futuros, sugerimos a aplicação de técnicas de aprendizado de máquina para aprimorar as previsões do ciclo solar, bem como a análise de componentes adicionais, como o ciclo de Gleissberg, para um entendimento mais abrangente das dinâmicas solares.

Referências bibliográficas

- [3] MORETTIN, P. A.. **Ondas e Ondaletas: Da Análise de Fourier à Análise de Ondaletas de Séries Temporais**. Editora da Universidade de São Paulo (EDUSP), São Paulo, 2014.
- [5] MEYER, Y.. **Wavelets: Algorithms and Applications**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1993.
- [6] COOLEY, J. W.; TUKEY, J. W.. **An algorithm for the machine calculation of complex fourier series**. *Mathematics of Computation*, 19(90):297–301, 1965.
- [7] CLETTE, F.; SVALGAARD, L. V. J. M. C. E. W.. **Revisiting the sunspot number: A 400-year perspective on the solar cycle**. *Space Science Reviews*, 186(1-4):35–91, 2014.
- [8] NASA. **What is the solar cycle?**, 2024. Acesso em: 25 nov. 2024.
- [9] SIDC. **Sunspot number monthly and yearly data**. Acesso em: 18 nov. 2024.
- [10] HARVEY, A. C.. **Forecasting, Structural Time Series Models and the Kalman Filter**. Cambridge University Press, Cambridge, UK, 1990.
- [12] EPA. **Radiation from solar activity**, 2024. Acesso em: 25 nov. 2024.
- [14] CHARBONNEAU, P.. **Dynamo models of the solar cycle**. *Living Reviews in Solar Physics*, 2:1–128, 2005.
- [15] LAB, T. S.. **Time series lab**. <https://www.timeserieslab.com>, 2025. Software utilizado para modelagem de séries temporais.
- [16] HATHAWAY, D. H.. **O ciclo solar**. arXiv preprint, 1502.07020v1, 2015.
- [22] SIDC. **Calculation of the international sunspot number**. <https://www.sidc.be/aboutSIDC/paper.php>, 2024. Acesso em: 05 dez. 2024.
- [23] NASA/ESA. **Sunspots - solar and heliospheric observatory (soho)**, 2024. Acesso em: 25 nov. 2024.

- [24] MORIN, D.. **Fourier analysis**, 2009. Capítulo de um livro potencial sobre ondas. Versão 1, 28 de novembro de 2009.
- [25] ECHER, E.; RIGOZO, N. R. N. D. J. R. E. A.. **O número de manchas solares, índice da atividade do sol**. Revista Brasileira de Ensino de Física, 25(2):157–163, 2003.