

Estimação de uma oscilação de baixa frequência em um sinal extraído de uma PMU

Naiara Maia dos Santos

Resumo—Este artigo, desenvolvido na disciplina de Probabilidade e Processos Estocásticos, investiga um evento real de oscilação de baixa frequência (LFO) medido por uma Unidade de Medição Fasorial (PMU). Partimos de uma linha de base probabilística tratando a frequência instantânea como v.a. i.i.d.; contudo, ACF/PACF revelam dependência temporal, o que motiva a modelagem como processo estocástico. Após análise espectral e separação da componente lenta, ajustamos um modelo ARIMA para a LFO, validamos por diagnóstico de resíduos e produzimos previsões. O repositório com o código que reproduz as análises está citado nas referências.

Index Terms—Palavras-chave: séries temporais; ARIMA; oscilação de baixa frequência (LFO); PMU; PSD; previsão

I. INTRODUÇÃO

Uma variável aleatória (v.a.) é uma função que atribui um número real a cada resultado de um experimento definido em um espaço amostral. Pelo Teorema do Limite Central (TLC), a soma ou a média de muitas v.a. *independentes e identicamente distribuídas* (i.i.d.), com média μ e variância $\sigma^2 < \infty$, aproxima uma distribuição Normal à medida que n cresce (convergência em distribuição). [1] [2].

O TLC é o resultado que explica por que, em muitos contextos, somas ou médias de muitas variáveis aleatórias independentes e identicamente distribuídas (i.i.d.), com média e variância finitas, apresentam comportamento aproximadamente normal. Em linguagem simples, ao agregar muitos efeitos pequenos e independentes, o formato da distribuição do agregado tende ao da curva gaussiana, o que permite aproximar probabilidades para somas e médias usando a normal. Essa é a razão de a normal “aparecer em toda parte”: fenômenos macroscópicos costumam resultar da adição de inúmeros processos microscópicos independentes (e, em aplicações de engenharia, médias e totais são onipresentes). Além disso, a formulação moderna do TLC é feita em termos de uma soma “padronizada” (isto é, centrada e escalada) cuja distribuição converge à normal padrão quando o número de termos cresce; na prática, isso justifica substituir a distribuição exata da soma por uma normal com a mesma média e variância para n suficientemente grande. Um ponto importante é a ampla generalidade do teorema: os somandos podem ter qualquer distribuição (contínua, discreta ou mista), desde que possuam média e variância finitas—daí a enorme utilidade do TLC como “ponte” entre modelos diversos e a normal [1] [2].

Em muitos experimentos aleatórios, o “resultado” não é um número isolado, mas uma função do tempo ou do espaço. É o caso de sistemas de reconhecimento de fala, em que

as decisões se baseiam numa forma de onda de tensão ao longo do tempo; de processamento de imagens, em que intensidade e cor variam sobre uma região do plano; de redes *peer-to-peer*, nas quais o número de pares ativos muda com o tempo; ou ainda de grandezas que evoluem em conjunto, como temperatura de uma cidade e demanda de energia elétrica local. Nesses cenários, falamos de grandezas numéricas que evoluem aleatoriamente: em vez de um único valor, observamos uma família de variáveis aleatórias indexada por tempo (ou espaço). [1] [2]

Contudo, nem todos os processos podem ser modelados como uma variável aleatória única. Quando aquilo que medimos “é uma v.a. que varia em função do tempo”, o objeto correto de estudo é um processo estocástico (ou processo aleatório). Modelar como uma única v.a. descreve apenas a distribuição marginal num instante, mas ignora o que é essencial em séries reais: dependência temporal/espacial, padrões oscilatórios, tendências e correlações entre instantes ou posições diferentes. Já o enquadramento como processo estocástico permite capturar essa estrutura (por exemplo, por meio de funções de correlação, conteúdo espectral ou modelos dinâmicos), oferecendo a base para análise, previsão e tomada de decisão em sistemas que mudam no tempo ou no espaço. [1] [2]

Portanto, artigo tem por objetivo utilizar uma oscilação de baixa frequência (LFO) em um sinal senoidal como estudo de caso para evidenciar que sinais que variam no tempo não podem ser modelados como uma simples variável aleatória (v.a.). Partimos do enquadramento dado pelo Teorema do Limite Central como linha de base — adequado para somas/médias de v.a.’s independentes — e, em seguida, mostramos por que é necessário tratar o sinal como processo aleatório. Para isso, modelamos a série temporal com modelos baseados em ARIMA amplamente empregados na previsão de séries temporais, principalmente para séries lineares. comparando-os ao baseline i.i.d. e discutindo ganhos de modelagem e previsão. [2] [3] [4] [5]

Este artigo está dividido como a seguir. Na Seção II são apresentados os materiais e métodos usados na modelagem do problema, o dataset escolhido bem como sua caracterização e a descrição do método escolhido. A apresentação e discussão dos resultados estão expostos na Seção III. Finalmente, a Seção IV traz as considerações finais do trabalho e trabalhos futuros.

II. MATERIAIS E MÉTODOS

Esta Seção é dividida em duas partes, apresentando os seguintes conteúdos: (A) descrição do *Dataset* e (B) análise do sinal e métodos estatísticos utilizados para processar os dados.

A. Descrição do Dataset

Os dados utilizados neste estudo foram obtidos no repositório *IEEE DataPort*, a partir do dataset intitulado *Phasor Measurement Data Recorded During Low Frequency Oscillation and Short Circuit Incidents in Actual Power Systems* [6]. Neste trabalho, selecionamos especificamente um distúrbio caracterizado por uma oscilação de baixa frequência (LFO) como estudo de caso, conforme ilustrado na Figura 1.

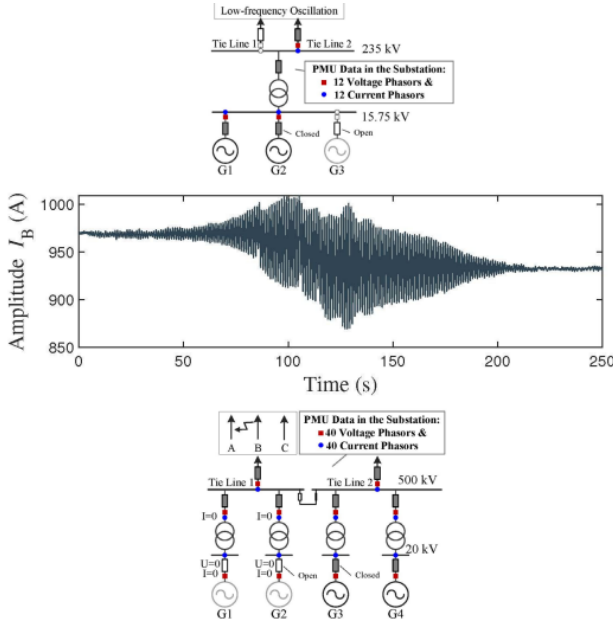


Figura 1: Oscilação de baixa frequência.

Constam dados de três PMUs (Unidade de Medição Fasorial, do inglês, *Phasor Measurement Unit*), e os dados escolhidos são da PMU1_I_2, onde usamos apenas uma das fases [6].

Uma PMU extrai, a partir dos sinais aplicados aos seus terminais, a magnitude, o ângulo de fase, a frequência e a derivada da frequência. Contudo, esses sinais frequentemente chegam contaminados por harmônicos, ruídos e alterações de estado provocadas por variações de carga ou por ações de controle e proteção, o que dificulta a estimação precisa desses parâmetros. [7]

B. Modelo ARIMA: construção, identificação e casos úteis

O **ARIMA** (*Autorregressivo Integrado de Médias Móveis*, do inglês *AutoRegressive Integrated Moving Average*) é uma família de modelos para séries temporais que combina três componentes: (i) parte autorregressiva (AR) de ordem p , (ii) diferenciações d para lidar com não estacionaridade, e (iii) médias móveis (MA) de ordem q . Usaremos a notação $ARIMA(p, d, q)$ e seguiremos o procedimento de Box–Jenkins (identificação, estimação e diagnóstico). [4] [3]

Estacionaridade e diferenciação: A prática usual é aplicar transformações (p.ex., log) e diferenças até obter uma série aproximadamente estacionária; sobre a série diferenciada ajusta-se um ARMA. Em geral, $d \in \{0, 1\}$ (raramente 2), pois sobrediferenciar inflaciona a variância e pode introduzir dependências artificiais. [3] [4]

ACF e PACF: Para um processo estacionário $\{X_t\}$, a autocovariância e a autocorrelação (ACF) são

$$\gamma(h) = \text{Cov}(X_t, X_{t-h}), \quad \rho(h) = \frac{\gamma(h)}{\gamma(0)}.$$

A *autocorrelação parcial* no defasamento h (PACF), denotada $\alpha(h)$, é definida por

$$\alpha(0) = 1, \quad \alpha(h) = \phi_{h,h} \quad (h \geq 1),$$

onde $\phi_{h,h}$ é o *último coeficiente* do modelo $AR(h)$ que resolve o sistema de Yule–Walker $\Gamma_h \phi_h = \gamma_h$, com

$$\Gamma_h = [\gamma(i-j)]_{i,j=1}^h \quad \text{e} \quad \gamma_h = (\gamma(1), \gamma(2), \dots, \gamma(h))^T.$$

Intuitivamente, $\alpha(h)$ mede a correlação entre X_t e X_{t-h} depois de remover o efeito linear dos valores intermediários $X_{t-1}, \dots, X_{t-h+1}$. [8]

Heurísticas de identificação (ACF/PACF): Usaremos “*corte*” para indicar que os coeficientes ficam essencialmente nulos após certo defasamento. [4]

- **AR(p):** a PACF apresenta *corte* em p (para $h > p$, $\alpha(h) \approx 0$); a ACF decai gradualmente (tipicamente de forma exponencial ou como onda amortecida).
- **MA(q):** a ACF apresenta *corte* em q ; a PACF decai gradualmente.
- **ARMA(p, q):** ACF e PACF decaem gradualmente (não há cortes nítidos).
- **Não estacionária:** ACF com *cauda longa* (decaimento muito lento) sugere aplicar diferenciação.

Identificação, estimação e diagnóstico: Sugere-se d (e D) por inspeção da ACF/testes de diferença; p, q (e P, Q) por padrões ACF/PACF e por critérios de informação (AIC/AICc/BIC). A estimação é por máxima verossimilhança sob *estacionaridade* (raízes de $\phi(B)$ fora do círculo unitário) e *invertibilidade* (raízes de $\theta(B)$ fora do círculo unitário). O diagnóstico verifica se os resíduos se comportam como ruído branco (ACF residual dentro das bandas; teste de Ljung–Box com p -valores altos). As previsões k -passos incluem bandas a partir da variância de previsão do modelo ajustado. [4] [3]

III. IMPLEMENTAÇÃO E DISCUSSÃO DOS RESULTADOS

A. Abordagem de modelagem: da variável aleatória ao processo

Extração da frequência instantânea (pré-processamento):

A partir do fasor complexo medido pela PMU, extraímos a fase desembrulhada e estimamos a *frequência instantânea* por derivação temporal, $f_{\text{inst}}(t) \approx \dot{\phi}(t)/(2\pi)$, em conformidade com a padronização IEC/IEEE 60255-118-1 para medições sincrofatorias (frequência e ROCOF) [7]. Para evitar transientes fora do evento, analisamos a janela específica de ocorrência (80–210 s) [9] e centralizamos f_{inst} (removendo o nível médio).

Hipótese nula: f_{inst} como variável aleatória i.i.d.:

Como *linha de base*, assumimos f_{inst} como uma v.a. i.i.d. À luz do TLC, essa hipótese justifica comparar a distribuição marginal da série com a Normal [1], [2]. Assim, construímos o *histograma* de f_{inst} e sobreposamos a PDF Normal com mesma média e variância, além do $Q-Q$ plot para avaliar

gaussianidade marginal. Essa etapa usa rotinas consolidadas de *NumPy*, *SciPy* e *Matplotlib*. Veja as figuras 2 e 3 a seguir:

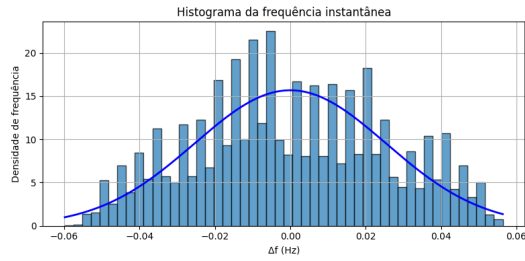


Figura 2: Histograma da LFO

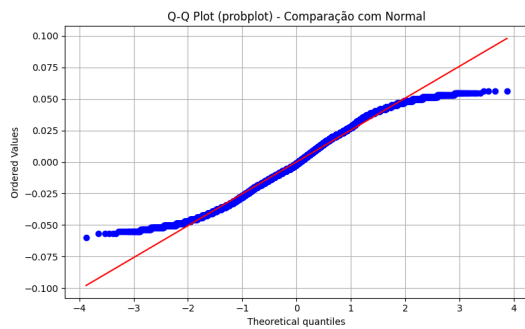


Figura 3: Q-Q Plot

Apesar de o histograma e a curva Normal ajustada sugerirem boa aderência na região central da distribuição, o *Q-Q plot* evidencia **caudas mais pesadas** do que as previstas pela Normal. Ademais, as funções de autocorrelação (ACF, Fig. 4) e autocorrelação parcial (PACF, Fig. 5) revelam **dependência seriada** expressiva: a ACF decai lentamente e permanece acima das bandas de confiança por diversos defasamentos, enquanto a PACF apresenta picos significativos nos primeiros lags. Em conjunto, esses achados refutam a hipótese de amostras i.i.d. para f_{inst} (isto é, não basta modelá-la como uma única variável aleatória) e motivam o tratamento como **processo estocástico**, a ser modelado adiante no arcabouço ARIMA.

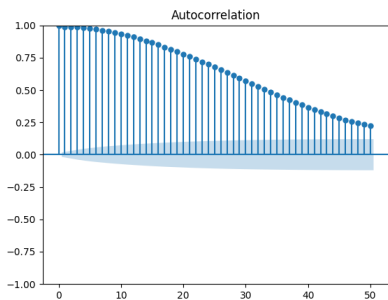


Figura 4: Autocorrelação

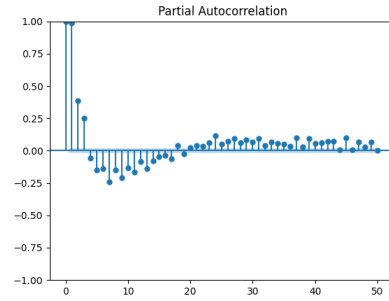


Figura 5: Autocorrelação Parcial

Espectro e separação LFO/rápida: Para compreender a estrutura em frequência, reduzimos a taxa de amostragem por *decimação* com filtragem de fase zero, diminuindo a variância do estimador espectral sem introduzir atraso. Em seguida, estimamos a PSD por Welch no sinal decimado e adotamos 1 Hz como linha de separação operacional entre a componente *lenta* (LFO) e a *rápida*. A decomposição é feita com um passa-baixas *Butterworth* de 4ª ordem (< 1 Hz) e filtragem zero-fase, porque o Butterworth oferece resposta monotônica na banda passante e transição suave, minimizando ondulações espúrias na componente lenta [10]. Observe as figuras 6 e 7.

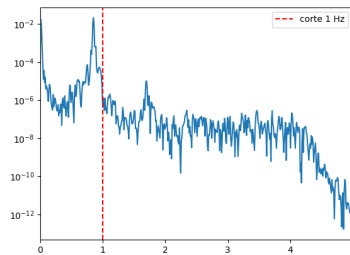


Figura 6: Densidade Espectral de Potência

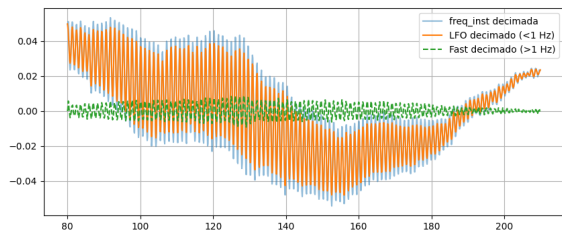


Figura 7: Decimação

Modelagem da LFO como ARIMA e diagnóstico.: Sobre a LFO, o teste ADF indicou não-estacionaridade (ADF = $-1,102$, $p = 0,714$), justificando usar $d = 1$. Ajustamos um ARIMA(2, d , 1), baseado no ACF/PACF. 8 e 9

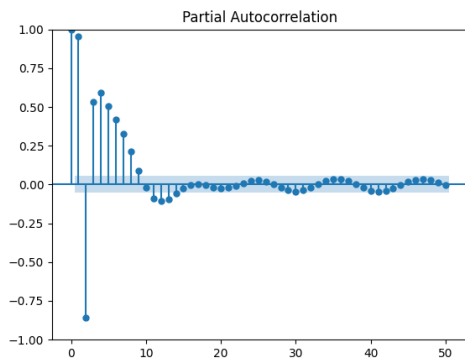


Figura 8: ACF LFO

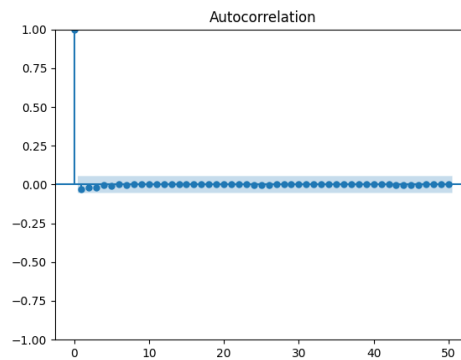


Figura 11: PAC dos Resíduos LFO

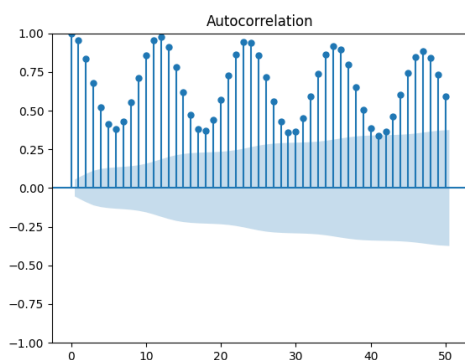


Figura 9: PAC LFO

Nas figuras, é possível ver p tendo dois lags fora da marcação e q não há um corte exato, então foram feitos testes com $q=0$ e $q=1$.

O diagnóstico de resíduos foi compatível com *ruído branco* (Ljung–Box $p = 0,991$), e a ACF/PACF residual não exibiu estrutura remanescente. Veja as figuras 10 e 11

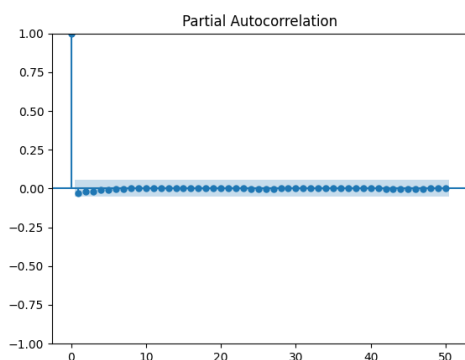


Figura 10: ACF dos Resíduos da LFO

Tratamento da componente rápida como ruído: Após a separação espectral (corte em 1 Hz) com filtro de Butterworth de 4ª ordem e filtragem *zero-phase* [10], a componente acima de 1 Hz apresentou amplitude máxima da ordem de 10^{-6} , substancialmente menor do que a variabilidade do componente LFO e do objetivo do estudo. Por essa razão, consideramos essa parcela como **ruído** operacional e a utilizamos apenas para verificação de pré-branqueamento (ACF residual pequena e teste de Ljung–Box sem evidência de autocorrelação). O foco da modelagem permaneceu no componente *lento*, coerente com a caracterização de LFO em sistemas de potência.

Previsão da LFO: A partir do ARIMA ajustado à componente lenta (< 1 Hz) da série decimada, geramos previsões multipasso (1000 passos) com `forecast`, sobrepondo-as ao histórico para evidenciar a continuidade da dinâmica lenta. A figura 12 mostra a curva observada e a trajetória prevista. Se compormos com a 1 vemos que se assemelha bastante ao comportamento real.

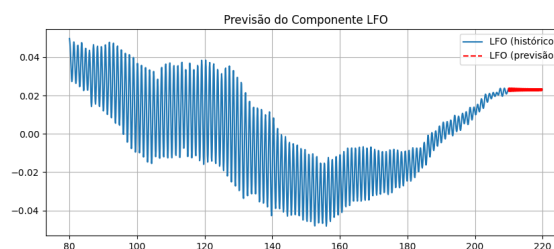


Figura 12: Previsão da Componente LFO

Simulação e avaliação preditiva.: Por fim, geramos uma trajetória sintética ancorada no fim da amostra a partir do ARIMA ajustado e comparamos com a LFO observada; num *hold-out* de 20%, obtivemos $MAE = 0,0079$ e $RMSE = 0,0098$, valores compatíveis com o nível de variabilidade da série, vide 13.

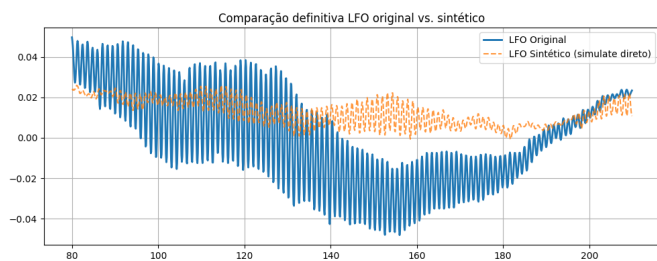


Figura 13: Geração de Um Sinal Sintético

Todo o código e os scripts que reproduzem as análises estão disponíveis no repositório [11].

IV. CONCLUSÕES

Os resultados indicam que a LFO foi adequadamente modelada por um ARIMA(2,1,1), capaz de capturar a dinâmica lenta do sinal. A análise dos resíduos pelo teste de Ljung–Box ($p > 0,05$) aponta ausência de autocorrelação remanescente, sugerindo especificação satisfatória do modelo. As previsões exibiram reversão suave à média de longo prazo, compatível com um comportamento estacionário no horizonte de previsão, enquanto os experimentos com dados sintéticos apresentaram MSE e RMSE baixos, evidenciando boa fidelidade na leitura do processo.

Como próximos passos, propomos a inclusão explícita de sazonalidade (SARIMA) de modo a preservar os ciclos característicos da LFO no *forecast*. O período sazonal pode ser definido a partir do pico dominante observado na ACF/periodograma da série, garantindo que o componente oscilatório seja mantido sem viés de fase. Além disso, recomendamos comparar a extração da LFO via filtro Butterworth e via decomposição STL, de forma a refinar a separação de componentes: enquanto o Butterworth oferece controle claro de banda (à custa de atraso de grupo), a STL é robusta a tendências e a *outliers* e preserva padrões locais. Essa comparação deve ser quantificada por métricas espectrais (energia do modo retida), coerência de fase e erro de previsão fora da amostra. Em conjunto, esses aprimoramentos tendem a produzir previsões mais fiéis e interpretáveis para o fenômeno oscilatório de interesse.

REFERÊNCIAS

- [1] C. A. Ynoguti, “Probabilidade, estatística e processos estocásticos,” Apostila T501, São Paulo, 2012, material de curso.
- [2] A. Leon-García, *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd ed. Upper Saddle River: Pearson Prentice Hall, 2008.
- [3] R. J. Hyndman and G. Athanasopoulos, “Forecasting: principles and practice,” <https://otexts.com/fpp2/>, Melbourne, 2018, acesso em: 24 ago. 2025.
- [4] Asian Institute of Management, “Time series analysis handbook,” https://phdinds-aim.github.io/time_series_handbook/Preface/Preface.html, 2020, por estudantes do PhD in Data Science (Batch 2023). Acesso em: 24 ago. 2025.
- [5] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, “A review of arima vs. machine learning approaches for time series forecasting in data driven networks,” *Future Internet*, vol. 15, 2023, acesso em: 25 ago. 2025. [Online]. Available: <https://doi.org/10.3390/fi15080255>

- [6] F. Zhang, “Phasor measurement data recorded during low frequency oscillation and short circuit incidents in actual power systems,” IEEE DataPort, Piscataway, NJ, USA, may 2020, acesso em: 25 ago. 2025. [Online]. Available: <https://doi.org/10.21227/1x22-r651>
- [7] International Electrotechnical Commission and Institute of Electrical and Electronics Engineers, “Iec/ieee 60255-118-1:2018 — measuring relays and protection equipment: Part 118-1: Synchrophasor for power systems — measurements,” IEC/IEEE Standard, Geneva; New York, 2018, acesso em: 25 ago. 2025. [Online]. Available: <https://standards.ieee.org/standard/60255-118-1-2018.html>
- [8] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 3rd ed. New York: Springer, 2016.
- [9] F. Zhang, X. Wang, Y. Yan, J. He, W. Gao, and G. Chen, “A synchrophasor data compression technique with iteration-enhanced phasor principal component analysis,” *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 2365–2376, may 2021, acesso em: 25 ago. 2025. [Online]. Available: <https://doi.org/10.1109/TSG.2020.3046666>
- [10] S. K. Mitra, *Digital Signal Processing: A Computer-Based Approach*, 4th ed. New York: McGraw-Hill, 2011.
- [11] N. M. Santos, “Trabalho_final_processos_estocasticos,” GitHub repository, 2025, accessed: Aug. 26, 2025. [Online]. Available: https://github.com/Naiara-Maia-Eng/Trabalho_Final_Processos_Estocasticos



Naiara Maia dos Santos é formada em Engenharia Elétrica pela UFJF em 2025. Ela é membro do Grupo de Pesquisa NIPS(Núcleo de Instrumentação e Processamento de Sinais). No período entre 2022 a 2025 fez parte do grupo devido a sua participação na iniciação científica. Atualmente, ele é um estudadnte de Mestrado em engenharia elétrica na UFJF, na área de Sistemas Eletrônicos.