

Universidade Federal do Paraná

Stephani Fernanda Soares Farinelli

Projeção da Energia Natural Afluente utilizando Séries Temporais

Curitiba

2023

Stephani Fernanda Soares Farinelli

Projeção da Energia Natural Afluente utilizando Séries Temporais

Trabalho de Conclusão de Curso apresentado à disciplina Laboratório B do Curso de Graduação em Estatística da Universidade Federal do Paraná, como exigência parcial para obtenção do grau de Bacharel em Estatística.

Orientador: Prof. Fernando Lucambio Pérez

Curitiba
2023

Agradecimentos

Primeiramente agradeço a Deus por ter me permitido chegar até aqui, por me ajudar a vencer todos os obstáculos encontrados ao longo do curso e por me dar forças para continuar.

Aos meus pais e irmã, por acreditarem em mim quando eu mesma não acreditei, em especial à minha querida mãe Fabiane, agradeço por todas as orações feitas em dias de provas, ou em dias em que eu estava triste e ansiosa vivendo a pressão da faculdade.

À Nayara que foi "*my person*" durante o curso, a amiga que a faculdade trouxe e persistirá para a vida. Foi quem dividiu comigo todas as alegrias e tristezas possíveis da fase universitária. Obrigada por tornar tudo mais leve.

À Luana e demais amigos que ao longo desses anos me acompanharam. De forma especial à minha irmã de coração, Carol, que por muitas vezes me escutou e aconselhou, sempre dando suporte emocional nos momentos difíceis.

À Electra Energy pelo suporte para que fosse possível me dedicar à faculdade quando necessário e uma menção especial ao Leonardo Salvi, dono de uma mente brilhante, que é uma das minhas inspirações para concluir esta graduação e por eu me interessar tanto pelo setor Elétrico Brasileiro.

Aos meus amigos da Gerência de Inteligencia de Mercado (GIM), que sempre me apoiaram e me ajudaram em momentos de surtos ou sofriam junto comigo, impossível esquecer esse setor.

Ao Professor Lucambio, pela excelente orientação, conselhos e paciência, por me incentivar em dias difíceis (e foram vários) a não desistir, sem ele eu não chegaria até aqui.

Por último, mas não menos importante, ao Professor Paulo Guimarães por aceitar fazer parte da banca avaliadora contribuindo com todo o seu conhecimento na área acadêmica.

Stephani Fernanda Soares Farinelli

"The carousel never stops turning. You can't get off."
(Série Grey's Anatomy, Temporada 2, Episódio 4)

Resumo

Neste trabalho é desenvolvida uma metodologia para a previsão da Energia Natural Afluente (ENA) que é um dos principais insumos para o cálculo do Preço de Liquidação das Diferenças (PLD) da energia elétrica, utilizando os dados históricos disponibilizados pelo Operador Nacional do Sistema Elétrico (ONS) e pela Câmara de Comercialização de Energia Elétrica (CCEE) através de séries temporais.

Essas técnicas simulam a ENA do Sistema Interligado Nacional (SIN) observando as séries históricas, para que seja possível fazer projeções. Também abordada-se a análise das séries históricas dos fenômenos meteorológicos *El Niño* e *La Niña* e os eventos de *ZCAS* afim de estudar as correlações que possam influenciar na ENA.

Este estudo trará as projeções de ENA para os próximos 12 meses subsequentes ao histórico abordado, através das técnicas de séries temporais, especificamente o modelo SARIMA que obteve o melhor resultado comparado com o modelo SSA.

Estes resultados são de grande importância para os estudos de planejamento da operação e de determinação do Preço de Liquidação das Diferenças adotado para valorar as exposições dos agentes no Mercado de Curto Prazo de energia e consequentemente o principal balizador em estudos de risco e estratégia de mercado.

Palavras-chave: Preço de Liquidação das Diferenças; Energia Natural Afluente; Comercialização de Energia; Previsão do Preço da Energia; Séries Temporais.

Lista de ilustrações

Figura 1 – Diagrama dos Segmentos de um Sistema de Energia Elétrica	13
Figura 2 – Índice Oceânico e a classificação de anos de ocorrência do El Niño (fonte na cor vermelha), La Niña (fonte na cor azul) e anos de neutralidade (fonte na cor preta), para o período de 1950 a 2022.	18
Figura 3 – Média Climatológica de Precipitação para Dezembro	20
Figura 4 – Submercados de Energia Brasileira.	23
Figura 5 – Dados de ENA da Região Sudeste/Centro-Oeste.	25
Figura 6 – Séries Históricas ENA para os Quatro Submercados.	26
Figura 7 – Série Histórica do Sudeste/Centro-Oeste.	28
Figura 8 – Séries Históricas de ENA (escala logarítmica), Eventos Meteorológicos e ZCAS.	28
Figura 9 – Ajuste através do modelo SARIMA (escala logarítmica).	33
Figura 10 – Série cortada do ajuste com o modelo SARIMA (escala logarítmica).	34
Figura 11 – Reconstrução da Séie através do SSA (escala logarítmica).	37
Figura 12 – Série cortada do ajuste com o modelo SSA (escala logarítmica)	38
Figura 13 – Modelo estimado com as bandas de confiança de 95% (escala logarítmica).	43
Figura 14 – Modelo estimado com as bandas de confiança de 95% (escala real).	44

Lista de tabelas

Tabela 1	–	CrITÉrios para a classificaÇ�o da intensidade do fen�meno ENOS.	19
Tabela 2	–	Previs�es dos Modelos.	42
Tabela 3	–	M�tricas de avalia��o para ambos os modelos (escala logar�tmica)	42
Tabela 4	–	Previs�o do modelo SARIMA com as bandas de confian�a.	45

Lista de abreviaturas e siglas

ACL	Ambiente de Contratação Livre
ACR	Ambiente de Contratação Regulada
AIC	Critério de informação de Akaike
Aneel	Agência Nacional de Energia Elétrica
AR	Auto Regressivo
ARIMA	Autoregressive Integrated Moving Averag
ARMA	Autoregressive Moving Average
CCEE	Câmara de Comercialização de Energia Elétrica
CEPEL	Centro de Pesquisas de Energia Elétrica
CMO	Custo Marginal de Operação
CVU	Custo Variável Unitário
ENA	Energia Natural Afluenta
ENOS	El Niño Oscilação Sul
FCF	Função de Custo Futuro
MA	Médias moveis
MCP	Mercado de Curto prazo
mm	Milímetros
MWh	Megawatt-hora
MWmed	Megawatt-médio
NOAA	National Weather Service Climate Prediction Center
NOS	Operador Nacional do Sistema Elétrico
ONS	Operador Nacional do Sistema Elétrico
PLD	Preço de Liquidação das Diferenças
PMO	Programa Mensal da Operação

QBOM	Vazão bombeado
QDEF	Vazão Defluente
QDES	Vazão Desviada
QTUR	Vazão Turbinada
QVER	Vazão Vertida
RDH	Relatório Diário da Situação Hidráulico-Hidrológica das Usinas Hidrelétricas do SIN
RED	Recursos Energéticos Distribuídos
REE	Reservatório Equivalente de Energia
RHA	Restrições de vazão afluyente
RHQ	Restrições de vazão afluyente
RHV	Volume Armazenado
S	Sazonal
SARIMA	Seasonal Autoregressive Integrated Moving Averag
SE/CO	Sudeste/Centro-Oeste (submercado)
SEE	Sistema de Energia Elétrica
SIN	SIN Sistema Interligado Nacional
SSA	Análise de Espetro Singular
ZCAS	Zona de Convergência do Atlântico Sul

Sumário

1	INTRODUÇÃO	10
1.1	Contextualização do Tema	10
1.2	Organização do Trabalho	11
2	FUNDAMENTAÇÃO TEÓRICA	12
2.1	Setor Elétrico Brasileiro	12
2.2	Matriz Elétrica	14
2.3	Usinas Hidrelétricas	15
2.4	Restrições Hidráulicas	15
2.4.1	Volume de Espera	16
2.4.2	Evaporação	16
2.5	Fenômenos	17
2.6	Zona de Convergência do Atlântico Sul (ZCAS)	19
2.7	Regime Anual de Chuva no Brasil	21
2.8	Energia Natural Afluente	21
2.9	Preço de Liquidação das Diferenças (PLD)	22
2.10	Comercialização de Energia	23
3	MATERIAIS E MÉTODOS	25
3.1	Material	25
3.1.1	Origem dos Dados	25
3.1.2	Banco de Dados	25
3.1.3	Recursos Computacionais	26
3.2	Métodos	26
3.2.1	Séries Históricas	26
3.2.2	Séries Temporais	27
3.2.3	Decomposição de Séries Temporais	29
3.2.4	Modelo <i>Autoregressive Integrated Moving Average</i> (ARIMA)	30
3.2.5	Análise Espectral Singular (SSA)	34
3.2.6	Métricas de Avaliação	38
4	RESULTADOS E DISCUSSÕES	42
5	CONSIDERAÇÕES FINAIS	46
	REFERÊNCIAS	47

1 Introdução

1.1 Contextualização do Tema

A energia elétrica é um dos insumos essenciais no mundo atual, logo, o mercado de energia cresce cada dia mais. Assim, tornam-se indispensáveis para os agentes de mercado ferramentas computacionais eficientes que permitam obter previsões viáveis e rigorosas da Energia Natural Afluente (ENA), do Custo Marginal da Operação (CMO), de Geração e outras variáveis que compõem o preço da Energia Elétrica do Mercado Livre. As previsões possibilitam desenvolver melhores estratégias no mercado energético, maximizando o lucro tendo em vista a compra e venda de energia otimizando a exploração dos recursos. O propósito deste trabalho é analisar o comportamento da ENA juntamente com os fenômenos e ZCAS.

O PLD é o equivalente ao preço *spot* do setor elétrico, sendo responsável por valorar toda energia gerada e consumida na etapa de contabilização da Câmara de Comercialização de Energia Elétrica (CCEE), resultando em valores a pagar ou a receber para cada agente, seja gerador ou consumidor. Esses valores de PLD são publicados diariamente pela CCEE em base horária e é calculado um dia antes o valor do preço horário que será praticado no dia seguinte (denotado por D+1). Seu cálculo consiste, basicamente, em reprocessar o CMO com aplicação de limites máximos e mínimos horários e um limite máximo em relação à média de variação diária, sendo os limites máximos e mínimos calculados anualmente. Como o PLD é um subproduto do CMO e este, por sua vez, depende do cálculo dos modelos computacionais com íntima relação hidráulica, a influência meteorológica na formação de preços é muito bem correlacionada com o clima e, portanto, com sazonalidade bem definida e previsível (Câmara de Comercialização de Energia Elétrica (CCEE) , 2023).

Já a ENA bruta representa a energia produzível pela usina, calculada pelo produto das vazões naturais aos reservatórios com as produtividades a 65% dos volumes úteis. A ENA armazenável considera as vazões naturais descontadas das vazões vertidas nos reservatórios.

Será então, apresentada uma metodologia para a previsão dos preços da energia elétrica, utilizando as séries históricas disponibilizadas pelo Operador Nacional do Sistema Elétrico (ONS), que é separado por submercados (Sudeste/Centro-Oeste, Sul, Nordeste e Norte). Primeiramente observando as séries históricas, para que seja possível fazer projeções e em seguida analisar através de séries temporais afim de estudar as influências que possam direcionar o provisionamento do PLD.

1.2 Organização do Trabalho

Após essa breve introdução no Capítulo 1, a estruturação deste trabalho é dividida em quatro partes. O Capítulo 2 apresenta uma fundamentação histórica do mercado de energia no Brasil, abordando as principais características que venham influenciar no presente trabalho. No Capítulo 3, encontram-se os detalhes a respeito dos dados obtidos, bem como os recursos computacionais e métodos estatísticos utilizados. Os resultados da análise exploratória e da previsão da série temporal da Energia Natural Afluente, são apresentados e discutidos no Capítulo 4. Por fim, no Capítulo 5 é trazida a discussão sobre os objetivos propostos, suas realizações e também são sugeridas ideias para possíveis pesquisas futuras.

2 Fundamentação Teórica

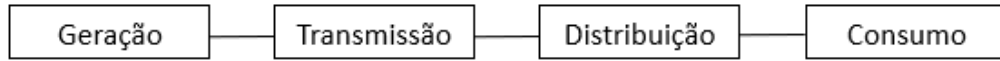
2.1 Setor Elétrico Brasileiro

O Sistema de Energia Elétrica (SEE) é hoje um dos principais e mais complexos setores no mundo, devido a sua capacidade de lidar com a matriz energética formada em sua maioria por hidrelétricas localizadas em diversas bacias hidrográficas com diferentes características climáticas, fontes altamente intermitentes e sazonais, com diferentes tipos de tecnologia e combustíveis.

Por muito tempo o setor elétrico foi composto em sua maioria por empresas estatais. Entretanto, a abertura do setor para o capital privado, aconteceu na primeira grande reforma do setor (1995 – 1998) (BNDES, DEZ. 1999), visando a necessidade de uma expansão e melhorias da rede elétrica, aliada ao momento de grandes investimentos estrangeiros no país. A partir disso, esse mercado foi separado em dois: o Ambiente de Contratação Livre (ACL) onde os agentes podem negociar livremente as condições de preço, quantidade, prazo e tipo da energia contratada; e o Ambiente de Contratação Regulada (ACR) das grandes distribuidoras que compram energia por meio de leilões públicos, com preços e quantidades definidos pela Agência Nacional de Energia Elétrica (ANEEL), agência responsável pela regulação do setor. Com essas reformas, sentiu-se a necessidade de órgãos para auxiliar o mercado que estava surgindo. Então, houve a criação dos agentes considerados fundamentais para a estruturação do mercado livre de energia: o (Operador Nacional do Sistema Elétrico, 2023b) (ONS) - um órgão responsável tecnicamente pelo sistema elétrico, desde o planejamento até a operação em tempo real; e a Câmara de Comercialização de Energia Elétrica CCEE (2023), atual responsável pelas contabilizações do mercado de energia elétrica no Brasil.

Devido a forte regulação e aos agentes reguladores que regem o mercado, o modelo vigente é um sucesso composto por empresas estatais e privadas em todos os segmentos que estão relacionados a geração, transmissão e distribuição, para que cheguem na etapa final, ou seja, o consumo conforme mostra a figura abaixo.

Figura 1 – Diagrama dos Segmentos de um Sistema de Energia Elétrica



Fonte: A autora (2023).

As usinas hidrelétricas que utilizam o potencial energético presente em rios e quedas d'água, juntamente com as usinas termelétricas que produzem energia por meio da queima de combustíveis, são as principais fontes de eletricidade em nível nacional e mundial.

Com o desenvolvimento da tecnologia e a busca pela sustentabilidade, foram criadas as novas fontes de energia, como a energia eólica, onde o vento movimentava pás de turbinas associadas a geradores, e a energia solar fotovoltaica, onde a radiação solar é convertida em energia elétrica.

A Transmissão é responsável por transferir grandes quantidades de energia vinculando a geração às distribuidoras que atendem os consumidores. A grande quantidade de energia transmitida e as maiores distâncias percorridas pelas linhas de transmissão são os diferenciais desse segmento em relação ao sistema de distribuição, já que a transmissão é responsável por interligar grandes centrais hidrelétricas, geralmente mais distantes dos centros urbanos, onde há maior consumo de energia. A transmissão também é responsável pela interligação entre grandes regiões do sistema, permitindo a troca de energia (intercâmbio) entre diferentes regiões.

Esse intercâmbio entre regiões pode reduzir os custos operacionais, aproveitando as condições hidrológicas favoráveis em uma região do sistema para reduzir os custos devido à menor disponibilidade de energia hidrelétrica em outra região, evitando custos de ativação de usinas termelétricas, por exemplo.

Essa transmissão de energia a longa distância resulta em uma perda parcial de energia devido ao efeito Joule¹. Para reduzir essas perdas, a transmissão de energia é realizada em altas tensões. Assim, a fronteira entre transmissão e distribuição também está definida: de acordo com a Aneel n.º 067/2004 (ANEEL, 2023), a transmissão inclui suas linhas, barramentos e transformadores com tensão igual ou superior a 230 kV.

¹ Joule: A perda por efeito joule ocorre quando há a passagem de uma corrente elétrica através de um condutor, com a conversão de energia elétrica em calor, sendo esta perda proporcional ao quadrado da corrente. Ao elevar-se a tensão na transmissão, para uma mesma potência, a corrente é reduzida.

O setor de distribuição de energia elétrica é responsável pelo fornecimento de energia aos consumidores finais. Isso inclui o processo de redução da tensão de transmissão e conexão aos consumidores. No Brasil, as distribuidoras também são responsáveis pela contratação de energia dos chamados consumidores cativos, repassando o custo de compra de energia a esses consumidores.

Por fim, a área do consumidor é o destino da energia elétrica no sistema. Nela, a energia elétrica é convertida em trabalho para iluminação, aquecimento, partida de equipamentos e máquinas elétricas e outras aplicações que requerem eletricidade.

Em um aspecto geral do sistema de energia, os consumidores recebem energia elétrica desde a geração até a distribuição, transmissão e consumo. Portanto, todos os consumidores dependem da conexão elétrica do sistema de distribuição.

Ressalte-se que o modelo regulatório brasileiro está em constante mudança, com a inclusão de novas fontes de energia no sistema elétrico, principalmente por meio da inserção dos chamados Recursos Energéticos Distribuídos (RED). Esses recursos estão relacionados à geração de pequena escala conectada ao sistema por meio da rede de distribuição. Além disso, a instalação da geração pelo consumidor, com foco na energia solar fotovoltaica, foi impulsionada pelos avanços tecnológicos e pela adoção de incentivos fiscais, dentre outras mudanças específicas que continuam a ocorrer para atender problemas situacionais.

2.2 Matriz Elétrica

Conforme apresentado, o SIN é um sistema de grande porte que depende da disponibilidade de fontes energéticas primárias. A matriz energética é composta pelas principais fontes de energia, são elas: eólica (13,41%); térmica (13,41%); solar (4,94%); biomassa (9,87%); nuclear (0,94 %) e com destaque para as capacidades instaladas hidrelétricas (51,89%), que totalizam 99,9% da capacidade instalada total no SIN, a partir dos dados disponibilizados pelo Operador Nacional do Sistema Elétrico (2023b) para o horizonte 2023-2027, tendo como referência o planejamento energético de junho de 2023.

Compreender a dependência das hidrelétricas para a geração de energia no SIN é essencial, sendo assim a operação interligada do sistema torna-se um desafio em diversos aspectos, sendo necessário o estabelecimento de uma política de planejamento que otimize o uso dos recursos energéticos disponíveis.

Aqui no Brasil, entende-se como necessária a figura do operador central, onde é realizada tal otimização que deve levar em conta minimizar o custo esperado total da operação, atendendo as restrições operativas e critérios de segurança, visando garantir o suprimento de energia elétrica no sistema, caso contrário, cada agente poderia operar em benefício próprio.

2.3 Usinas Hidrelétricas

Usinas hidrelétricas aproveitam o potencial energético de córregos que contenham queda d'água para gerar eletricidade.

Quando a água colide com a turbina, o grande volume de água empurra a turbina e o eixo é conectado a um gerador, que converte energia mecânica em energia elétrica. As usinas hidrelétricas podem ser divididas em usinas de reservatório que podem armazenar ou gerar energia de acordo com a necessidade do sistema, afetando a vazão de trechos de rios associados a usinas específicas e as usinas a fio d'água que são construídas e operadas de forma que o fluxo natural seja mantido, reduzindo o impacto da usina nos cursos d'água, não permitindo o armazenamento de grandes quantidades de água. Além disso cada usina hidrelétrica possui um reservatório, que afeta os cursos dos rios existentes de forma diferenciada conforme à sua capacidade.

A capacidade de armazenamento de energia é fundamental para o planejamento da operação para garantir o atendimento da demanda. A regularização está associada ao período de tempo onde um determinado reservatório pode ser esvaziado e reabastecido, podendo ser mensal, semanal ou diário.

Um aspecto que faz com que a geração hidrelétrica seja vantajosa para a operação do sistema, é que esta geração não cria custo adicional, no sentido de que não se paga mais para gerar mais energia como ocorre em uma usina térmica, onde a geração requer a queima de um combustível, que por sua vez tem custo. Em contrapartida, requerem a disponibilidade de um curso d'água com vazão considerável, uma vez que a energia gerada é proporcional a quantidade de água turbinada. Limitando as áreas nas quais a construção de uma usina hidrelétrica é viável.

2.4 Restrições Hidráulicas

As Restrições Hidráulicas permitem simular as restrições de operação de uma usina ou grupo de usinas hidrelétricas, estabelecendo limites máximos e mínimos para vazões defluentes ou afluentes, faixas de escoamento para um grupo de usinas e faixas de operação para reservatórios. Essas restrições devem ser incorporadas na simulação para representar operações de bacias hidrográficas que contém regras operacionais por questões ambientais, de cursos d'água ou de energia (LACERDA, 2021).

A restrições de vazão afluente (RHA) determinam os limites inferior e superior, expressos em metros cúbicos por segundo (m^3/s), da quantidade de água que chega a uma determinada usina hidrelétrica durante uma determinada etapa de simulação. A k -ésima restrição de vazão afluente associada a uma usina é dada por:

$$LI_k \leq Q_{inc,i} + \sum_{j \in M_i} Q_{defj} + \sum_{j \in D_i} Q_{desvj} \leq LS_k$$

onde:

$Q_{inc,i}$ = Vazão Incremental a usina i;

M_i = Conjunto de usinas a montante de i ;

Q_{defj} = Vazão defluente da usina da usina j;

D_i = Conjunto de usinas com canal de desvio d'água para i ;

Q_{desvj} = Vazão desviada da usina j.

As Restrições de vazão defluente (RHQ) definem o limite operacional defluente da usina e pode ser expressa em função das variáveis de vazão defluente (QDEF), turbinada (QTUR), vertida (QVER), desviada (QDES) e bombeada (QBOM) nas usinas. Por exemplo, por motivos estruturais, existem valores operacionais máximos e mínimos para a quantidade de água que pode ser despejada por um vertedouro. Para cada variável, a expressão geral é semelhantes à expressão das restrições elétricas dadas na equação acima. RHQ é dado por fase e limiar, os limites são em m^3/s .

As restrições de volume armazenado (RHV), limitado em hecto-metros cúbico (hm^3) e definido ao final de cada etapa da simulação, incluindo restrições de volume mínimo e máximo de processamento em usinas com reservatórios. Além de manter a manutenção dos níveis de armazenamento pelos motivos acima, esses limites podem ser usados para simular a operação do reservatório conforme a curva de referência.

2.4.1 Volume de Espera

O volume de água a ser retido define o percentual máximo do volume de espaço utilizável que o modelo pode usar para armazenar energia, sendo que a quantidade armazenada depende da tomada de decisão do modelo, sujeita a outros requisitos. Exemplos de processos, como defluência e armazenamentos (que possuem acomodação espacial e temporal), o processo ótimo pode levar à violação de uma restrição importante para atender o volume de espera. Com isso, o operador pode optar por relaxar a exigência de volume de modo que torne a operação simulada viável.

2.4.2 Evaporação

No processo de otimização, o modelo leva em consideração a taxa de evaporação do reservatório. O coeficiente de evaporação de cada reservatório é determinado mensalmente de acordo com a superfície da água e a quantidade de água evaporada. Assim, a taxa de evaporação oscila de acordo com as decisões do modelo, que alteram o volume armazenado no reservatório e, portanto, a quantidade de água disponível para evaporação.

Quando o coeficiente de evaporação do reservatório e seu volume contido na simulação são convertidos para a superfície da água, leva-se à perda do nível de evaporação. Isso faz com que o modelo não seja capaz de fornecer a quantidade de água necessária

para atender aos requisitos de evaporação. Nesses casos, o ONS flexibiliza o controle sobre o modelo e ignora qualquer evaporação que ocorra dentro do reservatório.

Cada reservatório possui um coeficiente de evaporação que varia de acordo com a tomada de decisão do modelo e que altera o volume armazenado nos reservatórios. Por conta disso, o volume “disponível” para evaporação depende das condições de resolução do problema. O balanço hídrico pode resultar na violação da evaporação, quando o modelo é incapaz de disponibilizar a quantidade de água necessária para atender a evaporação, o ONS relaxa a modelagem do reservatório desconsiderando a evaporação no mesmo.

Além das restrições hidráulicas que influenciam diretamente na vazão da água, temos também os fenômenos meteorológicos que influenciam no processo inicial da vazão, ou seja, a chuva.

2.5 Fenômenos

El Niño e La Niña são partes de um mesmo fenômeno atmosférico-oceânico importante, que ocorre no oceano Pacífico Equatorial. O El Niño Oscilação Sul (ENOS) causa consequências em todo o mundo. São eventos climáticos que acontecem periodicamente e que se caracterizam pelo aquecimento ou resfriamento das águas do oceano Pacífico, segundo Instituto Nacional de Pesquisas Espaciais - INPE (2023) , além de alterações na circulação atmosférica, gerando transformações no clima do Brasil, afetando o regime de chuvas de diferentes formas pelo planeta; a economia e a agricultura (positivamente ou negativamente).

El Niño e La Niña apresentam características controversas. Enquanto o El Niño consiste no aquecimento anormal das águas superficiais do Oceano Pacífico Equatorial, La Niña se caracteriza pelo resfriamento anormal das águas superficiais do Oceano Pacífico Equatorial. Em função do El Niño, os ventos sopram com menos força ou até invertem a direção em todo o centro do Oceano Pacífico. Por isso, acaba ocorrendo uma diminuição da ressurgência de águas profundas e na acumulação de água mais quente na costa oeste da América do Sul, acarretando altos índices pluviométricos e seca em outros países. No Brasil especificamente, obtemos reduções nas chuvas na região Norte, secas em partes do Nordeste e chuvas acima da média com aumento de temperatura na região Sul.

Já o fenômeno La Niña acontece quando há a intensificação dos ventos alísios², que sopram na faixa equatorial de leste para oeste. Dessa maneira, uma quantidade maior que o normal de águas quentes se acumula no Oceano Pacífico Equatorial Oeste e no Pacífico Leste verifica-se a presença de águas mais frias, ocasionando um aumento no desnível entre eles. No Brasil, alguns dos efeitos sentidos são: aumento de chuvas nas regiões Norte e Nordeste, estiagem no Centro-Oeste e no Sul.

² Os ventos alísios são, por definição, deslocamentos de massas de ar em direção à Zona de Convergência Intertropical do globo terrestre.

Segundo o National Oceanic and Atmospheric Administration (2023) os eventos de El Niño e La Niña foram identificados utilizando o Índice Oceânico Niño, que quantifica as anomalias na temperatura da superfície do mar do oceano Pacífico, obtendo uma média móvel de três meses consecutivos (Figura 2).

Figura 2 – Índice Oceânico e a classificação de anos de ocorrência do El Niño (fonte na cor vermelha), La Niña (fonte na cor azul) e anos de neutralidade (fonte na cor preta), para o período de 1950 a 2022.

Year	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
1950	-1.5	-1.3	-1.2	-1.2	-1.1	-0.9	-0.5	-0.4	-0.4	-0.4	-0.6	-0.8
1951	-0.8	-0.5	-0.2	0.2	0.4	0.6	0.7	0.9	1.0	1.2	1.0	0.8
1952	0.5	0.4	0.3	0.3	0.2	0.0	-0.1	0.0	0.2	0.1	0.0	0.1
1953	0.4	0.6	0.6	0.7	0.8	0.8	0.7	0.7	0.8	0.8	0.8	0.8
1954	0.8	0.5	0.0	-0.4	-0.5	-0.5	-0.6	-0.8	-0.9	-0.8	-0.7	-0.7
1955	-0.7	-0.6	-0.7	-0.8	-0.8	-0.7	-0.7	-0.7	-1.1	-1.4	-1.7	-1.5
1956	-1.1	-0.8	-0.6	-0.5	-0.5	-0.5	-0.6	-0.6	-0.5	-0.4	-0.4	-0.4
1957	-0.2	0.1	0.4	0.7	0.9	1.1	1.3	1.3	1.3	1.4	1.5	1.7
1958	1.8	1.7	1.3	0.9	0.7	0.6	0.6	0.4	0.4	0.4	0.5	0.6
1959	0.6	0.6	0.5	0.3	0.2	-0.1	-0.2	-0.3	-0.1	0.0	0.0	0.0
Year	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
1960	-0.1	-0.1	-0.1	0.0	0.0	0.0	0.1	0.2	0.3	0.2	0.1	0.1
1961	0.0	0.0	0.0	0.1	0.2	0.3	0.1	-0.1	-0.3	-0.3	-0.2	-0.2
1962	-0.2	-0.2	-0.2	-0.3	-0.3	-0.2	0.0	-0.1	-0.1	-0.2	-0.3	-0.4
1963	-0.4	-0.2	0.2	0.3	0.3	0.5	0.9	1.1	1.2	1.3	1.4	1.3
1964	1.1	0.6	0.1	-0.3	-0.6	-0.6	-0.6	-0.7	-0.8	-0.8	-0.8	-0.8
1965	-0.6	-0.3	-0.1	0.2	0.5	0.8	1.2	1.5	1.9	2.0	2.0	1.7
1966	1.4	1.2	1.0	0.7	0.4	0.2	0.2	0.1	-0.1	-0.1	-0.2	-0.3
1967	-0.4	-0.5	-0.5	-0.4	-0.2	0.0	0.0	-0.2	-0.3	-0.4	-0.3	-0.4
1968	-0.6	-0.7	-0.6	-0.4	0.0	0.3	0.6	0.5	0.4	0.5	0.7	1.0
1969	1.1	1.1	0.9	0.8	0.6	0.4	0.4	0.5	0.8	0.9	0.8	0.6

Fonte: *National Oceanic and Atmospheric Administration (2023)* (NOAA)

Sendo assim para os valores obtidos maiores do que $+0,5^{\circ}\text{C}$, o período é considerado de El Niño, e quando for menor que $-0,5^{\circ}\text{C}$, caracteriza-se como La Niña, e os critérios para classificação da intensidade estão apresentados na Tabela 1.

Tabela 1 – Critérios para a classificação da intensidade do fenômeno ENOS.

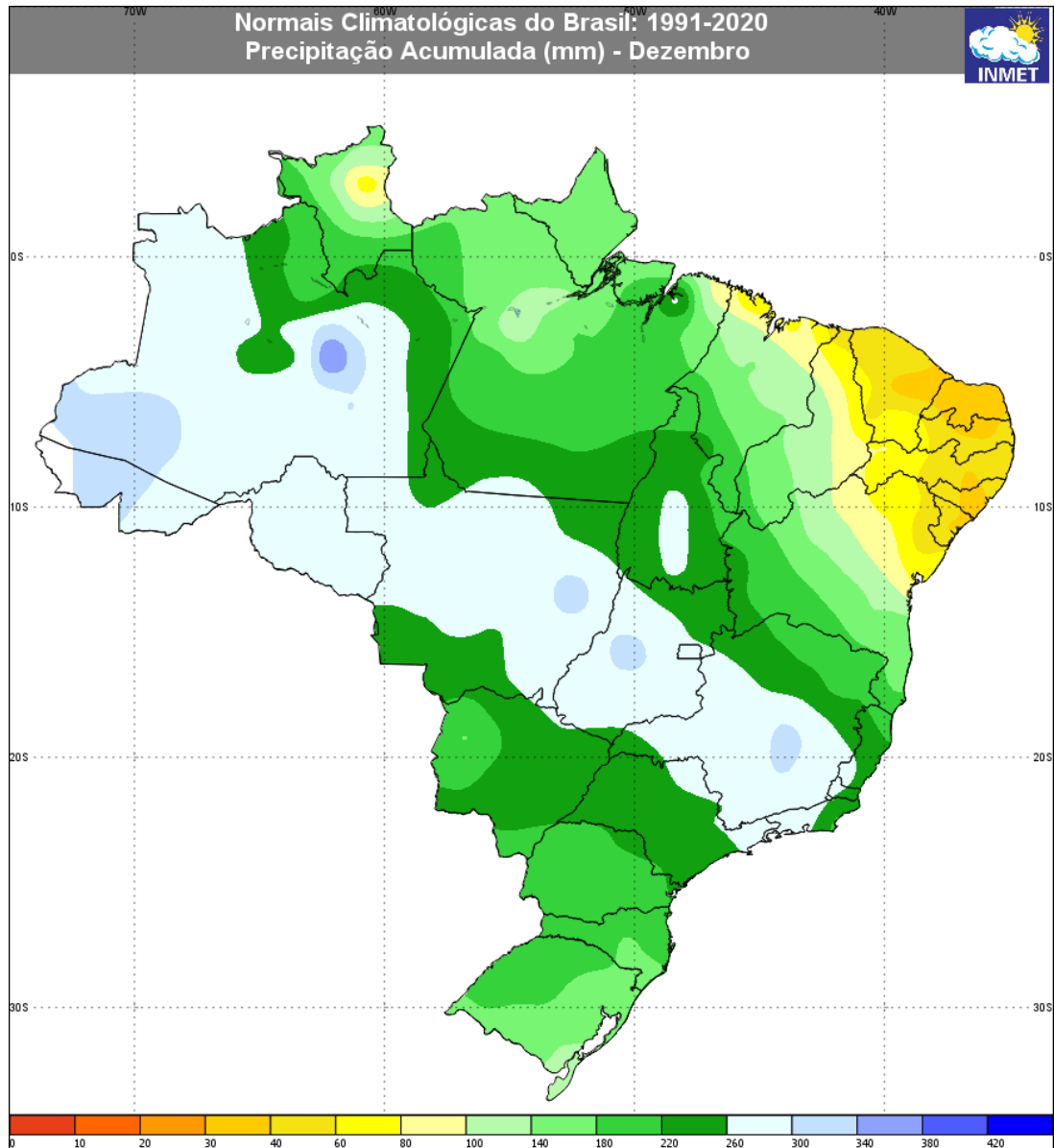
Evento	Índice Oceânico Niño	Classificação da Intensidade
El Niño	0,5 a 0,9	Fraca
	1,0 a 1,4	Moderada
	$\geq 1,5$	Forte
La Niña	-0,5 a -0,9	Fraca
	-1,0 a -1,4	Moderada
	$\geq -1,5$	Forte

Fonte: A autora (2023).

2.6 Zona de Convergência do Atlântico Sul (ZCAS)

Segundo o Instituto Nacional de Meteorologia - INMET (2023) a Zona de Convergência do Atlântico Sul (ZCAS) é um dos principais sistemas climáticos responsáveis pelas chuvas do final da primavera e do verão no Centro-Oeste e Sudeste. A ZCAS é definida como o corredor de nuvens pelo Brasil, desde a região Sul da Amazônia até o Oceano Atlântico, passando pela faixa central do país. Eventualmente a ZCAS também pode se mover para cima ou para baixo, formando uma frente fria, um centro de baixa pressão ou uma área alongada de baixa pressão superficial. Esse evento é facilmente identificável em imagens de satélite devido à organização da nuvens.

Figura 3 – Média Climatológica de Precipitação para Dezembro



Fonte: Climatologia de referência: 1991 - 2020 (Instituto Nacional de Meteorologia - INMET, 2023)

Conforme a imagem anterior é possível notar que os maiores acumulados de chuva referente à média climatológica ocorrem desde o Amazonas até o Rio de Janeiro, passando por Mato Grosso, Goiás, Minas Gerais e norte de São Paulo, superando os 250mm em dezembro. O período de ocorrência da ZCAS é definido entre outubro e março, ou seja, dentro da estação monçônica brasileira.

Quando o Fenômeno El Niño ocorre durante a primavera-verão, devido à mudança na circulação de ventos em altitude, são esperados eventos mais fracos de ZCAS. O contrário ocorre quando há atuação da La Niña que favorece ZCAS mais amplas e com chuva bem distribuída.

Ressalta-se que a grande quantidade de chuva alimenta alguns dos principais reservatórios do país e garante importantes reservas de água durante a estação seca do ano, trazendo vantagens para o setor elétrico. Por outro lado, as chuvas trazidas pela ZCAS também podem causar desastres naturais como: enchentes, deslizamentos de terra e transbordo de rios.

2.7 Regime Anual de Chuva no Brasil

É visto que a matriz elétrica do Brasil é fortemente dependente da disponibilidade hídrica. Devido à sua vasta extensão territorial, o Brasil contém diversas bacias hidrográficas, que são essenciais para a geração de energia que se encontram em diferentes localidades tendo assim diferentes tipos de climas, tornando as condições de chuva, temperatura e vento altamente sazonais.

Sendo assim, a região Sudeste/Centro-Oeste é onde predominam-se os reservatórios do país, que respondem por 70% da capacidade de armazenamento, e sua energia armazenada tem forte capacidade de regularização, contribuindo para a operação durante todo o ano, independente da estação.

Já no Sul obtemos chuvas uniformemente distribuídas, e suas características do solo trazem um comportamento hidrológico com grandes variações ao longo do ano.

Na região Nordeste encontram-se ventos constantes o ano todo tornando-a a maior produtora de energia eólica no país. A região Norte apresenta chuvas bastante sazonais, com intensidade entre fevereiro e maio. Como predominam as usinas de fio d'água, sem reservatório de energia, a energia gerada nesse período é exportada para a região Sudeste/Centro-Oeste.

O início do período chuvoso no Brasil é muito esperado pelo setor de energia elétrica, pois os níveis de armazenamento tendem a ser baixos no final do período seco devido a pouca chuva no outono e no inverno. Com isso, as acilações da Energia Natural Afluente (ENA) interferem diretamente no Preço de Liquidação de Diferenças (PLD).

2.8 Energia Natural Afluente

Energia Natural Afluente (ENA) é um conceito teórico utilizado nas áreas de recursos hídricos e hidrelétricas. Refere-se à quantidade de energia aproveitada do fluxo de água em uma bacia hidrográfica em um determinado período. É um indicador essencial para a gestão e planejamento da operação de usinas hidrelétricas (Esfera Inteligência é Energia,). A ENA é afetada por muitos fatores, sendo o principal fator a precipitação na bacia hidrográfica. A quantidade de chuva em uma determinada área determina a quantidade de água disponível para o sistema de abastecimento de água, que por sua vez afeta diretamente a geração de energia hidrelétrica. Outros fatores que podem afetar

a ENA incluem a evaporação, a infiltração no solo, o escoamento superficial e o fluxo dos rios afluentes. A medição e previsão da ENA são atividades complexas que requerem grandes quantidades de dados e análises. Órgãos responsáveis pela gestão de recursos hídricos e operação de usinas hidrelétricas utilizam modelos matemáticos e estatísticos para estimar a ENA com base em dados históricos e em tempo real. Essas estimativas são críticas para planejar as operações. A ENA desempenha um papel vital na determinação da disponibilidade de energia hidrelétrica. A partir disso, pode-se calcular uma energia fixa, que representa a energia garantida que uma usina hidrelétrica pode produzir em um determinado período, levando em consideração a variação da ENA ao longo do tempo. Ressaltando as variações sazonais e climáticas que a ENA está sujeita, que afetam significativamente a disponibilidade de energia hidrelétrica, destacando a importância da gestão integrada e sustentável dos recursos hídricos.

2.9 Preço de Liquidação das Diferenças (PLD)

O Preço de Liquidação das Diferenças (PLD) é um valor calculado pela Câmara de Comercialização de Energia Elétrica (CCEE) (2023) para valorar a diferença entre a energia gerada/contratada e o total que foi consumido pelo mercado.

O PLD serve como métrica para definir o valor da energia no Mercado de Curto Prazo (MCP) e é influenciado por fatores como: o volume de energia gerado; as condições climáticas; o preço dos combustíveis e a demanda de energia dos consumidores. O PLD segundo o site Mercado Livre de Energia Elétrica (2023) é usado para valorar os volumes de energia liquidados na CCEE e tem como base o custo marginal de operação (CMO), calculado pelo Operador Nacional do Sistema (ONS), que utiliza os modelos matemáticos NEWAVE (para médio prazo), DECOMP (curto prazo) e DESSEM e DECOMP (curto prazo), fornecendo o CMO para determinar quanto custará um MWh (megawatt hora) de produção adicional para o Sistema Elétrico.

O PLD é calculado diariamente para cada hora do dia seguinte, considerando a aplicação dos limites máximos e mínimos vigentes. E seu processo de precificação é dividido por submercados (Sul, Sudeste/Centro-Oeste, Nordeste e Norte), conforme a figura abaixo:

Figura 4 – Submercados de Energia Brasileira.



Fonte: Coelho (2017).

Desde 2020, o Preço de Liquidação das Diferenças (PLD) é limitado por valores mínimo e máximo de acordo com a Resolução Normativa ANEEL nº 858/19, com validade entre a primeira e a última semana operativa de cada ano. Conforme é especificado no site da CCEE (2023)

2.10 Comercialização de Energia

Conforme mencionado anteriormente, a CCEE realiza estudos com base no planejamento do ONS para identificar o PLD. Este preço é utilizado para avaliar a diferença entre o recurso do agente e a demanda do mercado de energia. O recurso seria a quantidade de energia que um agente tem o direito de consumir ou vender para outros agentes com base na quantidade de eletricidade gerada pela usina ou contrato de compra de energia. Já a demanda refere-se à quantidade de energia consumida ou alocada em um contrato de venda, ou seja, seu agente possui obrigações de provimento de seus recursos (CCEE, 2023).

Como a produção e o consumo diferem, essas quantidades tendem a diferir, resultando em um balanço energético positivo ou balanço energético negativo.

Quando positivo, o agente deve ser pago por fornecer mais energia do que realmente está disponível no contrato. Contrapartida o agente terá um balanço energético negativo devido ao descumprimento de suas obrigações contratuais. Essa diferença é valorada ao PLD, gerando um risco financeiro positivo ou negativo para o agente. Do ponto de vista prático, a CCEE mantém registros dos contratos de agência, como medições de geração e consumo, e no final do período contábil com base no mês-calendário determina-se o valor a ser cobrado ou pago ao agente, o chamado Mercado de Curto Prazo (MCP).

Na prática, as relações são mais complexas, uma vez que existem diversos agentes interagindo entre si, além de questões como a inadimplência no setor, por conta da judicialização de pagamentos e da incidência de encargos aos agentes.

Como o PLD depende das condições de energia elétrica, principalmente condições hidrológicas futuras, os agentes operam contratos de energia se antecipando às possíveis exposições ao PLD, com base em sua pesquisa de percepção de mercado e previsão de preço usando cadeias de modelos de planejamento.

3 Materiais e Métodos

3.1 Material

3.1.1 Origem dos Dados

Para o desenvolvimento das análises propostas neste projeto, utiliza-se o banco de dados do histórico de ENA fornecido pelo ONS e por recursos computacionais, ambos são descritos a seguir.

3.1.2 Banco de Dados

Utiliza-se o banco de dados disponibilizado pelo Operador Nacional do Sistema Elétrico (2023a). Este banco possui séries históricas de ENA entre 1931 e 2022, com dados mensais para cada submercado. Tais séries são disponibilizadas em MWm¹ com respectivos valores para cada região. E para a realização do estudo do comportamento da variação da ENA, observa-se o subsistema Sudeste/Centro-Oeste, especificamente. Abaixo, apresenta-se um recorte de como a informação de ENA está armazenada.

Figura 5 – Dados de ENA da Região Sudeste/Centro-Oeste.

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1931	69275	103169	107109	80141	54954	40720	31951	25395	25905	27641	31824	47506
1932	69258	76880	68506	51780	40197	35184	27353	21307	17871	25287	31530	58457
1933	79751	68472	58364	53285	38380	27967	21735	17153	16137	20972	22639	44319
1934	57465	52690	52066	43230	30940	22874	18612	13734	13849	15110	18309	46481
1935	71982	91162	80227	71577	44893	33543	25131	23870	21139	38047	30633	39384
1936	46920	43730	66094	47686	33473	25475	19525	16899	19508	16641	21192	40376
1937	75736	58491	58074	52146	38219	29558	21203	16943	14672	24745	39243	60712
1938	64557	58958	53144	41733	33494	25449	22208	18411	16052	21779	26635	46258
1939	65621	67847	45816	40913	32650	23993	18695	14943	13464	15927	28872	44921
1940	66583	89807	86029	53597	40789	30215	23901	20428	18156	18034	35915	42549
1941	57728	51892	55636	50857	32476	25642	21662	19812	21695	27415	34126	49665
1942	56098	66616	78374	61011	43497	40729	30711	21301	21112	23624	32223	47666
1943	90086	85972	84589	56099	38240	31825	23879	18816	17225	27114	38937	49318
1944	45944	61852	68117	46764	31572	23477	18838	15118	13007	14680	28584	35208
1945	52340	86472	80088	78516	46746	32587	28282	19214	16902	20710	35966	68317
1946	91187	76444	85868	59992	41128	34384	33071	22512	18312	23111	27727	38036
1947	64711	75437	109208	70184	43678	31931	26532	22673	26357	27890	30140	52220
1948	61991	68406	72331	53009	34856	29361	22099	20183	15799	17580	26687	58535
1949	62998	87208	68262	50473	37566	29052	22231	17538	14519	18226	25349	41902
1950	53914	78625	72218	53858	38242	28056	22245	16412	13779	20566	35233	51882
1951	65895	80116	82149	60355	39120	30391	23923	19080	15992	18288	21152	32730

Fonte: CCEE (2023).

¹ Megawatt Médio: Unidade de produção energética igual a energia produzida pela operação contínua de um megawatt de capacidade durante um período de tempo.

3.1.3 Recursos Computacionais

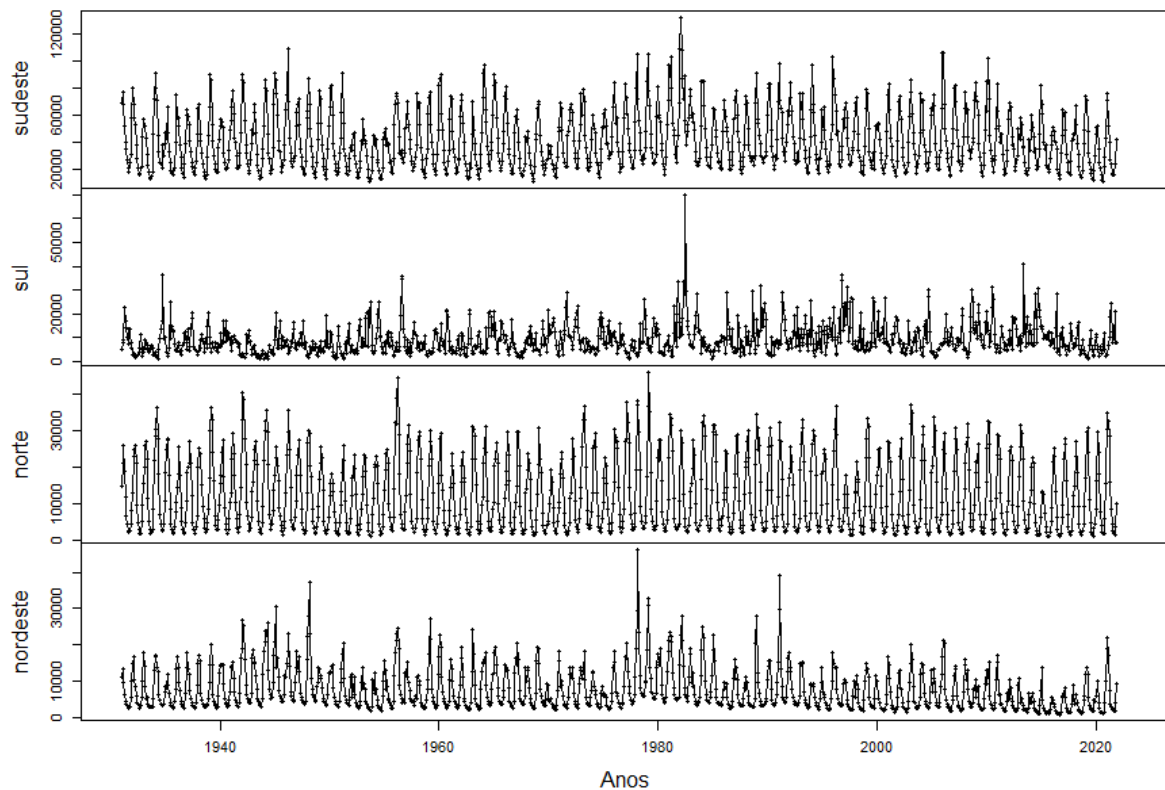
Utiliza-se o *software* R, versão 3.1.1 (R CORE TEAM, 2021) para realizar a aplicação das séries temporais e as bibliotecas utilizadas foram o *forecast* para realizar a análise através do modelo SARIMA e o *Rssa* para a análise através do modelo SSA.

3.2 Métodos

3.2.1 Séries Históricas

Essa análise é separada por submercado, tendo foco no submercado Sudeste/Centro-Oeste. Levando em conta também as estações do ano e eventos climáticos. A Figura 6 apresenta o comportamento das séries para os quatro submercados. Pode-se perceber que a quantidade de energia natural afluyente oscila entre os meses de cada ano e a diferença entre os submercados.

Figura 6 – Séries Históricas ENA para os Quatro Submercados.



Fonte: A autora (2023).

Pode-se observar a volatilidade dos dados de ENA e diferenças significativas no comportamento para os quatro submercados, por exemplo, a diferença de escala do Norte e Nordeste em relação ao Sudeste/Centro-Oeste e Sul. Isso é decorrência dos eventos climáticos discutidos anteriormente.

3.2.2 Séries Temporais

Uma série temporal é um conjunto de observações ordenadas no tempo. Essas técnicas podem ser divididas em dois grupos: contínuas e discretas. As séries discretas podem ser obtidas através de amostras retiradas em períodos de tempo equidistantes, e as contínuas são aquelas que podem assumir valores infinitos dentro de um intervalo específico como: altura, peso, temperatura, entre outras. Os objetivos das análises de séries temporais são de desenvolver modelos paramétricos, onde o número de parâmetros é finito, ou não-paramétricos, onde o número de parâmetros é infinito, fazendo projeções de valores futuros ou apenas relatando o comportamento da série, construindo modelos probabilísticos ou estocásticos, ou seja, controlados por leis probabilísticas MORETTIN P.A.; TOLOI (2006). Uma série temporal de comprimento T será representada por:

$$Z_T = z_1, z_2, \dots, z_T \quad (3.2.2)$$

onde:

z_1 = É a variável aleatória que representa o valor obtido pela série no primeiro momento;

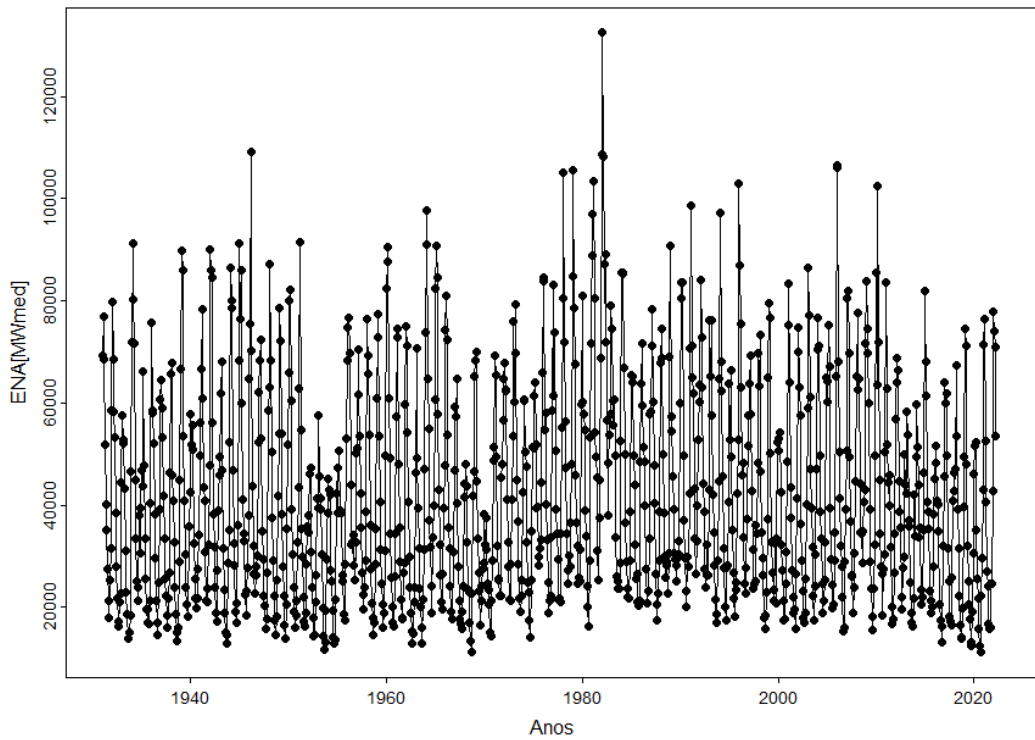
z_2 = Indica o valor do segundo período de tempo;

z_3 = Representa o valor do terceiro período de tempo e assim por diante.

Em geral, o conjunto de variáveis aleatórias Z_t , indexado por t é chamado de processo aleatório.

Uma série temporal é normalmente exibida graficamente plotando o valor de uma variável aleatória no eixo horizontal ou vertical, com a escala de tempo como abscissa. Muitas vezes é conveniente concatenar valores de períodos de tempo adjacentes para reconstruir visualmente algumas séries temporais contínuas que podem ter gerado esses valores como amostras discretas. Na figura 7, tem-se em destaque, a série de ENA para o submercado Sudeste/Centro-Oeste, são 91 anos de dados mensais. A princípio, não vê-se padrões ou comportamentos previsíveis.

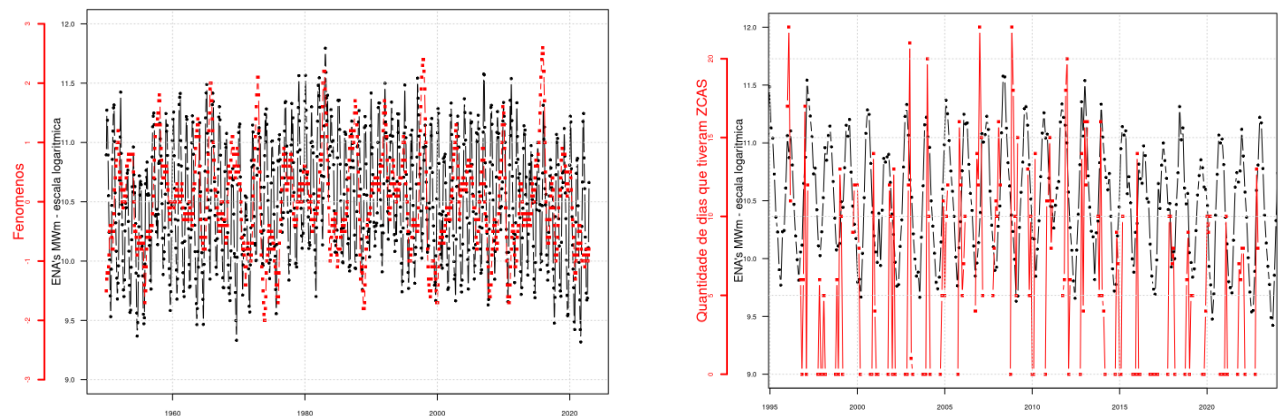
Figura 7 – Série Histórica do Sudeste/Centro-Oeste.



Fonte: A autora (2023).

Nas próximas figuras, fez-se a sobreposição das séries no mesmo período de tempo na intenção de detectar comportamentos correlacionados. À esquerda têm-se a série de ENA com a série dos fenômenos ENOS, e à direita, a mesma série de ENA com a série de dados das ZCAS.

Figura 8 – Séries Históricas de ENA (escala logarítmica), Eventos Meteorológicos e ZCAS.



a) Série de Fenômenos e ENA.

b) Série de ZCAS e ENA.

Observando as séries sobrepostas, nota-se que a ENA possui um comportamento independente, tanto dos fenômenos ENOS quanto das ZCAS, este resultado é diferente do se esperava. Com isso decidiu-se utilizar na análise apenas a própria série de ENA na aplicação dos métodos, com foco na região Sudeste/Centro-Oeste.

3.2.3 Decomposição de Séries Temporais

A decomposição de séries temporais é uma técnica para quebrar uma série temporal através de seus padrões básicos: tendência, sazonalidade, ciclo e irregularidade (GAMBONI, 2017).

- Tendência em função do tempo ($T(t)$): É o movimento de longo prazo de uma série, indicando se ela está aumentando, diminuindo ou permanecendo estável;
- Sazonalidade em função do tempo ($S(t)$): É um padrão que se repete regularmente, como meses ou estações, e está relacionado a fatores externos, como: clima, feriados ou festas populares;
- Ciclos em função do tempo ($C(t)$): Os ciclos são flutuações mais longas do que os ciclos sazonais e estão relacionados a fatores econômicos, políticos ou climáticos, como os fenômenos já citados anteriormente;
- Irregularidade em função do tempo $I(t)$: Irregularidades são partes aleatórias da sequência que não podem ser explicadas por outros componentes.

A decomposição da série permitirá identificar quais componentes estão atuando naquele conjunto, além de possibilitar obter índices e/ou equações para realizar previsões para períodos futuros da série. Dessa forma, é possível decompor a série em:

$$Z(t) = T(t) + S(t) + R(t)$$

Onde $T(t)$ e $S(t)$ representam a tendência e a sazonalidade, respectivamente, e $R(t)$ é uma componente ruidosa (aleatória que pode incluir os elementos irregulares e alguns elementos identificados nas variações cíclicas). Para se estimar a tendência, deve ser usada algum tipo de suavização.

Existem diferentes métodos de decomposição de séries temporais, como média móvel, decomposição clássica e decomposição de Tendência Sazonal usando LOESS² (STL). Uma média móvel é um método simples que envolve o cálculo da média dos valores da série durante um período de tempo para suavizar as flutuações e estimar a tendência, sendo melhor para sazonalidade estocástica.

² LOESS é uma técnica de regressão não paramétrica usada para estimar a relação entre duas variáveis. É uma sigla em inglês para “*Local Regression Scatterplot Smoothing*” (suavização de gráfico de dispersão de regressão local) e é usada para ajustar modelos de regressão em dados com ruído.

Já a decomposição clássica, é um método que assume que os componentes da série são aditivos ou multiplicativos, ou seja, que a série é igual à soma ou ao produto dos componentes. A decomposição STL é um método mais flexível que usa algoritmos de suavização local para estimar os componentes da série, permitindo variações na tendência e na sazonalidade.

Na sequência, são apresentados os métodos escolhidos para a aplicação do conjunto de dados de ENA.

3.2.4 Modelo *Autoregressive Integrated Moving Average* (ARIMA)

Segundo BAYER F.M.; SOUZA (2010) a classe de modelos tradicionalmente mais utilizada em análise de séries temporais é a dos modelos ARIMA. Esse modelo tem sido pouco utilizado na área hidrológica, apesar de seu grande poder de modelagem e de previsão. Alguns trabalhos antigos introduziram o uso de séries temporais na hidrologia, como pode ser visto em LEDOLTER (1976).

ARIMA é um modelo estatístico que utiliza a regressão linear para prever valores futuros de uma série temporal. Ele é uma extensão do modelo ARMA (*Autoregressive Moving Average*) que consiste em uma técnica para entender e, talvez, prever valores futuros da série. Conforme SOUZA (2011) o mesmo consiste de duas partes, uma parte auto regressiva (AR) e uma parte de média móvel (MA). A parte AR envolve regressar a variável em seus próprios valores passados. A parte MA envolve modelar o termo de erro como uma combinação linear de termos que ocorrem contemporaneamente e em vários momentos no passado.

A diferença se dá pelo modelo ARIMA ser utilizado para dados que apresentam comportamento de não estacionariedade, e assim é aplicado um processo de diferenciação para transformar os dados em estacionários (BRESSAN, 2004). Sendo possível aplicar os modelos auto-regressivos (AR) e médias móveis (MA).

Muitas vezes, a dependência do passado tende a ocorrer mais fortemente em múltiplos de alguns intervalos (chamados de *lags*) sazonais subjacentes. Por exemplo, com dados mensais de ENA, há um forte componente anual ocorrendo em *lags* que são múltiplos de $s = 12$ devido fortes correspondentes às estações do ano que também incluem fenômenos naturais e a temperatura que tendem a combinar com as flutuações sazonais. Por causa disso, é apropriado introduzir polinômios autorregressivos de médias móveis que se identifiquem com as defasagens sazonais.

Como citado anteriormente, uma série que não apresenta característica de estacionariedade pode tornar-se estacionária pela utilização de um processo de diferenciação. Sendo assim, se uma série se torna estacionária após d diferenças, a série é dita integrada (I) de ordem d . Com isso, o modelo ARMA integrado passa a ser denominado como ARIMA(p, d, q), sendo p , d , e q números inteiros não negativos, onde p é o número de defasagens do modelo AR, d é o índice de diferenciação e q é o grau de média móvel.

Em estudos hidrológicos é comum encontrar sazonalidade nas séries. O que torna necessário a utilização de modelos que incorporem essa importante característica temporal. Neste contexto, utilizando então o modelo SARIMA que é a extensão do modelos ARIMA e inclui um passo adicional de diferenciação sazonal da série temporal.

- Modelo *Seasonal Autoregressive Integrated Moving Average* (SARIMA): o modelo multiplicativo sazonal autorregressivo integrado de médias móveis ou modelo SARIMA, é composto por três partes: pelo AR, MA e uma parte sazonal S. A parte autorregressiva é responsável por modelar a dependência serial, a parte de média móvel é responsável por modelar o ruído da série temporal e a parte sazonal é responsável por modelar a sazonalidade.

O modelo SARIMA(p, d, q)(P, D, Q) é dado por:

$$\phi(B)\phi(B)(1-B)^d(1-B^s)^dZ_t = \Theta(B)\Theta(B)\epsilon_t$$

onde (p, d, q) são as ordens do modelo referentes à dinâmica ordinal já mencionadas e (P, D, Q) são as ordens da parte sazonal. Os parâmetros $\phi_1, \dots, \phi_p$ e $\Phi_1, \dots, \Phi_p$ são os parâmetros autorregressivos sazonais e de médias móveis sazonais, respectivamente.

A seguir é apresentado um exemplo de aplicação do modelo SARIMA no *software* R com os dados de ENA em escala logarítmica e a justificativa da escolha dos parâmetros.

A escala logarítmica foi utilizada para ajustar a série a uma série estacionária e assim ser possível utilizar a teoria dos modelos SARIMA.

A utilização do algoritmo nos permite perceber o modelo escolhido como mais adequado; retornando os coeficientes autorregressivos (ARIMA) e sazonais (SARIMA) com a 1ª diferenciação de um ano atrás, para isto utilizou-se o critério AIC, ou seja, dentre todos os modelos SARIMA, de ordem até SARIMA (p=5, q=5, d=2, P=2, Q=2, D=1) resultando em 94 modelos diferentes e àquele com o menor AIC foi o selecionado.

```

1
2 library ( forecast )
3 Sudeste.ARIMA = auto.arima(Sudeste)
4 Sudeste.ARIMA
5
6 ## Series: Sudeste
7 ## ARIMA(0,0,2)(0,1,1)[12] with drift
8 ##
9 ## Coefficients:
10 ##          ma1      ma2      sma1  drift
11 ##          0.6199  0.3057  -0.8774  0e+00
12 ## s.e.    0.0339  0.0296   0.0226  1e-04
13 ##
14 ## sigma^2 = 0.03181: log likelihood = 254.05
15 ## AIC=-498.1   AICc=-498.03   BIC=-474.34

```

```

16
17 best.model.ARIMA = arima(Sudeste, order = c(0,0,2), seasonal = list(order =
    c(0,1,1), period = 12))

```

Listing 3.1 – Modelo SARIMA.

Dessa forma, esse modelo foi armazenado na variável `best.model.ARIMA` para que fosse possível dar sequência ao estudo.

```

1
2 best.model.ARIMA = arima(Sudeste, order = c(0,0,2), seasonal = list(order =
    c(0,1,1), period = 12))
3
4 summary(best.model.ARIMA)
5
6 ##
7 ## Call:
8 ## arima(x = Sudeste, order = c(0, 0, 2), seasonal = list(order = c(0, 1,
    1), period = 12))
9 ##
10 ## Coefficients:
11 ##          ma1      ma2      sma1
12 ##       0.6198  0.3057  -0.8771
13 ## s.e.    0.0339  0.0296  0.0224
14 ##
15 ## sigma^2 estimated as 0.03167:  log likelihood = 254.04,  aic = -500.09
16 ##
17 ## Training set error measures:
18 ##
19 ##           ME      RMSE      MAE      MPE      MAPE
20 ## MASE
21 ## Training set 0.0009419772 0.1767184 0.1378935 -0.01887441 1.309973
    0.5524781
20 ##           ACF1
21 ## Training set 0.08783801

```

Listing 3.2 – Modelo SARIMA.

Para verificar a qualidade do ajuste, utilizou-se as linhas de comando exibidas para gerar as estimativas e resíduos, que são visualizados a seguir.

```

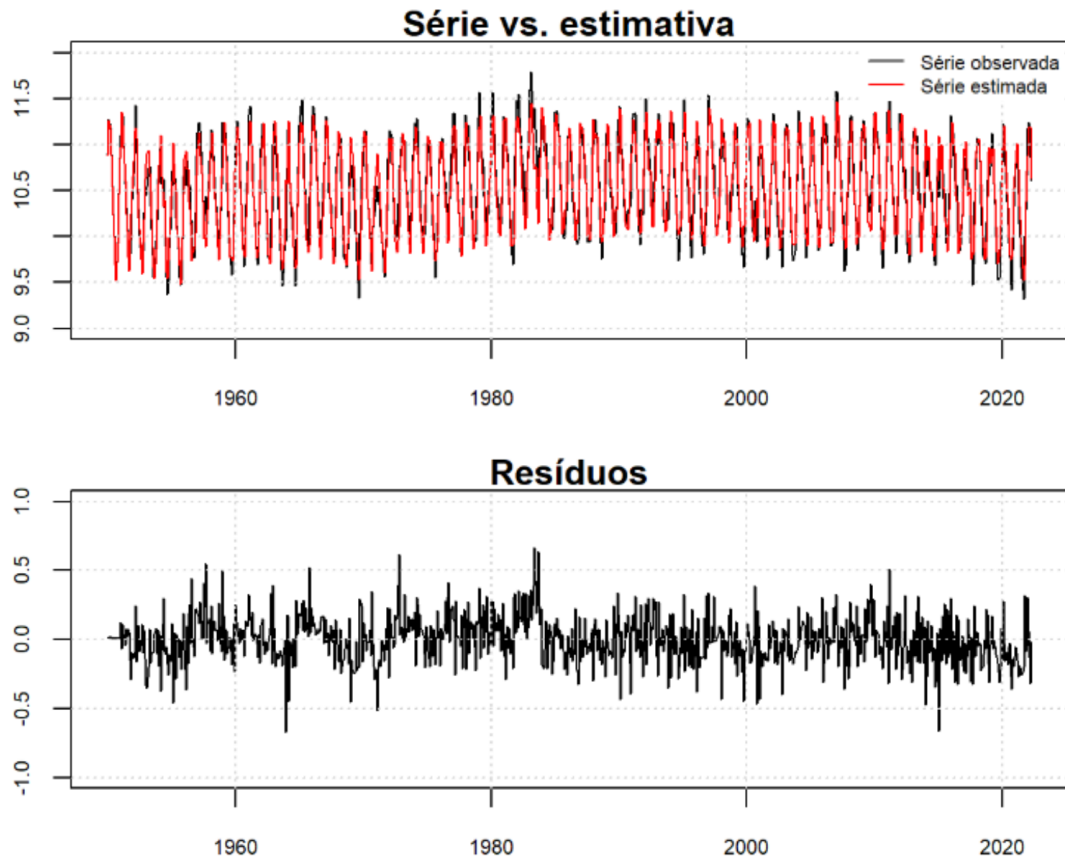
1
2 par(mar=c(3, 6, 1, 1) + 0.1, cex.axis=0.7, cex=0.7, mfrow=c(2,1))
3 plot(Sudeste, pch=16, ylim=c(9,12), xlab="", ylab="", col="black", main="
    Serie vs. estimativa")
4 lines(fitted(best.model.ARIMA), col = "red")
5 grid()
6 plot(resid(best.model.ARIMA), main = "Resíduos", xlab="", ylab="", ylim=c
    (-1,1), col="black")
7 grid()

```

Listing 3.3 – Modelo SARIMA

Na figura 9, têm-se as séries originais e estimadas, seguidas da série dos resíduos calculados da diferença entre as duas séries:

Figura 9 – Ajuste através do modelo SARIMA (escala logarítmica).



Fonte: A autora (2023).

Pode-se perceber uma excelente sobreposição da série de estimativas em relação à série original. Para melhor observarmos o comportamento do modelo estimado faremos uma aproximação no gráfico anterior. Para esta aproximação escolheu-se a parte superior do gráfico, ou seja, somente visualizaremos as observações e suas estimativas na faixa entre 10.5 e 12.

```

1
2 Sudeste1 = ifelse(Sudeste > 10.5, Sudeste, NA)
3 Sudeste2 = ifelse(fitted(best.model.ARIMA) > 10.5, fitted(best.model.ARIMA)
4               , NA)
5 par(mar=c(3, 6, 1, 1) + 0.1, cex.axis=0.7, cex=0.7)

```

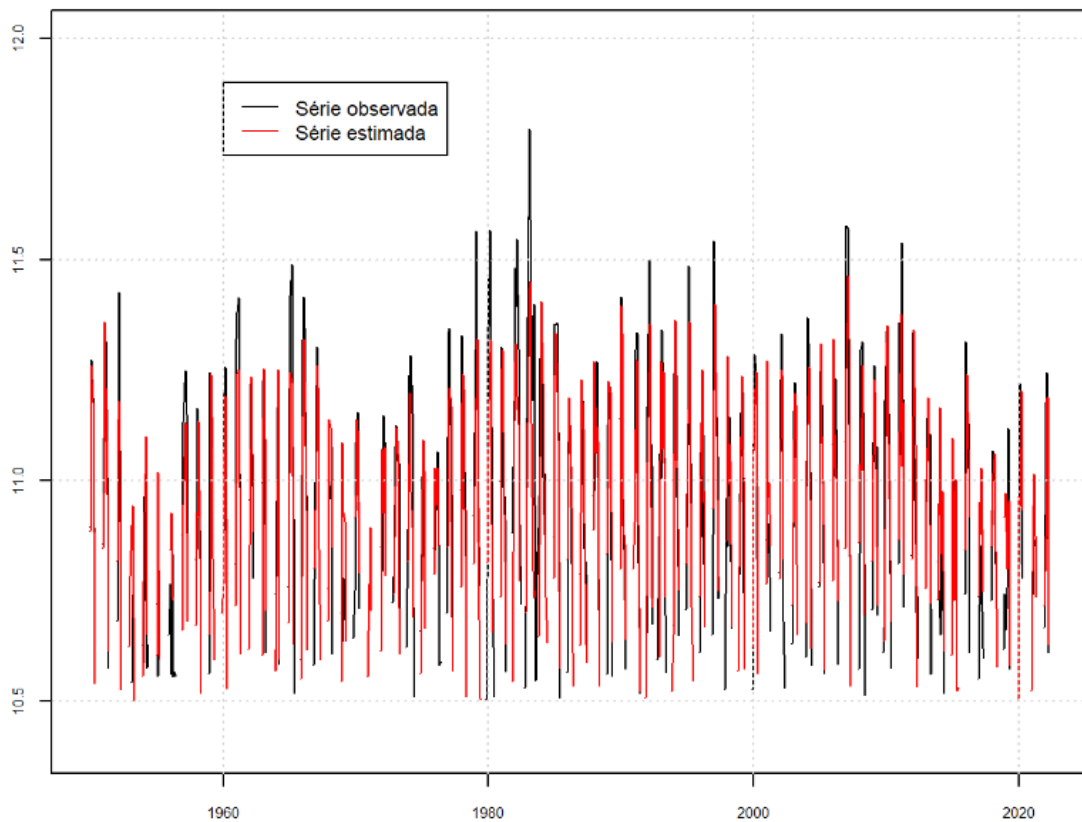
```

5 plot(Sudeste1, pch=16, ylim=c(10.4,12), xlab="", ylab="", col="black", main
   = "")
6 lines(Sudeste2, col = "red")
7 legend(1960,11.9, legend = c("S rie observada", "S rie estimada"), col =
   c("black", "red"), lty = 1)
8 grid()

```

Listing 3.4 – Modelo SARIMA.

Figura 10 – Série cortada do ajuste com o modelo SARIMA (escala logarítmica).



Fonte: Autora (2023).

Percebe-se mais claramente a diferença entre as observações e as estimativas segundo o melhor modelo estimado no conjunto de dados de treinamento, lembrando que os dados estão em escala logarítmica e, portanto, as estimativas também.

3.2.5 Análise Espectral Singular (SSA)

A Análise Espectral Singular (SSA) ou Análise Estrutural de Séries Temporais, é uma técnica poderosa aplicável a muitos problemas práticos, como o estudo de séries

temporais clássicas, estatísticas multivariadas, geometria multivariada, sistemas dinâmicos e processamento de sinais.

O método SSA básico consiste em duas etapas complementares: decomposição e reconstrução; essas duas fases consistem em duas etapas independentes. No primeiro estágio decompõem-se a série, no segundo estágio reconstrói-se a série original e usa-se a série reconstruída para prever novos pontos de dados. O principal conceito no estudo das propriedades SSA é a separabilidade, que descreve o grau de separação entre diferentes componentes.

Observe que, embora os métodos baseados em SSA empreguem alguns conceitos de probabilidade e estatística, não precisamos fazer nenhuma suposição estatística, como estacionariedade da série ou normalidade dos resíduos.

SSA é uma ferramenta muito útil que pode ser usada para resolver diferentes problemas como: encontrar tendências; alisamento; extrair componentes sazonais; extrair periodicidades com diferentes magnitudes; extrair tendências simultaneamente e periodicidade complexa; encontrar estrutura em séries temporais curtas e teste de causalidade.

No campo da análise de séries temporais não lineares, a SSA é considerada uma técnica que pode competir com os métodos clássicos (ARIMA, *Holt-Winters*, etc.). Existem muitos estudos olhando para SSA como um método de filtragem, como estudado por exemplo, por Hassani (2010).

A motivação para se usar a SSA no presente trabalho é por ser uma técnica não paramétrica que trabalha com processos estatísticos arbitrários, sejam lineares ou não lineares, estacionários ou não estacionários, gaussianos ou não gaussianos.

Além disso, ao contrário dos métodos tradicionais de previsão de séries temporais (modelos autorregressivos e estruturais que assumem normalidade e estacionariedade das séries), o método SSA é não paramétrico e não faz suposições prévias sobre os dados, superando as técnicas clássicas. O método também decompõe a série em suas partes constituintes e a reconstrói.

Deve-se notar que, embora o método baseado em SSA use probabilidade e conceitos estatísticos, suposições estatísticas como a estacionariedade da série ou a normalidade do resíduo não são necessárias, e o SSA usa *bootstrapping*³ para obter os intervalos de confiança para as previsões.

Considera-se uma série temporal $Y_T = (y_1, \dots, y_T)$ unidimensional. Seja L um número inteiro fixo tal que $L \leq T/2$ representando o comprimento do intervalo e também seja $K = T - L + 1$.

1º Passo: Computando a matriz de trajetória: Isso transfere a série temporal Y_T para a série multidimensional $X_1, \dots, X_K$, com vetores $X_i = (y_i, \dots, y_{i+L-1}) \in \mathbb{R}^L$. O

³ Bootstrapping é uma técnica de reamostragem que estima a distribuição de amostragem de uma estatística coletando várias amostras com reposição de uma amostra aleatória única. Essa técnica é usada para calcular intervalos de confiança, testar hipóteses ou estimar parâmetros populacionais sem fazer suposições sobre a forma da distribuição da população.

parâmetro único da incorporação é o comprimento da janela L . O resultado desta etapa é a matriz de trajetória $X[X_1, \dots, X_K]$:

$$X = (y_{ij})_{i,j=1}^{L,K} = \begin{bmatrix} y_{11} & y_{12} & y_{13} & \cdots & y_K \\ y_{21} & y_{22} & y_{23} & \cdots & y_{K+1} \\ y_{31} & y_{32} & y_{33} & \cdots & y_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{L+1} & y_{L+2} & y_{L+3} & \cdots & y_T \end{bmatrix}$$

Observe que a matriz de trajetória X é uma matriz de Hankel, o que significa que todos os elementos ao longo da diagonal $i + j = \text{constante}$ são iguais.

2º Passo: Construindo uma matriz para aplicar SVD: Calcular a matriz XX^T .

3º Passo: Decomposição em valores singulares (SVD) da matriz XX^T : Calcular os autovalores e autovetores da matriz XX^T e representá-la como $XX^T = P\Lambda P^T$. Aqui $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_L)$ é a matriz diagonal de autovalores de XX^T : de modo que $\lambda_1 \leq \lambda_2 \leq \dots, \leq \lambda_L \leq 0$ e $P = (P_1, P_2, \dots, P_L)$ é a matriz ortogonal correspondente de autovetores de XX^T .

4º Passo: Seleção de autovetores: Selecione um grupo de $l, 1 \leq l \leq L$ autovetores $P_{I_1}, P_{I_2}, \dots, P_{I_l}$. O passo de agrupamento corresponde a dividir as matrizes elementares X_i em vários grupos e somando as matrizes em cada grupo. Seja $I = i_1, \dots, i_l$ um grupo de índices. Em seguida, a matriz X_I correspondente ao grupo I é definida como

$$X_i = X_{i_1} + \dots + X_{i_l}$$

. Passo 5: Reconstrução da série unidimensional: Calcule a matriz

$$\tilde{X} = ||\tilde{x}_{ij}|| = \sum_{k=1}^l P_{ik} P_{ik}^T X,$$

como uma aproximação para X . A transição para a série unidimensional pode agora ser alcançada pela média das diagonais da matriz $\tilde{X}$.

Abaixo, apresenta-se a forma como se deu a aplicação do modelo SSA no *software* R com os dados de ENA em escala logarítmica.

```

1 library(Rssa)
2 modelo.ssa = ssa(Sudeste, L= 12, kind = "toeplitz-ssa")
3 modelo.ssa1 <- reconstruct(modelo.ssa, groups = list(Trend = 1, Seasonality
4   = 2:11))
5 modelo.ssa2 <- forecast(modelo.ssa, groups = list(1:6), method = "recurrent
6   ", len = 12,
7   interval = "confidence", only.intervals = FALSE, R
8   = 100, level = 0.95)
9 plot(Sudeste, type = "l", main = "")
10 modelo.ssa3 = modelo.ssa1$Trend+modelo.ssa1$Seasonality

```

```

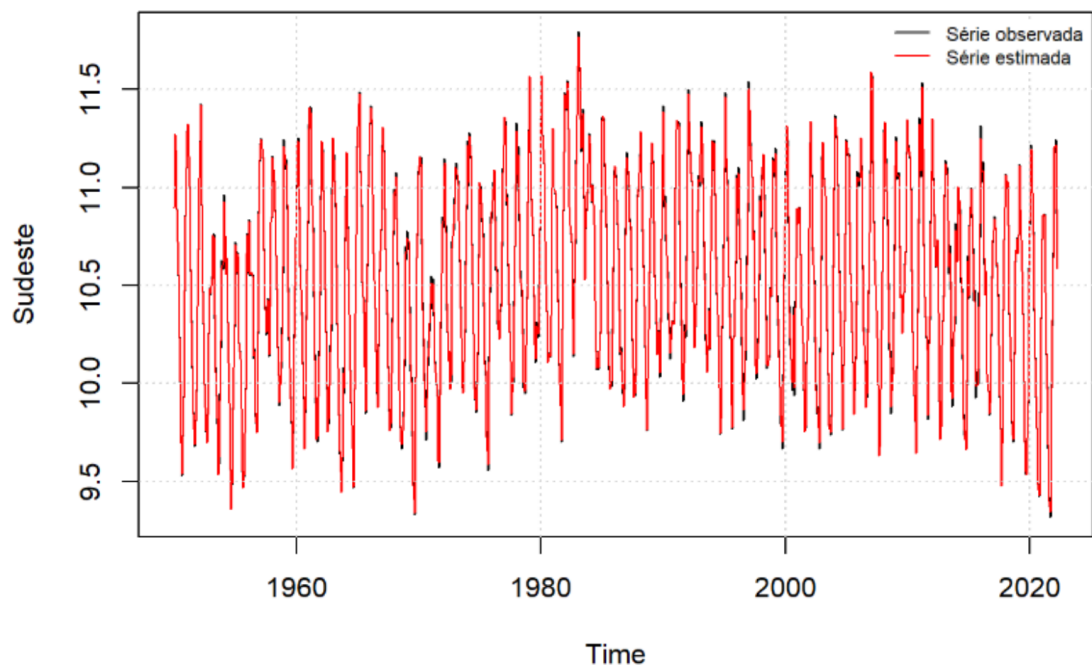
9 lines(modelo.ssa3, col = "red")
10 grid()

```

Listing 3.5 – Modelo de Análise da Espectro Singular (SSA).

Na figura 11, apresenta-se a série reconstruída pelo método SSA sobreposta a série original. Observa-se que as séries são bastante similares.

Figura 11 – Reconstrução da Séie através do SSA (escala logarítmica).



Fonte: A autora (2023).

Novamente para esta aproximação escolhemos a parte superior do gráfico, ou seja, somente visualizamos as observações e suas estimativas na faixa entre 9 e 12 para melhor visualização das sobreposição das séries.

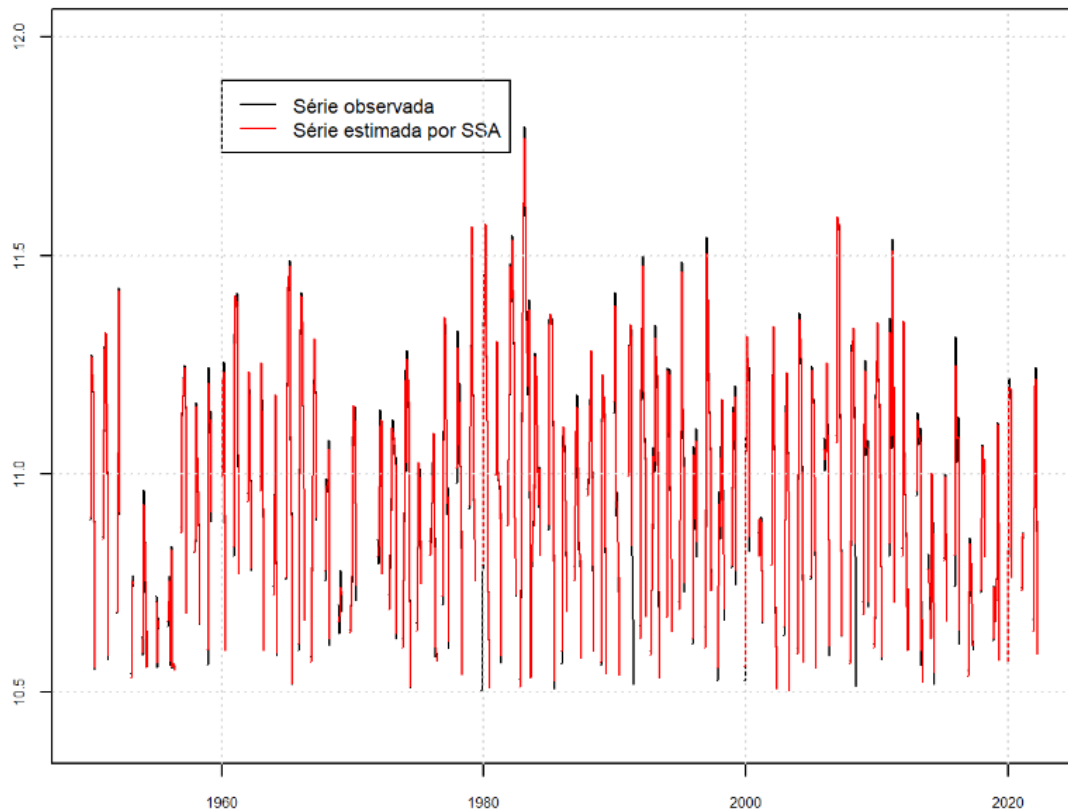
```

1 Sudeste3 = ifelse(modelo.ssa3 > 10.5, modelo.ssa3, NA)
2 par(mar=c(3, 6, 1, 1) + 0.1, cex.axis=0.7, cex=0.7)
3 plot(Sudeste1, pch=16, ylim=c(10.4,12), xlab="", ylab="", col="black", main="" )
4 lines(Sudeste3, col = "red")
5 legend(1960,11.9, legend = c("S rie observada", "S rie estimada por SSA"),
6       , col = c("black", "red"), lty = 1)
7 grid()

```

Listing 3.6 – Modelo de Análise da Espectro Singular (SSA).

Figura 12 – Série cortada do ajuste com o modelo SSA (escala logarítmica)



Fonte: A autora (2023).

3.2.6 Métricas de Avaliação

Para medir um modelo, é necessário analisar os erros que ele apresenta, ou seja, vamos comparar Y e $\hat{Y}$ (Y real e Y previsto, respectivamente) e dar atenção à esses resíduos. (Karoline Penteado,)

Dado que não é confiável analisar apenas o tamanho dos resíduos para saber o quão grande podem ser os verdadeiros erros de previsão. Por esse motivo é interessante avaliar a precisão da previsão utilizando outras métricas. Sendo assim a precisão das previsões pode ser determinada considerando o desempenho de um modelo em novos dados, que não foram usados ao ajustá-lo. Com isso, é necessário dividir a base de dados em duas partes, uma para o treinamento dos modelos e a outra parte é usada para avaliar sua precisão, fornecendo então uma indicação confiável de quão bem o modelo provavelmente preverá novos dados.

Para isso, o conjunto de teste deve ser pelo menos tão grande quanto o horizonte máximo de previsão, para que se possa ter um treinamento adequado. Porém, não necessariamente um modelo que ajusta bem os dados de treinamento irá prever bem novos dados. Em contrapartida um ajuste excessivo de um modelo aos dados é tão ruim

quanto não identificar um padrão sistemático nos dados, segundo o livro Hyndman R.J. Athanasopoulos (2018). A seguir, abordam-se algumas das métricas que utilizamos para avaliar a precisão de ambos os modelos de séries temporais descritos anteriormente.

O primeiro passo é a utilização da função `window()` para extrair uma parte de uma série temporal, criando então um conjunto de treinamento e outro de teste. Neste caso utilizamos os dados a partir de 1951 até abril de 2022.

Para o modelo ARIMA, apresentam-se os resultados em escala logarítmica da previsão de ENA, com seus limites inferiores e superiores em um intervalo de confiança de 80% e 95%. Observa-se o período de teste entre maio de 2022 a abril de 2023.

```

1
2 Sudeste.final = window(sudeste_dados_log, start = c(1951,0))
3 Sudeste.final.ARIMA = forecast(best.model.ARIMA, h = 12)
4 Sudeste.final.ARIMA
5 ##          Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
6 ## May 2022      10.389938  10.161888  10.617988  10.041166  10.73871
7 ## Jun 2022      10.332898  10.064601  10.601195   9.922573  10.74322
8 ## Jul 2022      10.053007   9.775800  10.330214   9.629056  10.47696
9 ## Aug 2022       9.824193   9.546986  10.101400   9.400242  10.24814
10 ## Sep 2022      9.696615   9.419408   9.973822   9.272664  10.12057
11 ## Oct 2022      9.899858   9.622651  10.177065   9.475906  10.32381
12 ## Nov 2022     10.249790   9.972582  10.526997   9.825838  10.67374
13 ## Dec 2022     10.671547  10.394340  10.948754  10.247595  11.09550
14 ## Jan 2023     10.960656  10.683449  11.237863  10.536704  11.38461
15 ## Feb 2023     11.040826  10.763619  11.318033  10.616874  11.46478
16 ## Mar 2023     11.035826  10.758619  11.313034  10.611875  11.45978
17 ## Apr 2023     10.783407  10.506200  11.060614  10.359456  11.20736

```

Listing 3.7 – Estimativas obtidas pelo modelo ARIMA (escala logarítmica).

Então, retornam-se os dados em sua escala original para que seja possível a comparação com os dados de ENA.

```

1
2 exp(Sudeste.final.ARIMA$mean)
3
4 ##          Jan          Feb          Mar          Apr          May          Jun          Jul
5 ##          Aug
6 ## 2022          32530.66  30727.03  23225.52
7 ##          18475.36
8 ## 2023  57564.20  62369.12  62058.11  48214.12
9 ##          Sep          Oct          Nov          Dec
10 ## 2022  16262.47  19927.54  28276.59  43111.57
11 ## 2023

```

Listing 3.8 – Estimativas obtidas pelo modelo SARIMA em escala original.

Para realizar o cálculo dessas métricas basta fazer a diferença entre um valor observado e sua previsão. Retornando então o erro, que significa a parte imprevisível de uma observação. Segundo o livro citado, o erro é descrito como:

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T},$$

onde os dados de treinamento são $y_1, \dots, y_T$ e os dados de teste são $y_{T+1}, y_{T+2}, \dots$

Com o resultado do modelo podemos então proceder com as métricas de verificação. Existem várias técnicas, que podem medir a precisão da previsão, resumindo os erros de previsão de diferentes maneiras.

As duas medidas dependentes de escala mais usadas são baseadas nos erros absolutos ou quadráticos:

- Erro Absoluto Médio: $MAE = mean(|e_T|)$
- Raiz do Erro Quadrático Médio: $RMSE = \sqrt{mean(e_T^2)}$

O MAE é uma métrica amplamente utilizada para comparação de técnicas de previsão quando aplicadas a uma ou múltiplas séries temporais com as mesmas unidades. Isso se deve à sua facilidade de compreensão e cálculo. Por outro lado, minimizar o RMSE levará a previsões da média, razão pela qual o RMSE também é amplamente utilizado.

O erro percentual é dado por $p_t = 100e_t/y_t$, tendo como vantagem de não ter unidades, e portanto, são frequentemente usados para comparar desempenhos de previsão entre conjuntos de dados. A medida mais utilizada é:

- Erro Percentual Médio Absoluto: $MAPE = mean(|p_t|)$

Essa medida tem a desvantagem de ser infinita ou indefinida se $y_t = 0$ para qualquer t no período de interesse, e tendo valores extremos, se houver y_t próximo de zero. Outra desvantagem é que eles impõem uma penalidade mais pesada aos erros negativos do que aos erros positivos.

Outro erro proposto por Armstrong (1978) é o sMAPE:

- Erro percentual Médio Absoluto Simétrico: $sMAPE = mean(200|y_t - \hat{y}_t|/(y_t + \hat{y}_t))$

No entanto, se y_t é próximo de zero, $\hat{y}_t$ também é provável que seja próximo de zero. Devido a divisão por um número quase zero, a medição permanece instável. Além disso, o valor sMAPE pode assumir valores negativos, o que significa que não é um verdadeiro indicador de "erros percentuais absolutos".

Dito isso calculou-se para o modelo ARIMA, os erros descritos acima através dos dados em escala logarítmica.

```

1
2 erro.SARIMA = Sudeste.final-Sudeste.final.ARIMA$mean
3
4 mean(abs(erro.SARIMA)) # MAE
5 ## [1] 0.164751
6
7 sqrt(mean(abs(erro.SARIMA^2))) # RMSE
8 ## [1] 0.1895735
9
10 pt = 100*erro.SARIMA/Sudeste.final.ARIMA$mean
11 mean(abs(pt)) #MAPE
12 ## [1] 1.582685
13
14 mean( 200*abs(erro.SARIMA)/(Sudeste.final+Sudeste.final.ARIMA$mean)) #sMAPE
15 ## [1] 1.585541

```

Listing 3.9 – Métricas de avaliação obtidas do modelo ARIMA.

O mesmo foi feito para o modelo SSA.

```

1
2 erro.SSA = Sudeste.final-modelo.ssa.final$mean
3 mean(abs(erro.SSA)) #MAE
4 ## [1] 0.2364241
5
6 sqrt(mean(abs(erro.SSA^2))) #RSME
7 ## [1] 0.2623865
8
9
10 pt = 100*erro.SSA/modelo.ssa.final$mean
11 mean(abs(pt)) #MAPE
12 ## [1] 2.300299
13
14 mean( 200*abs(erro.SSA)/(Sudeste.final+modelo.ssa.final$mean)) SMAPE
15 ## [1] 2.283165

```

Listing 3.10 – Métricas de avaliação obtidas do modelo SSA.

4 Resultados e Discussões

A análise exploratória inicial dos dados ocorreu com a produção de gráficos apresentados no Capítulo 3 de modo a explorar as relações entre as variáveis. No entanto, não foram identificadas possíveis correlações entre as variável ENA e o comportamento dos fenômenos e ZCAS, sendo assim, optou-se por realizar o estudo utilizando apenas a ENA como variáveis decisória, fazendo aplicações para a técnica SARIMA e o modelo SSA, obtendo como resultado as seguintes previsões.

Tabela 2 – Previsões dos Modelos.

Modelos								
SARIMA	Mai/22	Jun/22	Jul/22	Ago/22	Set/22	Out/22	Nov/22	Dez/22
	32530	30727	23225	18475	16262	19927	28276	43111
	Jan/23	Fev/23	Mar/23	Abr/23				
	57564	62369	62058	48214				
SSA	Mai/22	Jun/22	Jul/22	Ago/22	Set/22	Out/22	Nov/22	Dez/22
	26387	18430	13591	12010	14078	21244	35700	54606
	Jan/23	Fev/23	Mar/23	Abr/23				
	66285	61771	48613	34574				
Valores Reais								
	Mai/22	Jun/22	Jul/22	Ago/22	Set/22	Out/22	Nov/22	Dez/22
	26960	24225	16810	15856	16114	24562	24616	42681
	Jan/23	Fev/23	Mar/23	Abr/23				
	77841	73925	70963	53417				

Fonte: A autora (2023).

A partir do logarítmico desses valores, fez-se para ambas as técnicas a avaliação utilizando as métricas tabeladas. Como pode-se notar, na Tabela 3 o melhor modelo com os menores erros, ou seja, onde as métricas de avaliação retornaram os menores valores, foi o modelo SARIMA. Sendo assim, escolheu-se este como o melhor modelo para o estudo em questão.

Tabela 3 – Métricas de avaliação para ambos os modelos (escala logarítmica) .

Modelos	MAE	RMSE	MAPE	sMAPE
SARIMA	0.164751	0.1895735	1.582685	1.585541
SSA	0.2364241	0.2623865	2.300299	2.283165

Fonte: A autora (2023).

Com isso, exibem-se agora os valores preditos do modelo SARIMA, o intervalo de confiança da predição e os valores reais da série.

```

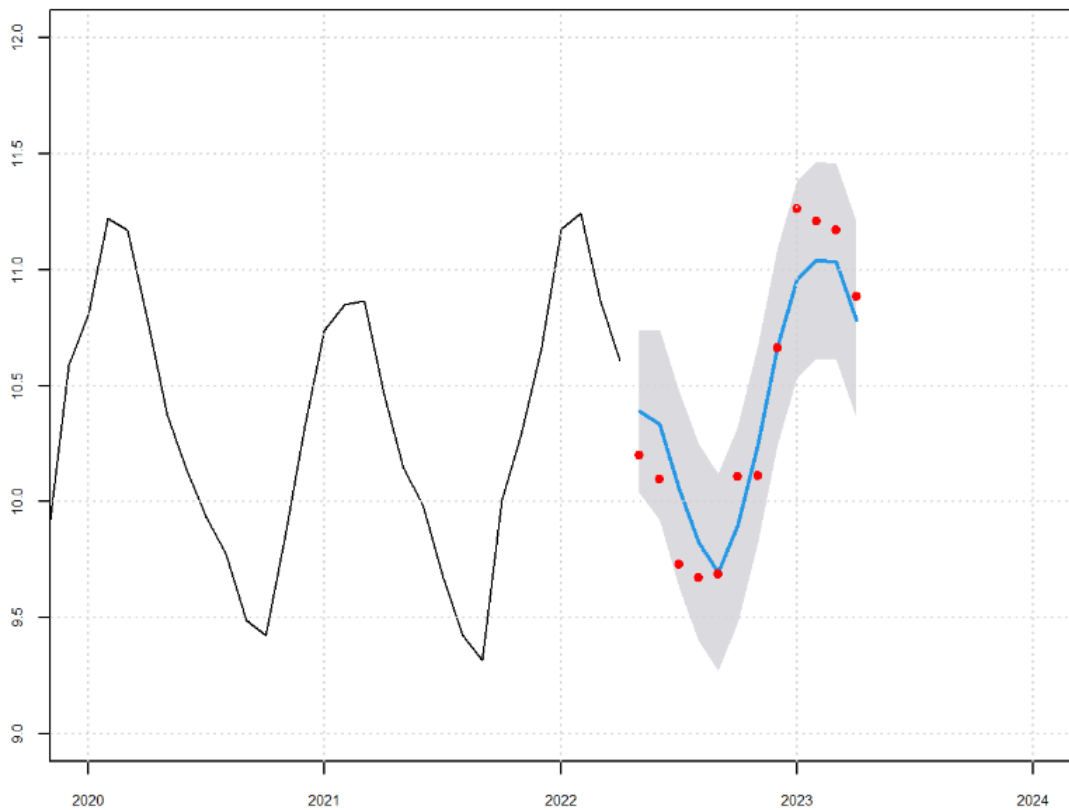
1
2 par(mar=c(3, 6, 1, 1) + 0.1, cex.axis=0.7, cex=0.7)
3 plot(forecast(best.model.ARIMA, h=12, level = 95), pch=16, ylim=c(9,12),
4       xlab="", ylab="",
5       xlim = c(2020,2024), col="black", main="")
6 points(Sudeste.final, col = "red", pch = 19)
7 grid()

```

Listing 4.1 – Estimativas obtidas pelo modelo SARIMA

A seguir, têm-se na Figura 13 os valores reais em escala logarítmica da série e os valores estimados pelo modelo SARIMA, para os últimos 12 meses de dados computados. Também tem-se no gráfico uma banda com 95%.

Figura 13 – Modelo estimado com as bandas de confiança de 95% (escala logarítmica).



Fonte: A autora (2023).

Observa-se então, que os valores reais (em vermelho) ficaram dentro da banda de confiança de 95%, ou seja, um resultado satisfatório para a proposta deste estudo.

```

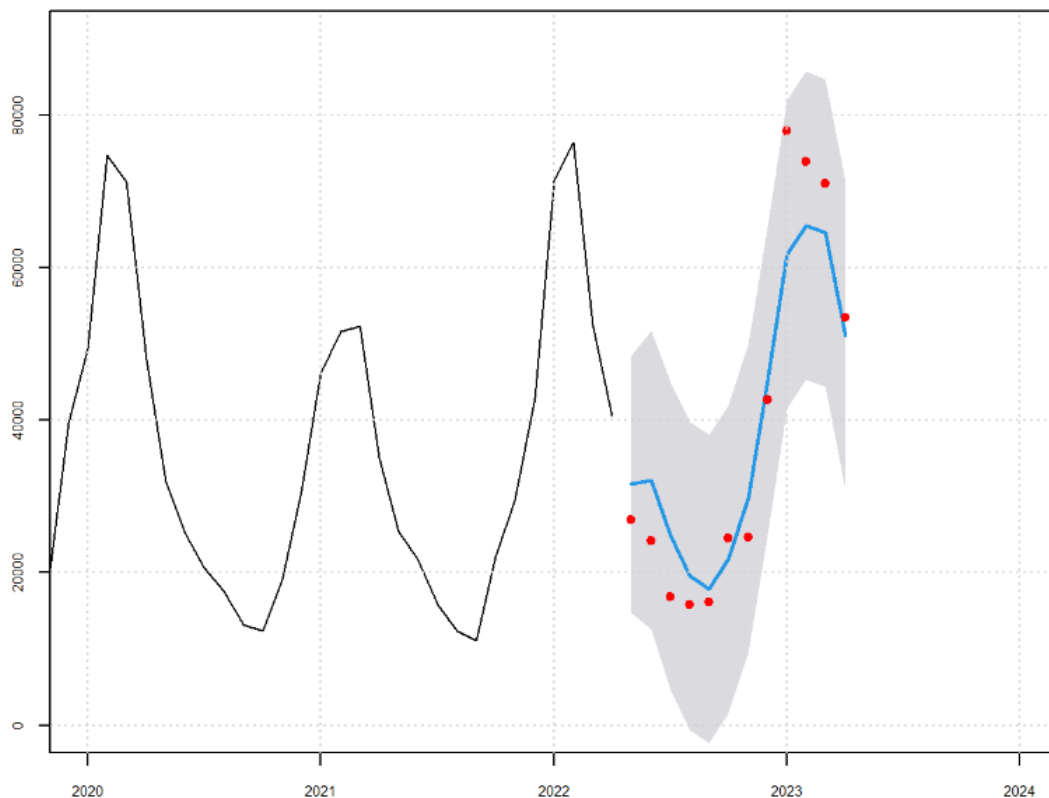
1
2 estimativas = forecast(best.model.ARIMA, h=12, level = 95)
3 estimativas
4 ##          Point Forecast      Lo 95      Hi 95
5 ## May 2022      10.389938  10.041166  10.73871
6 ## Jun 2022      10.332898   9.922573  10.74322
7 ## Jul 2022      10.053007   9.629056  10.47696
8 ## Aug 2022       9.824193   9.400242  10.24814
9 ## Sep 2022       9.696615   9.272664  10.12057
10 ## Oct 2022       9.899858   9.475906  10.32381
11 ## Nov 2022      10.249790   9.825838  10.67374
12 ## Dec 2022      10.671547  10.247595  11.09550
13 ## Jan 2023      10.960656  10.536704  11.38461
14 ## Feb 2023      11.040826  10.616874  11.46478
15 ## Mar 2023      11.035826  10.611875  11.45978
16 ## Apr 2023      10.783407  10.359456  11.20736

```

Listing 4.2 – Estimativas obtidas pelo modelo SARIMA.

Agora, trazendo os valores para a escala original da variável ENA para melhor comparação gráfica.

Figura 14 – Modelo estimado com as bandas de confiança de 95% (escala real).



Fonte: A autora (2023).

E os valores em tabela, para melhor visualização numérica.

Tabela 4 – Previsão do modelo SARIMA com as bandas de confiança.

Limite Inferior	Mai/22	Jun/22	Jul/22	Ago/22	Set/22	Out/22	Nov/22	Dez/22
	22952	20385	15200	12091	10643	13041	18505	28214
	Jan/23	Fev/23	Mar/23	Abr/23				
	37673	40817	40614	31554				
Estimativa	Mai/22	Jun/22	Jul/22	Ago/22	Set/22	Out/22	Nov/22	Dez/22
	32530	30727	23225	18475	16262	19927	28276	43111
	Jan/23	Fev/23	Mar/23	Abr/23				
	57564	62369	62058	48214				
Limite Superior	Mai/22	Jun/22	Jul/22	Ago/22	Set/22	Out/22	Nov/22	Dez/22
	46106	46315	35488	28230	24848	30449	43206	65873
	Jan/23	Fev/23	Mar/23	Abr/23				
	87957	95299	94824	73670				
Valores Reais	Mai/22	Jun/22	Jul/22	Ago/22	Set/22	Out/22	Nov/22	Dez/22
	26960	24225	16810	15856	16114	24562	24616	42681
	Jan/23	Fev/23	Mar/23	Abr/23				
	77841	73925	70963	53417				

Fonte: A autora (2023).

5 Considerações Finais

Observando os resultados encontrados para os modelos propostos na previsão de ENA, para um período de 12 meses, considera-se que os modelos propostos são promissores para este tipo de aplicação. Em especial, o modelo ARIMA que retornou erros menores quando comparado com o modelo SSA.

Pode-se sugerir como continuação para esta pesquisa, a comparação dos resultados com outras técnicas aplicadas à previsões, por exemplo, algumas redes neurais artificiais. Também propõem-se organizar os dados de outra maneira para aplicar os modelos estudados, por exemplo, separando-os em estações do ano, tentando obter um modelo ARIMA para cada período. Por último, sugere-se continuar a busca por outras variáveis que possam ser integradas com a ENA para enriquecer o estudo.

De maneira geral, o trabalho contribuiu para aplicar os conhecimentos aprendidos ao longo do curso de Estatística, tal como uma nova reflexão sobre a pesquisa científica. Foi necessária a leitura de diversos artigos e conversas com profissionais da área para podermos ter uma boa percepção do assunto investigado. Do mesmo modo, é necessário ressaltar a importância da divulgação de resultados, mesmo que a princípio apresentando uma aproximação com um intervalo de confiança alto, implicando em limitações, porém são resultados que podem ajudar na evolução das pesquisas a longo prazo.

Referências

- ANEEL. Instalações de transmissão. 2023. Acessado em 09 de Junho de 2023. Disponível em: <<https://www.ons.org.br/paginas/sobre-o-ons/o-que-e-ons>>.
- ARMSTRONG, J. Previsão de longo prazo: da bola de cristal ao computador. *John Wiley Sons*, p. 348, 1978.
- BAYER F.M.; SOUZA, A. *Revista Brasileira de Biometria*, Wavelets e modelos tradicionais de previsão: um estudo comparativo, v. 28, n. 2, p. 40–61, 2010.
- BNDES, B. N. de Desenvolvimento Econômico e S. Revista do bndes. *REVISTA DO BNDES*, v. 6, n. 12, p. 137–168, DEZ. 1999.
- BRESSAN, A. A. *RAE-eletrônica, SciELO Brasil*, Tomada de decisão em futuros agropecuários com modelos de previsão de séries temporais, v. 3, n. 1, 2004.
- CCEE. Contratos ccee. 2023. Acessado em 09 de Junho de 2023. Disponível em: <<https://www.ccee.org.br/mercado/contratos>>.
- CCEE, C. de Comercialização de energia E. Conceito de preço. 2023. Acessado em 09 de Junho de 2023. Disponível em: <<https://www.ccee.org.br/precos/conceitos-precos>>.
- Câmara de Comercialização de Energia Elétrica (CCEE) . Conceito de preço. 2023. Acessado em 09 de Junho de 2023. Disponível em: <<https://www.ccee.org.br/precos/conceitos-precos>>.
- Esfera Inteligência é Energia. O que É energia natural afluenta? entenda a importância desse dado. Acessado em 09 de Junho de 2023. Disponível em: <<https://blog.esferaenergia.com.br/fontes-de-energia/energia-natural-afluenta>>.
- GAMBONI, L. T. B.
Previsão de Energia Natural Afluenta do Substema Sudeste sob a Abordagem Singular Spectrum Analysis — Universidade Federal Fluminense, 2017.
- HASSANI, H. Singular spectrum analysis based on the minimum variance estimator. *Nonlinear Analysis: Real World Applications*, v. 11, n. 3, p. 2065–2077, 2010.
- HYNDMAN R.J. ATHANASOPOULOS, G. Forecasting: principles and practice. *John Wiley Sons*, v. 2nd edition, 2018. Acessado em 09 de Junho de 2023. Disponível em: <<https://otexts.com/fpp2/accuracy.html>>.
- Instituto Nacional de Meteorologia - INMET. Normais climatológicas do brasil 1991-2020. 2023. Acessado em 09 de Junho de 2023. Disponível em: <https://clima.inmet.gov.br/NormaisClimatologicas/1961-1990/precipitacao_acumulada_mensal_anual>.
- Instituto Nacional de Pesquisas Espaciais - INPE. Condições atuais dos enos: Enfraquecimento do la niña e inicio de condições de neutralidade. 2023. Acessado em 09 de Junho de 2023. Disponível em: <<http://enos.cptec.inpe.br/>>.

Karoline Penteado. Métricas de avaliação para séries temporais. Acessado em 09 de Junho de 2023. Disponível em: <[https://www.alura.com.br/artigos/metricas-de-avaliacao-para-series-temporais#e-a-pergunta-e-como-podemos-medir-a-qualidade-do-ajuste-de-um-modelo-de-previsao-para-uma-serie-temporal?](https://www.alura.com.br/artigos/metricas-de-avaliacao-para-series-temporais#e-a-pergunta-e-como-podemos-medir-a-qualidade-do-ajuste-de-um-modelo-de-previsao-para-uma-serie-temporal?>)>

LACERDA, W. F. R.

ASPECTOS TÉCNICO-ECONÔMICOS DA FLEXIBILIZAÇÃO DE CASOS NO MODELO DECOMP — Escola Politécnica, Universidade Federal do Rio de Janeiro, 2021.

LEDOLTER, J. *IIASA Research Memorandum RM-76-069*, ARIMA Models and their Use in Modelling Hydrologic Sequences, p. 45, 1976.

Mercado Livre de Energia Eletrica. Conceito de preço. 2023. Acessado em 09 de Junho de 2023. Disponível em: <<https://www.mercadolivredeenergia.com.br/pld/>>.

MORETTIN P.A.; TOLOI, C. Análise de series temporais. *São Paulo: Egard Blucher*, 2^a ed., 2006.

National Oceanic and Atmospheric Administration. Climate prediction center: Historical el nino / la nina episodes (1950-2023). 2023. Acessado em 09 de Junho de 2023. Disponível em: <https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php>.

Operador Nacional do Sistema Elétrico. Energia natural afluyente por subsistema. 2023. Acessado em 09 de Junho de 2023. Disponível em: <https://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/energia_afluyente_subsistema.aspx>.

Operador Nacional do Sistema Elétrico. Evolução da capacidade instalada no sin - junho 2023/ dezembro 2027. 2023. Acessado em 09 de Junho de 2023. Disponível em: <<https://www.ons.org.br/paginas/sobre-o-sin/o-sistema-em-numeros>>.

R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2021. Disponível em: <www.R-project.org/>.

SOUZA, G. d. S. e. a. *Revista de Economia e Sociologia Rural, SciELO Brasil*, Previsões para o mercado de carnes, v. 49, n. 2, p. 473–492, 2011.