



UNESP - Universidade Estadual Paulista

“Júlio de Mesquita Filho”

Instituto de Química de Araraquara



Vinícius Ragnini da Silva

**MODELOS DE SÉRIES TEMPORAIS APLICADOS AO CONSUMO DE
ENERGIA ELÉTRICA NA REGIÃO SUDESTE**

Araraquara

2023



UNESP - Universidade Estadual Paulista

Instituto de Química de Araraquara



Vinícius Ragnini da Silva

**MODELOS DE SÉRIES TEMPORAIS APLICADOS AO CONSUMO DE
ENERGIA ELÉTRICA NA REGIÃO SUDESTE**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia Química do Instituto de Química de Araraquara, da Universidade Estadual Paulista, para a obtenção do grau de bacharelado em engenharia química.

**Orientador: Professor Dr. Jorge Manuel
Vieira Capela**

Araraquara

2023

S586a

Silva, Vinícius Ragnini da

Aplicação de Séries de Tempo na modelagem do consumo de energia na região sudeste do Brasil / Vinícius Ragnini da Silva. -- Araraquara, 2023

57 f. : il.

Trabalho de conclusão de curso (Bacharelado - Engenharia Química) - Universidade Estadual Paulista (Unesp), Instituto de Química, Araraquara

Orientador: Jorge Manuel Vieira Capela

1. Análise de séries temporais. 2. Consumo de energia. 3. Suavização | (Estatística). 4. Previsão. 5. Modelagem de informações. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Química, Araraquara. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Vinícius Ragnini da Silva

**MODELOS DE SÉRIES TEMPORAIS APLICADOS AO CONSUMO DE
ENERGIA ELÉTRICA NA REGIÃO SUDESTE**

Trabalho de Conclusão de Curso
apresentado ao Curso de Graduação em
Engenharia Química do Instituto de Química
de Araraquara, da Universidade Estadual
Paulista, para a obtenção do grau de
bacharelado em engenharia química.

**Orientador: Professor Dr. Jorge Manuel
Vieira Capela**

Araraquara, 15 de dezembro de 2023

Banca examinadora

Professor Dr. Jorge Manuel Vieira Capela

Prof. Dr. Elias de Souza Monteiro Filho

Prof. Dr. Ossamu Hojo

AGRADECIMENTOS

A minha mãe, Aguida Aparecida Ragnini, que sempre me apoiou e incentivou, mesmo quando nem ela sabia o que eu deveria fazer para superar meus problemas.

A minha irmã, Sophia Ragnini, por me mostrar que eu poderia dividir não só o carinho dos meus pais, mas sobretudo o mundo com outra pessoa.

Ao restante da família, que mesmo sem saber me tornava mais consciente da minha força a cada abraço, beijo, conselho, discussão e repreensão.

A minha segunda família, República Goteira, que me acolheu mesmo sem saber quem eu era e me mostrou que ali eu teria um lugar seguro para ser quem eu sou.

A meus amigos e amigas de Jundiaí, que são os mesmos desde que me entendo por gente... Obrigado por continuarem até aqui.

Aos muitos amigos e amigas que fiz na UNESP, que pretendo levar para o restante da vida... Obrigado por começarem daqui.

A todos os professores e membros da universidade com quem me relacionei, por todos os ensinamentos, conselhos, concordâncias e desacordos.

Ao professor Jorge Capela, em especial, por confiar em mim como seu orientando para desenvolver um trabalho acerca de um tema que tem tanto apetite.

RESUMO

O presente trabalho tem como objetivo utilizar técnicas de análise de séries temporais e a linguagem de programação R para modelar o consumo de energia na região sudeste do Brasil. Inicialmente, foram coletados dados históricos de consumo de energia na região sudeste do Brasil, abrangendo o período de janeiro de 2004 até julho de 2023. Esses dados foram importados e tratados no Rstudio, software que trabalha com linguagem R. Em seguida, foram exploradas técnicas de visualização gráfica para identificar padrões temporais e sazonalidades. Com base na análise exploratória, foram aplicados modelos de séries temporais adequados para a previsão do consumo de energia na região sudeste, envolvendo métodos como a decomposição da série em componentes, modelagem de tendência e sazonalidade, além da consideração de fatores externos que possam influenciar o consumo. A linguagem R foi utilizada para implementar e ajustar os modelos de séries temporais. A linguagem R é amplamente utilizada para análise estatística e possui pacotes e funções específicos para lidar com séries temporais. O estudo buscou determinar a precisão dos modelos de suavização exponencial com base na avaliação dos parâmetros de suavização e dos limites superiores e inferiores de um intervalo de 95% de confiança.

Palavras – chave: Análise de séries temporais; Consumo de energia; Suavização; Previsão; Modelagem de Informações.

ABSTRACT

The present work aims to use time series analysis techniques and the R programming language to model energy consumption in the southeast region of Brazil. Initially, historical energy consumption data was collected for the southeast region of Brazil, covering the period from January 2004 to July 2023. This data was imported and processed in Rstudio, a software that works with the R language. Next, graphical visualization techniques were explored to identify temporal patterns and seasonalities. Based on the exploratory analysis, suitable time series models were applied to forecast energy consumption in the southeast region. These models involved methods such as series decomposition into components, trend and seasonal modeling, and consideration of external factors that may influence consumption. The R language was used to implement and adjust the time series models. R is widely used for statistical analysis and has specific packages and functions for handling time series data. The study aimed to determine the accuracy of exponential smoothing models based on the evaluation of smoothing parameters and the upper and lower limits of a 95% confidence interval.

Keywords: Time series analysis; Energy consumption; Smoothing; Forecast; Information Modeling.

LISTA DE FIGURAS

Figura 1.1: Sistema Interligado Nacional.....	13
Figura 1.2: Os quatro dos Subsistemas do SIN.....	14
Figura 1.3: Evolução da capacidade instalada no SIN Setembro de 2023/ Dezembro 2027.....	15
Figura 1.4: Extensão da rede básica de distribuição 2022 x projeção 2027.....	17
Figura 4.1- Decomposição da série histórica de 2002 a 2021 do consumo residencial de energia elétrica da região Sudeste (conjunto treino).....	40
Figura 4.2: Ajuste do modelo de suavização exponencial da série temporal de treino de 01/2004 a 12/2021 e da série de teste de 01/2022 a 07/2023 do consumo residencial de energia elétrica da região Sudeste.....	41
Figura 4.3: Ajuste do modelo ARIMA da série temporal de treino de 01/2004 a 12/2021 e da série de teste de 01/2021 a 07/2023.....	44
Figura 4.4 Previsão utilizando o ajuste do modelo ETS da série temporal até 2025....	47

LISTA DE TABELAS

Tabela 4.1: Parâmetros de suavização, acurácia e variância residual usando o modelo ETS.....	42
Tabela 4.2: Valores previstos e limites de 95% de confiança usando Suavização Exponencial (modelo ETS).....	43
Tabela 4.3: Parâmetros, acurácia e variância residual usando o modelo ARIMA(p,d,q)..	45
Tabela 4.4: Valores previstos e limites de 95% de confiança usando ARIMA (modelo auto arima).....	46
Tabela 4.5: Valores previstos até 2025 e limites de 95% de confiança usando modelo ETS.....	48

LISTA DE SÍMBOLOS

x_t : Valor observado na série temporal no instante t

T_t : Estimador da tendência da série temporal no instante t

S_t : Estimador da sazonalidade da série temporal no instante t

C_t : Estimador do componente cíclico da série temporal no instante t

E_t : Estimador dos resíduos da série temporal no instante t

m: É o total de observações em um ciclo completo

N: Total de observações na amostra

Processos Estacionários

h: número de defasagens da série temporal

ρ_h : Autocorrelação defasada em h instantes, onde: $-1 \leq \rho_h \leq 1$

μ : Média da amostra

a_h : autocorrelação parcial defasada h instantes, onde: $-1 \leq a_h \leq 1$

σ^2 : Variância característica de um ruído branco

Suavização Exponencial

α : Fator de suavização

F_t : Previsão da série temporal para o instante t

\hat{x}_{t+1} : Previsão da série temporal para o instante t + 1

β : Fator de suavização da tendência

γ : Fator de suavização da sazonalidade

Modelos ARIMA e SARIMA

ϕ_p : Coeficiente autoregressivo de ordem p

θ_q : Coeficiente de médias móveis de ordem q

X_t : Previsão da série temporal para o instante t

∞ : Intercepção média

Δdx : Série temporal diferenciada d vezes

ΔDs : Série temporal diferenciada D vezes sazonalmente

θ_Q : Coeficiente de médias móveis sazonal de ordem P

ϕ_P : Coeficiente autorregressivo sazonal de ordem Q

σ^2 : Variância residual

\hat{y}_i : Dados originais

y_i : Valor médio do conjunto de dados originais

y_i : Valores gerados pelo modelo

k: Número total de parâmetros a serem estimados

LISTA DE ABREVIATURAS E SIGLAS

ADF – Augmented Dickey-Fuller

AIC – Akaike's Information Criterion

AR – Autoregressive

ARIMA – Autoregressive Integrated Moving Average

ARMA – Autoregressive Moving Average

BIC – Bayesian's Information Criterion

Corr - Correlação

DFT – Discrete Fourier Transform

EPE - Empresa de Pesquisa Energética

FFT – Fast Fourier Transform

I – Integrated

LSE – Least Squares Estimation

MA – Moving Average

MIN – Minimize function

SARIMA – Seasonal Autoregressive Integrated Moving Average

SIN - Sistema Interligado Nacional

SSE – Sum Squared Error

Sumário

1. INTRODUÇÃO.....	12
1.1 A energia elétrica no Brasil e o SIN.....	12
1.1.1 A Usina Hidrelétrica de Itaipu.....	15
1.1.2 A Usina Hidrelétrica de Belo Monte.....	15
1.1.3 Sistemas de transmissão e distribuição.....	15
1.2 O Subsistema SE/CO e o cenário energético nas regiões envolvidas.....	17
1.3 Problemática.....	18
1.4 Objetivo Geral.....	18
1.4.1 Objetivos Pontuais.....	18
2. REVISÃO BIBLIOGRÁFICA.....	20
2.1 Séries Temporais.....	20
2.1.1 Tendência.....	20
2.1.2 Sazonalidade.....	21
2.1.3 Ciclos.....	22
2.1.4 Componentes Não-sistemáticas.....	22
2.2 Estacionariedade.....	22
2.3 Previsão e Séries Temporais.....	23
2.4 Modelos de Previsão.....	24
2.4.1 Suavização Exponencial Simples (SES).....	24
2.4.2 Suavização Exponencial de Holt (SEH).....	26
2.4.3 Suavização Exponencial Sazonal de Holt-Winters (HW).....	27
2.4.4 Modelo ARIMA.....	29
2.4.5 Validação dos métodos de previsão.....	31
3. METODOLOGIA.....	34
3.1 Suavização Exponencial no R.....	34
3.2 Ajuste e Validação de um Modelo ARIMA em R.....	36
3.3 Conjunto de Treino e Conjunto de Teste.....	36
4. RESULTADOS E DISCUSSÃO.....	38
4.1 Decomposição da Série.....	38
4.2 Suavização Exponencial.....	39
4.3 Ajuste de um Modelo ARIMA.....	43
4.4 Previsão pelo Melhor Ajuste.....	46
6. CONCLUSÃO E SUGESTÕES PARA TRABALHOS FUTUROS.....	48
7. REFERÊNCIAS BIBLIOGRÁFICAS.....	50
APÊNDICES.....	53
Modelo ETS.....	53
Modelo ARIMA.....	54

1. INTRODUÇÃO

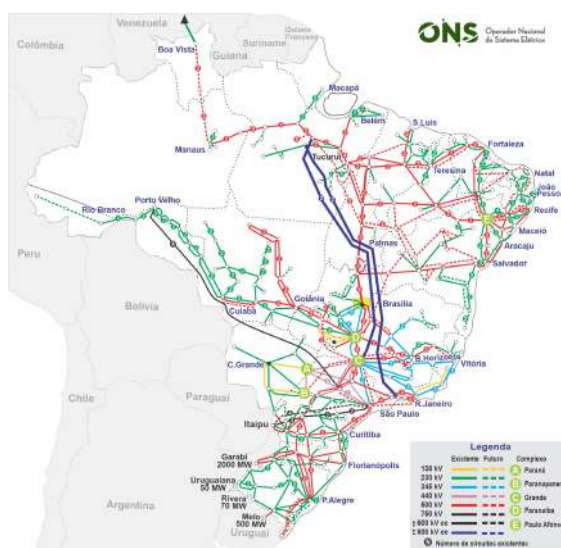
1.1 A energia elétrica no Brasil e o SIN

Se no início dos anos 2000, o Brasil tinha cerca de 180 milhões de habitantes, em 2022 esse número já ultrapassava os 203 milhões (IBGE, 2022 [2]). Este crescimento populacional e industrial resultou em alguns momentos críticos no cenário energético, como a crise do apagão em 2001, a crise energética de 2015 e a crise dos anos 2020-2021, que afetou principalmente a região Norte do país.

Mas o crescimento populacional não é o único motivo pelas crises energéticas do país. Segundo dados do Ministério de Minas e Energia do Brasil, em setembro de 2021, cerca de 61% da matriz elétrica brasileira era proveniente de usinas hidrelétricas, aproveitando a vasta rede de rios do país - percentual que já foi próximo a 90% no início do século XXI. (UFMG, 2023 [3])

Além disso, a dimensão territorial do país torna bastante desafiadora a missão de estabelecer um sistema energético eficiente. Por isso, em 1998, foi criado o Sistema Interligado Nacional (SIN). O SIN consiste em uma estrutura de equipamentos e instalações interligados por meio de conexões elétricas com o propósito de viabilizar o abastecimento de energia do país. (ONS, 2023 [4])

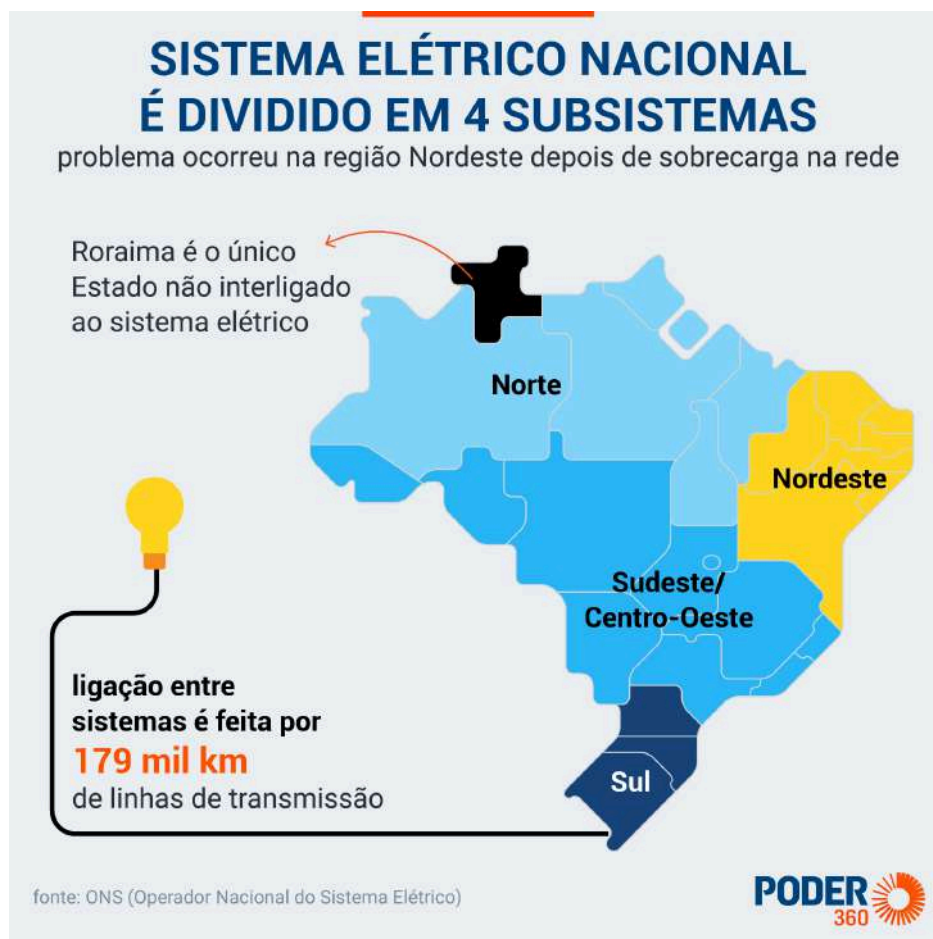
Figura 1.1: Sistema Interligado Nacional



Fonte: ONS (Operador Nacional do Sistema Elétrico), 2023

O SIN é constituído por quatro subsistemas: Subsistema Sul (S); Subsistema Nordeste (NE); Subsistema Norte (N); e o Subsistema Sudeste/Centro-Oeste (SE/CO), objeto de estudo do presente trabalho. (ONS, 2023 [4])

Figura 1.2: Os quatro dos Subsistemas do SIN



Fonte: ONS, 2023

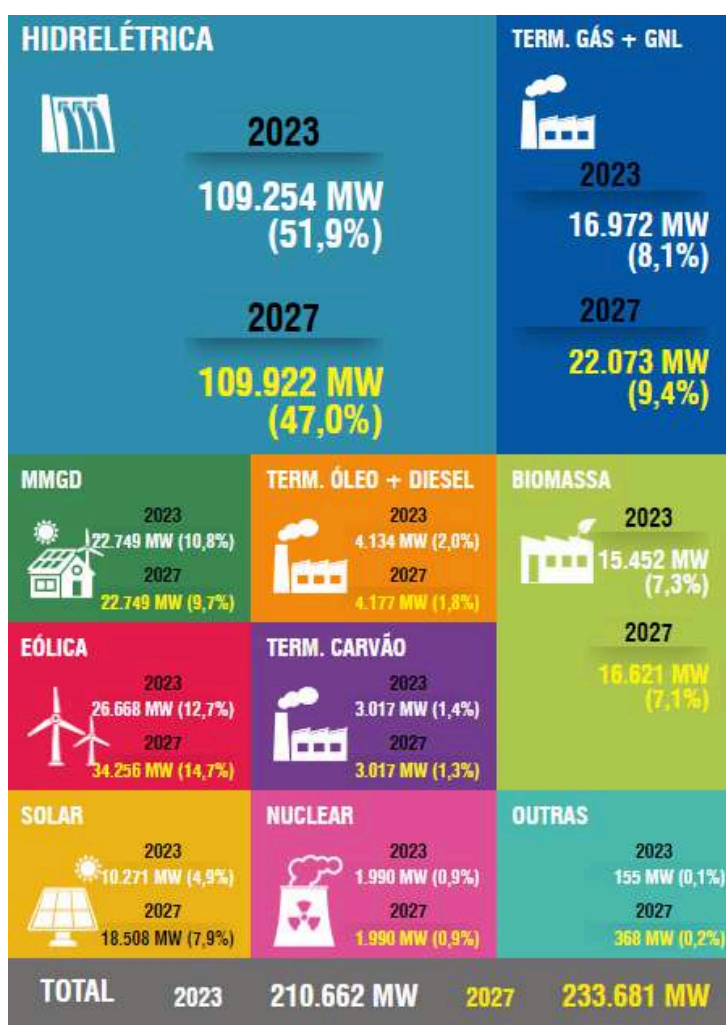
O SIN conta com uma operação interligada de mais de 180 mil quilômetros de extensão, sendo o maior sistema integrado do mundo. A principal vantagem desta operação é o intercâmbio energético entre os quatro subsistemas, buscando assegurar suprimento energético durante todo o ano, na tentativa de contornar a sazonalidade de cada região. Desta forma, o potencial hidrelétrico nacional é melhor aproveitado. Contudo, como alguns subsistemas geram mais energia que outros, em certos casos é gerada uma carência energética. (ONS, 2023 [4])

E esse não é o único problema do Sistema Interligado Nacional: devido à enorme extensão e aos inúmeros aparelhos e processos ligados ao funcionamento

eficaz do sistema, perturbações pontuais podem comprometer a estabilidade dos SIN como um todo. (SANTOS, 2013 [5]).

No tocante à geração de energia, a fonte hidrelétrica ainda é predominante. Dentre as principais usinas hidrelétricas responsáveis por este abastecimento, merecem destaque as usinas de Itaipu e Belo Monte, além daquelas localizadas ao longo dos rios Paraná, Tocantins e São Francisco. (SANTOS, 2013 [5]).

Figura 1.3: Evolução da capacidade instalada no SIN Setembro de 2023/ Dezembro 2027



Fonte: PMO Setembro/2023 *No momento, o valor de MMGD exibido neste painel dinâmico não leva em conta a expansão atualmente em curso. O número apresentado refere-se à capacidade instalada em M-2, conforme a metodologia em vigor, até que a expansão seja devidamente incorporada.(ONS, 2023)

1.1.1 A Usina Hidrelétrica de Itaipu

A Usina de Itaipu, finalizada em 1964, é considerada uma das maiores usinas hidrelétricas do mundo, estando localizada no Rio Paraná, próxima à fronteira do Brasil com o Paraguai. Por conta disso, a usina é operada em conjunto pelos dois países e a sua produção energética é também compartilhada por ambos, recebendo o Brasil a maior parte da produção. (CAVALCANTI, 1976 [6])

Segundo dados do site oficial da Itaipu Binacional, em 2021 a capacidade instalada da Usina de Itaipu era de cerca de 14 mil megawatts (MW). A localização no Rio Paraná favorece esta alta potência de operação, que utiliza turbinas gigantes para transformar a energia potencial das quedas d'água em energia elétrica. Essa energia é significativa para os dois países envolvidos na sua operação.

1.1.2 A Usina Hidrelétrica de Belo Monte

A Usina Hidrelétrica de Belo Monte está localizada no Rio Xingu, no estado do Pará, no Brasil. Sua construção é mais recente em comparação com Itaipu, tendo a primeira unidade geradora entrado em operação em 2016. A usina é projetada para ser uma das maiores hidrelétricas do Brasil e uma das maiores do mundo em capacidade instalada.

A capacidade instalada da Usina de Belo Monte, quando totalmente concluída, é estimada em cerca de 11.200 MW. A usina é composta por várias unidades geradoras e barragens, e a energia é gerada a partir da água do Rio Xingu, que é desviada para alimentar as turbinas. (BERMANN, 2012 [7])

1.1.3 Sistemas de transmissão e distribuição

Dada a dimensão do sistema hidrelétrico, surge o questionamento sobre os meios de transmissão da energia gerada nas usinas. A transmissão é realizada por uma extensa rede de linhas de alta tensão, interligando as subestações e usinas de todo o Brasil. Este sistema de transmissão é monitorado e operado pelo ONS (Operador Nacional do Sistema Elétrico), órgão de direito privado sem fins lucrativos, cuja função é assegurar a estabilidade do sistema elétrico nacional. (ONS, 2023 [4])

Figura 1.4: Extensão da rede básica de distribuição 2022 x projeção 2027

230 kV	2022 67.137 km	2027 75.115 km
345 kV	2022 10.491 km	2027 11.406 km
440 kV	2022 6.934 km	2027 7.027 km
500/525 kV	2022 70.044 km	2027 97.039 km
600 kV	2022 12.816 km	2027 12.816 km
750 kV	2022 2.683 km	2027 2.683 km
800 kV	2022 9.204 km	2027 10.677 km
TOTAL	179.311 km	216.759 km

Fonte: Para a totalização do quantitativo de linhas em operação foram utilizados os totais contabilizados na última reunião de consolidação de quantitativos de linhas de transmissão e subestações concluídos no ano de 2022 no Sistema de Transmissão Nacional realizada em 24/01/2023 coordenada pelo MME com participação do ONS, EPE, CCEE e ANEEL. Para as linhas futuras foram considerados os totais de linhas indicadas no PARPEL 23/27 além das outorgadas ainda não em operação na data de 31/10/2022. Números atualizados em 11/04/2023.

Por fim, a distribuição de energia fica a cargo de concessionárias de distribuição de energia, que atendem áreas urbanas e também rurais. Algumas das principais do mercado que atuam na região Sudeste/Centro-Oeste são a Eletropaulo, Cemig e Light. (ONS, 2023 [4])

1.2 O Subsistema SE/CO e o cenário energético nas regiões envolvidas

O Subsistema SE/CO compreende, além dos estados das regiões Sudeste e Centro-Oeste, Rondônia e Acre. Ele é bastante robusto e interconectado, desempenhando um papel fundamental no atendimento às necessidades energéticas dessas regiões. O sistema elétrico é gerido e regulado por várias instituições e empresas, com destaque para a Agência Nacional de Energia Elétrica (ANEEL) e o Operador Nacional do Sistema Elétrico (ONS).

Falando mais especificamente da região Sudeste, seu sistema elétrico é composto por uma rede básica que abrange mais de 35.000 quilômetros de linhas, operando em tensões de 750, 500, 440, 345 e 230 kV, além de um sistema adicional operando em 138, 88 e 69 kV para as Demais Instalações de Transmissão (DIT). (KAGAN, 2021 [8])

Essa região abrange os estados de São Paulo, Minas Gerais, Rio de Janeiro, Espírito Santo, Goiás, Mato Grosso, Distrito Federal e, a partir de 2009, também inclui os estados do Acre e Rondônia, de acordo com o Plano Decenal de Energia de 2023 (PDE 2023).

A região Sudeste lidera o consumo de energia no Brasil, contribuindo com 48,7% do consumo total em 2022. Nesse contexto, o setor industrial mantém-se como o principal consumidor, correspondendo a 36,2% do consumo, seguido pelo setor residencial, que representa 30%. Em relação ao consumo médio de energia por habitante, cada pessoa na região consumiu uma média de 2.363 kWh em 2022. (ONS, 2023 [4])

A malha elétrica interligada nessa região é a maior do país e desempenha um papel crucial, atendendo aproximadamente 62% da demanda de energia do Sistema Interligado Nacional (SIN) no ano de 2013, conforme informado pelo Operador Nacional do Sistema Elétrico (ONS). Os principais centros de consumo estão situados nas áreas metropolitanas de São Paulo, Rio de Janeiro e Minas Gerais, que estão distantes das principais fontes de geração de energia. Isso resulta na necessidade de uma extensa rede de transmissão de alta tensão para garantir o fornecimento de eletricidade a essas áreas. (CASTRO, 2021 [9])

Assim sendo, o subsistema SE/CO se destaca como o mais extenso do Brasil e desempenha um papel fundamental tanto na demanda quanto na geração de energia elétrica. Ele é responsável por cerca de 60% da produção de energia no SIN no ano de 2013, evidenciando sua importância estratégica no contexto nacional, conforme relatório do ONS.

1.3 Problemática

Ainda que se tratando de uma fonte renovável de energia, as fontes hidrelétricas tornam o setor energético do país refém de mudanças climáticas repentinas (cada vez mais recorrentes em decorrência do aquecimento global) e, principalmente, dos períodos de seca.

Paralelamente, o desenvolvimento de infraestrutura das redes hidrelétricas não acompanhou o crescimento populacional, gerando um *gap* entre oferta e demanda. Isso, sem mencionar os danos consideráveis das hidrelétricas aos ecossistemas aquáticos do país.

Desta forma, analisar o consumo energético dos últimos anos e a partir disso entender os fatores que atuam neste consumo é fundamental para realizar projeções futuras e evitar situações como as crises energéticas já mencionadas no presente trabalho.

1.4 Objetivo Geral

Estudar o comportamento da série do consumo de energia elétrica dos Sistema Elétrico do Sudeste fazendo uso, para tanto, de séries de tempo.

1.4.1 Objetivos Pontuais

Explorando mais a fundo o objetivo geral, podemos enxergar novos objetivos, desta vez mais específicos, a serem alcançados com o presente estudo. São eles:

- i) Apontar as componentes de tendência e sazonalidade evidenciadas na série;

- ii) Analisar e descrever o comportamento das principais componentes da série;
- iii) Encontrar o melhor ajuste da série para um modelo que favoreça a análise de resultados;
- iv) Utilizar do comportamento do modelo de série para realizar previsões de comportamento futuro;
- v) Identificar problemáticas com base na análise de comportamento futuro da série e propor medidas que solucionem essas problemáticas.

2. REVISÃO BIBLIOGRÁFICA

Nesta seção serão descritos os conceitos mais importantes para o entendimento de séries temporais e seus desdobramentos.

2.1 Séries Temporais

As séries temporais são conceitos estatísticos que descrevem o comportamento de um conjunto de dados de determinado fato no tempo. Desta forma, os dados são ordenados em intervalos comumente iguais do tempo, sendo este a única variável do processo. Em contrapartida, o fato analisado e o local onde o fato acontece se mantêm invariáveis.

Em 1966, Wiener definiu séries do tempo como “sequências de dados quantitativos relativos a momentos específicos e estudados segundo sua distribuição no tempo” (WIENER, 1949 [10]). Desta forma, pode-se desprender a noção da vasta aplicabilidade das séries do tempo nas mais variadas áreas de estudo: descrição de performance operacionais em indústrias, mercado financeiro, projeções epidemiológicas e afins.

As séries temporais também são chamadas séries históricas, uma vez que o estudo delas se baseia efetivamente em valores do passado, ou seja, dados históricos, para realização de predições futuras. E, para compreender melhor a dispersão dos dados é necessário conhecer as componentes das séries temporais: sazonalidade, tendência, ciclo e componentes não-sistemáticas, além dos conceitos atrelados a estes. (DEB, 2017 [11])

2.1.1 Tendência

A tendência pode ser definida como a componente que aponta uma inclinação de direção positiva ou negativa dos dados durante um determinado período de tempo. Ou seja, ela pode indicar um crescimento ou declínio do conjunto de dados durante o recorte temporal observado. E, é importante frisar que a tendência não precisa ser

necessariamente linear, podendo ser categorizada como a linha em torno da qual os dados flutuam. (ANTUNES, 2015 [12])

Essa componente pode ser observada pela variação dos valores da série ao longo do tempo, perpetuando-se assim mesmo após a remoção das demais componentes (ciclos, erros e sazonalidade) da análise da série. Levando isso para o cenário do consumo de energia elétrica, a tendência está ligada à mudança na quantidade de consumidores (aumento ou decaimento populacional) e/ou variações no perfil destes consumidores (substituição de ventiladores por ar condicionados, por exemplo). (ANTUNES, 2015 [12])

Para evidenciar a tendência em uma série de tempo, existem algumas técnicas e testes apropriados, tais como Regressão Linear Simples, Teste do Sinal (Cox-Stuart), Teste de sequências (Wald-Wolfowitz) e Teste do coeficiente de correlação de Spearman. (RODRIGUES, 2021 [13])

2.1.2 Sazonalidade

A sazonalidade é outra componente importante de uma série temporal, especialmente quando falamos de energia elétrica. Ela é caracterizada como um padrão não decorrente de fatores econômicos, que é recorrente e repetitivo num dado período de tempo intra-anual, isto é, um período inserido em um ano (e, portanto, menor que um ano). Exemplos de períodos inter-anuais são os períodos diários, mensais, bimestrais, trimestrais e afins. (ANTUNES, 2015 [12])

No contexto do consumo de energia elétrica, o principal exemplo de sazonalidade são as estações do ano, que costumam definir padrões de consumo a depender das condições climáticas referentes a cada estação.

Existem duas categorias de sazonalidade: a aditiva e a multiplicativa. A primeira está ligada a padrões sazonais relativamente constantes sem levar em conta o nível global da série. A multiplicativa, por sua vez, refere-se a quando a dimensão das flutuações varia havendo dependência com o nível global da série. (BASSETO, 2015 [14])

Para atestar a presença de sazonalidade em uma série histórica, existem alguns testes que podem ser realizados. Esses testes podem ser divididos em testes

paramétricos e não-paramétricos. Dentre os testes não-paramétricos, destacam-se o Teste de Kruskal-Wallis e o Teste de Friedman. (PONTES, 2000 [15])

2.1.3 Ciclos

Assim como a sazonalidade, os ciclos são padrões repetitivos ao longo do tempo. Contudo, apesar dessa similaridade, os ciclos se diferenciam das flutuações sazonais, pois ocorrem em períodos de tempo maiores ou iguais a um ano, além de que podem ser causados por fatores econômicos (e geralmente, essas são as maiores causas). (DEB, 2017 [11])

Variações cíclicas são aquelas que exibem oscilações regulares em torno de uma linha reta ou uma curva de tendência, e estão relacionadas às flutuações de longo prazo. Esses ciclos podem ser caracterizados como periódicos ou não, ou seja, podem ou não repetir padrões semelhantes após intervalos de tempo iguais. (ANTUNES, 2015 [12])

2.1.4 Componentes Não-sistemáticas

Diferentemente das componentes anteriores, as componentes não-sistemáticas não representam períodos, inclinações ou padrões bem definidos. Pelo contrário: apresentam variações irregulares, desgovernadas e esporádicas. Na literatura, é possível encontrar como componentes não-sistemáticas definições como “erro”, “ruído”, “noise” ou ainda “componente irregular ou aleatória”. (DEB, 2017 [11])

A importância da componente aleatória está intimamente ligada à capacidade de influenciar a direção da tendência, bem como à extensão dos ciclos presentes. Isso resulta na acurácia das estimativas relacionadas a fenômenos nos quais a componente aleatória é insignificante em comparação com as componentes sistemáticas. (FARIA, 2018 [16])

2.2 Estacionariedade

A estacionariedade é uma característica de séries temporais que se desenvolvem no tempo aleatoriamente ao redor de uma média constante, de forma a apresentar uma espécie de equilíbrio e estabilidade. Em outras palavras, a média da série temporal e as variâncias entre os dados dispostos não mudam com o tempo. (MÜLLER, 1998 [17])

A estacionariedade pode estar presente em todo o período temporal de análise da série, como também em períodos mais enxutos, caracterizando a estacionariedade homogênea. Além disso, é importante ressaltar que a presença da componente tendência em uma série anula a possibilidade desta se apresentar estacionariedade. A presença da tendência contrapõe a premissa de uma série estacionária, já que dependendo da inclinação da direção de tendência, a média dos dados irá variar com o tempo, comprometendo o equilíbrio esperado de uma série estacionária. (DEB, 2017 [11])

A premissa de estacionariedade é fundamental na análise de uma série temporal e, em casos em que não é possível assumir de antemão essa premissa, é necessário transformar os dados para que a série de estudo passe a apresentar comportamento estacionário, conforme a expressão abaixo: (DEB, 2017 [11])

$$y'_t = y_t - y_{t-1}$$

(equação 2.1)

sendo:

- y'_t a diferença; e
- y_t e y_{t-1} observações consecutivas.

2.3 Previsão e Séries Temporais

Nos diversos campos em que as séries temporais podem ser aplicadas, a principal finalidade de sua utilização como ferramenta é a previsão de dados futuros, seja para seguir tendências de mercado, evitar catástrofes naturais ou calcular a receita esperada para determinado período.

Em todos os casos em que a ideia é se realizar uma previsão, dois tipos de análise podem ser empregados: a análise qualitativa e a quantitativa. A análise

qualitativa é baseada no entendimento do analista sobre a matéria de estudo, bem como a experiência na temática. Devido à falta de dados absolutos para embasamento, a confiabilidade desta análise não é consistente. Contudo, pode ser útil em situações onde não há disponibilidade de dados históricos. (SOEIRO, 2021 [18])

Em contrapartida, a análise quantitativa é baseada em dados e valores numéricos, que indicam padrões de comportamento e facilitam a realização de previsões. A previsão por séries temporais é uma análise quantitativa, que se dá a partir da observação gráfica e numérica de particularidades relevantes como tendência, sazonalidade e aleatoriedade. Não obstante, é necessária a aplicação de ferramentas estatísticas, como testes, técnicas e algoritmos para uma previsão confiável. (SOEIRO, 2021 [18])

2.4 Modelos de Previsão

Métodos de suavização exponencial são técnicas empregadas na previsão de séries temporais, baseando-se na ponderação das observações passadas, com maior ênfase nas mais recentes. Essas abordagens são especialmente valiosas para séries temporais que exibem tendências, sazonalidades ou flutuações aleatórias.

Existem diversos métodos de suavização exponencial que dependem do número de parâmetros envolvidos. Por exemplo, o caso da Suavização Exponencial Simples (SES) envolve a utilização de apenas um parâmetro de suavização, denotado por α . O método SES é especialmente adequado para séries temporais que não apresentam tendência ou sazonalidade. A Suavização Exponencial de Holt (SEH) emprega dois parâmetros de suavização, α e β , sendo apropriado para séries temporais com tendência, mas sem sazonalidade. Para o caso de séries temporais que apresentam comportamento sazonal pode ser empregada a Suavização Exponencial de Holt-Winters (HW) que envolve três parâmetros de suavização, α , β e γ . (MORETTIN, 2018 [19]) (MONTGOMERY, 2015 [20]) (MAKRIDAKIS, 2008 [21])

2.4.1 Suavização Exponencial Simples (SES)

Seja $Z_1, Z_2, (\dots), Z_N$ uma série temporal livre de tendência e sazonalidade, composta por seu nível mais um ruído aleatório, isto é

$$Z_t = \mu_t + \epsilon_t, \quad t = 1, 2, (\dots), N$$

onde $E(\epsilon_t) = 0$, $\text{Var}(\epsilon_t) = \sigma_\epsilon^2 e^{\mu_t}$ é um parâmetro desconhecido que pode variar lentamente com o tempo.

No método da suavização exponencial simples (SES) de define-se o valor exponencialmente suavizado \bar{Z}_t por

$$\bar{Z}_t = \alpha Z_t + (1 - \alpha) \bar{Z}_{t-1}, \quad t = 1, 2, (\dots), N$$

(equação 2.2)

onde $Z_0 = Z_1$, e $0 \leq \alpha \leq 1$ representa a constante de suavização.

O valor exponencialmente suavizado Z_t também pode ser obtido utilizando a Eq. (2.3) apresentada a seguir:

$$\bar{Z}_t = \alpha \sum_{k=0}^{t-1} (1 - \alpha)^k Z_{t-k} + (1 - \alpha)^t \bar{Z}_0, \quad t = 1, 2, (\dots), N$$

(equação 2.3)

A expansão da Eq.(2.3) resulta em

$$\bar{Z}_t = \alpha Z_t + \alpha(1 - \alpha)Z_{t-1} + \alpha(1 - \alpha)^2 Z_{t-2} + (\dots) + \alpha(1 - \alpha)^{t-1} Z_1 + (1 - \alpha)^t \bar{Z}_0,$$

$$t = 1, 2, (\dots), N$$

(equação 2.4)

Observa-se que a SES é uma média ponderada que atribui pesos maiores às observações mais recentes. Tais pesos decaem exponencialmente com o passar do tempo.

Previsão

A previsão dos valores futuros é dada pelo último valor exponencialmente suavizado

$$\left\{ \begin{array}{l} \widehat{Z}_t(h) = \bar{Z}_t, \text{ para todo } h > 0 \\ \widehat{Z}_t(h) = \alpha Z_t + (1 - \alpha)\widehat{Z}_{t-1}(h + 1) \end{array} \right.$$

(equação 2.5)

onde h representa o número de passos à frente a ser previsto. Note que utiliza-se apenas o parâmetro de suavização exponencial, sua observação mais recente e a previsão feita no passo anterior.

2.4.2 Suavização Exponencial de Holt (SEH)

O modelo de SES da seção anterior não é adequado para séries de dados que exibem uma tendência linear, sendo que a sua aplicação nesses casos resultaria em previsões que sistematicamente subestimariam ou superestimariam os valores reais da série ao longo do tempo. O modelo de Suavização Exponencial de Holt (SEH), desenvolvido por Holt em 1957, oferece uma solução que permite fazer previsões em séries que apresentam uma tendência linear. (MORETTIN, 2018 [19]) (PELLEGRINI, 2001 [22]) (WINTERS, 1960 [23])

O modelo SEH se diferencia da Suavização exponencial simples pela introdução de um novo parâmetro para lidar com a componente de tendência da série. Seja Z_t uma série temporal não sazonal, composta da soma de nível, tendência e resíduo aleatório com média zero e variância constante σ_ϵ^2 , isto é

$$Z_t = \mu_t + \epsilon_t, \quad t = 1, 2, (\dots), N$$

Os valores do nível e da tendência da série no instante t serão estimados, respectivamente, por

$$\begin{cases} L_t = \alpha Z_t + (1 - \alpha)(L_{t-1} + T_{t-1}), & 0 \leq \alpha \leq 1, t = 2, 3, (\dots), N \\ T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, & 0 \leq \beta \leq 1, t = 2, 3, (\dots), N \end{cases}$$

(equação 2.6)

onde L_t estima o nível da série no tempo t , T_t estima a tendência no tempo t e α e β são, respectivamente, os parâmetros de suavização para o nível e para a tendência.

Previsão

A previsão dos valores futuros Z_{t+h} , com origem em t é dada pelo último valor exponencialmente suavizado

$$\widehat{Z}_t(h) = L_t + h\bar{T}_t, \text{ para todo } h > 0$$

(equação 2.7)

ou seja a previsão é realizada adicionando-se ao valor do nível L_t a tendência multiplicada pelo número de passos h à frente que se deseja prever.

2.4.3 Suavização Exponencial Sazonal de Holt-Winters (HW)

O modelo de Holt-Winters (HW) é especialmente recomendado para séries temporais que exibem padrões sazonais distintos. Winters em 1960 introduziu esta extensão ao modelo original de Holt com o propósito de capturar os efeitos sazonais. A estrutura geral do modelo é a mesma discutida na seção anterior, mas incorpora uma equação adicional para acomodar a sazonalidade. A principal vantagem do modelo HW é a sua capacidade de lidar com séries temporais de natureza mais genérica, uma vez que na prática a maioria das séries apresenta algum grau de

sazonalidade. (MORETTIN, 2018 [19]) (PELLEGRINI, 2001 [22]) (WINTERS, 1960 [23])

Existem dois diferentes métodos de análise por HW, que se diferenciam de acordo com o modo como a sazonalidade será modelada. Se a série tem uma sazonalidade multiplicativa, isso significa que os picos e vales sazonais têm um efeito multiplicativo sobre a média da série. Esse modelo é adequado quando a amplitude das variações sazonais aumenta ou diminui em relação à média ao longo do tempo. Se a série tem uma sazonalidade aditiva, isso significa que os picos e vales sazonais têm um efeito aditivo à média da série. Esse modelo é apropriado quando a amplitude das variações sazonais permanece relativamente constante. É importante identificar corretamente o tipo de sazonalidade presente nos dados para fazer previsões precisas usando o método de Holt-Winters.

Modelo Aditivo

Se a componente sazonal for tratada de forma aditiva, então o modelo será dado por:

$$\left\{ \begin{array}{l} L_t = \alpha(Z_t - S_{t-m}) + (1 - \alpha)(L_{t-1} + T_{t-1}), \quad 0 \leq \alpha \leq 1 \\ T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, \quad 0 \leq \beta \leq 1 \\ S_t = \gamma(Z_t - L_{t-1} - T_{t-1}) + (1 - \gamma)S_{t-m}, \quad 0 \leq \gamma \leq 1 \end{array} \right.$$

(equação 2.8)

$$\hat{Z}_{t+h} = L_t + h\bar{T}_t + S_{t+h-m}$$

(equação 2.9)

sendo $t = m + 1, m + 2, (\dots), N$. Os parâmetros de suavização para o nível, para a tendência e sazonalidade são, respectivamente, α , β e γ . A Eq.(2.9) apresenta a formulação para o modelo de previsão h passos à frente.

Modelo Multiplicativo

Se a componente sazonal for tratada de forma multiplicativa, então o modelo será dado por:

$$\left\{ \begin{array}{l} L_t = \alpha \frac{Z_t}{S_{t-m}} + (1 - \alpha)(L_{t-1} + T_{t-1}), \quad 0 \leq \alpha \leq 1 \\ T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}, \quad 0 \leq \beta \leq 1 \\ S_t = \gamma \frac{Z_t}{L_{t-1} - T_{t-1}} + (1 - \gamma)S_{t-m}, \quad 0 \leq \gamma \leq 1 \end{array} \right.$$

(equação 2.10)

$$\hat{Z}_{t+h} = (L_t + h T_t)S_{t+h-m}$$

(equação 2.11)

sendo $t = m + 1, m + 2, (\dots), N$. As equações para a tendência são as mesmas em ambos os métodos, sendo diferentes as equações que apresentam componente sazonal. Neste caso estas equações aparecem com operações de multiplicação e divisão ao invés de soma e subtração.

2.4.4 Modelo ARIMA

O modelo ARIMA (*AutoRegressive Integrated Moving Average*) é amplamente utilizado na análise de séries temporais, sendo especialmente útil quando os dados exibem tendências, sazonalidades ou autocorrelação. Ele é composto por três componentes principais:

a) AR (*AutoRegressive*): Esta parte do modelo captura a relação linear entre a observação atual e um ou mais valores passados, conhecidos como *lags*. Tem-se como hipótese que a observação atual é uma combinação linear ponderada das observações anteriores. Um modelo autorregressivo de ordem p é um tipo de modelo estatístico que tenta prever o valor de uma variável como função do tempo t usando os valores anteriores dessa mesma variável. Por exemplo, se houver interesse em prever o consumo de energia elétrica em um determinado mês, podemos usar os dados dos meses anteriores como variáveis explicativas. A ordem p indica quantos valores passados são usados na previsão. Um modelo autoregressivo de ordem p pode ser escrito da seguinte forma:

$$Z_t = c + \Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + (\dots) + \Phi_p Z_{t-p} + \epsilon_t$$

(equação 2.12)

onde Z_t é o valor da variável no tempo t , c é uma constante, $\Phi_1, \Phi_2, (\dots), \Phi_p$ são os coeficientes do modelo e ϵ_t é um termo de erro aleatório.

b) I (*Integrated*): A etapa de integração é necessária quando a série temporal não é estacionária, ou seja, quando suas propriedades mudam ao longo do tempo. A diferenciação é aplicada para tornar a série estacionária, subtraindo o valor atual do valor anterior. O número de diferenciações necessárias é indicado pela letra d . O valor d refere-se a quantas vezes seria necessário tirar a derivada de sua tendência de série temporal para obter uma ou constante. Por exemplo, se a tendência for linear, então tem-se $d = 1$. Se a tendência for quadrática, então $d = 2$. A integração ou diferenciação é usada para tornar os dados mais estacionários, ou seja, com média e variância constantes ao longo do tempo.

c) MA (*Moving Average*): Esta parte do modelo considera a relação entre a observação atual e um ou mais valores de erro passados (resíduos) da parte AR. Ela assume que a observação atual é uma combinação linear ponderada dos resíduos passados. Um modelo de média móvel de ordem q é um tipo de modelo estatístico que tenta capturar o comportamento de uma série temporal usando os valores passados de um termo de erro aleatório. Um modelo de média móvel de ordem q pode ser escrito da seguinte forma:

$$Z_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + (\dots) + \theta_q \epsilon_{t-q}$$

(equação 2.13)

onde Z_t é o valor da série no tempo t , μ é uma constante, ϵ_t é um termo de erro aleatório com distribuição normal e média zero e variância σ^2 , e $\theta_i, i = 1, 2, (\dots) q$ são os coeficientes do modelo. Os modelos de média móvel são usados quando há uma dependência entre os erros da série temporal, ou seja, quando o valor do erro no tempo t é influenciado pelo seu valor no tempo anterior $t - 1$ e por um termo aleatório.

O modelo ARIMA é então expresso como ARIMA(p, d, q), onde:

- p : Representa a ordem do componente AR (número de lags usados).
- d : Representa o número de diferenciações necessárias para tornar a série temporal estacionária.
- q : Representa a ordem do componente MA (número de *lags* dos resíduos usados).

Um *lag* ou atraso representa uma observação anterior em relação à observação atual em uma sequência de dados temporais. Eles capturam a dependência temporal nas séries temporais, sendo usados para modelar a autocorrelação nas séries temporais, ou seja, como os valores passados influenciam os valores futuros.

A seleção apropriada dos valores de p , d e q é o passo mais importante no uso do modelo ARIMA e geralmente envolve análise exploratória dos dados e técnicas de otimização. Uma vez que o modelo é ajustado aos dados, ele pode ser usado para fazer previsões futuras com base nas características passadas da série temporal. O ARIMA é amplamente utilizado em previsão financeira, econômica, climatológica, entre outras aplicações.

Modelo SARIMA

O modelo SARIMA (Seasonal Autoregressive Integrated Moving Average) é uma extensão do modelo ARIMA que permite a modelagem de séries temporais com componentes sazonais. O modelo é definido pelos parâmetros $(p,d,q)(P,D,Q)[S]$, onde (p,d,q) são os parâmetros do modelo ARIMA e (P,D,Q) são os parâmetros do modelo sazonal. O número S representa o período sazonal. Portanto, a diferença entre o modelo SARIMA e o modelo ARIMA é que o primeiro inclui termos adicionais para modelar a sazonalidade

2.4.5 Validação dos métodos de previsão

Para avaliar a precisão e validar um modelo, são calculadas métricas que medem o quão distantes os valores previstos estão dos valores reais.

Seja Z_1, Z_2, \dots, Z_N uma série temporal em que a parte de treino é denotada por Z_1, Z_2, \dots, Z_M e a parte de teste por $Z_{M+1}, Z_{M+2}, \dots, Z_N$. O modelo, obtido com base no conjunto de treino, fornecerá os próximos $N - M$ valores previstos e o procedimento de validação consiste em compará-los com os valores de teste.

Os erros são definidos por

$$\epsilon_t = Z_t - \hat{Z}_t, \text{ para } t = M + 1, \dots, N$$

(equação 2.14)

onde Z_t é o valor real e \hat{Z}_t é o valor previsto.

Um dos critérios mais simples para avaliação de um modelo é o erro médio absoluto (Mean Absolute Error), a média da soma dos erros absolutos encontrados:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\epsilon_i|$$

(equação 2.15)

A MAE utiliza a mesma escala de dados, por isso não é indicada para experimentos que estejam utilizando séries de dados com diferentes escalas. Quanto mais próximo de 0 (zero) for, melhores são os resultados.

Uma outra métrica que pode ser empregada para a validação é a raiz quadrada da média da soma dos quadrados dos erros (Root Mean Squared Error):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \epsilon_i^2}$$

(equação 2.16)

O RMSE atribui um maior peso aos erros de maior magnitude, sendo útil em situações onde se procura evitar os maiores erros. Por isso, ele é bastante sensível a outliers e nunca é negativo, sendo que quanto menor for melhores são os resultados.

O erro percentual médio absoluto, definido por

$$MAPE = \frac{1}{n} \sum_{I=1}^n \left| \frac{\epsilon_i}{Z_i} \right|$$

(equação 2.17)

utiliza valores percentuais e por isso é útil em experimentos onde há dados com escalas diferentes.

A acurácia dos métodos de suavização exponencial pode ser avaliada usando diferentes medidas de erro e técnicas de validação cruzada. A validação cruzada, que consiste em dividir os dados em conjuntos de treinamento e teste, ajusta o modelo no conjunto de treinamento e avalia o seu desempenho no conjunto de teste. Esses métodos podem ajudar a escolher o melhor modelo para cada caso e a melhorar as previsões.

3. METODOLOGIA

Os métodos de previsão em séries temporais são técnicas estatísticas que usam os dados históricos de uma variável para estimar o seu valor futuro. Esses métodos são baseados na ideia de que existe um padrão ou uma estrutura na série temporal que pode ser identificado e projetado para o futuro.

Neste trabalho serão considerados os modelos de suavização exponencial, que são métodos não causais, isto é baseiam-se na existência de uma dependência temporal entre as observações da série, ou seja, que o valor da variável no tempo atual depende do seu valor no tempo anterior e de um termo aleatório. Os métodos não causais usam modelos de séries temporais para capturar o padrão da série e depois usam esse padrão para fazer previsões. (MORETTIN, 2018 [19]) (FERREIRA, 2018 [24]) (HYNDMAN, 2018 [26])

Para tanto, foi utilizado o Rstudio, que consiste em uma das versões de software mais avançadas para a programação em linguagem R, pois oferece uma ampla gama de recursos e painéis que são extremamente úteis para a análise de Big Data."

3.1 Suavização Exponencial no R

Uma forma de utilizar os métodos de suavização para realizar ajustes e previsões no R é através da função `HoltWinters()`. (FERREIRA, 2018 [24]) (MCLEOD, 2012 [25])

- No caso SES:

```
ajuste<-HoltWinters(serie,beta=FALSE,gamma=FALSE)
```

- No caso SEH:

```
ajuste<-HoltWinters(serie,gamma=FALSE)
```

- No caso HW:

```
ajuste<-HoltWinters(serie)
```

As previsões para a série temporal serão obtidas através da função `forecast()` do pacote `forecast`:

```
ajuste_prev<-forecast(ajuste,h = n,level=95)
```

onde h é o número de períodos a serem previstos e `level` é o intervalo de confiança para a predição.

Além dessa, para a suavização exponencial no R também pode ser utilizada a função `ETS` para ajustar um modelo de suavização exponencial. O modelo `ETS` no R pode ser especificado de diferentes maneiras, incluindo a combinação de diferentes letras para indicar quais componentes estão presentes no modelo. O parâmetro `model` é uma *string* que pode conter uma combinação de até três letras, cada uma correspondendo a uma parte diferente do modelo. A primeira letra denota o tipo de erro ("A","M" ou "Z"), a segunda letra denota o tipo de tendência ("N","A","M" ou "Z") e a terceira letra denota o tipo de sazonalidade ("N","A","M" ou "Z"). Em todos os casos, "N" significa nenhum, "A" aditivo, "M" multiplicativo e "Z" automaticamente selecionado. (HYNDMAN, 2018 [26]) (HYNDMAN, 2008 [27]) (HYNDMAN, 2020 [28])

A combinação dessas letras determina o modelo de suavização exponencial a ser ajustado aos dados da série temporal. Por exemplo, se for utilizado o valor "AAA" para o parâmetro `model`, isso significa que está sendo ajustado um modelo de suavização exponencial com erros aditivos, tendência linear e sazonalidade aditiva:

```
ajuste <-ets(serie,"AAA")
```

A escolha do modelo depende da natureza dos dados da série temporal que está sendo analisada. Um procedimento é experimentar diferentes combinações de letras no parâmetro `model` para encontrar o modelo que melhor se ajusta aos dados e fornece as previsões mais precisas.

Acurácia em R

Uma forma de medir a acurácia dos modelos de suavização exponencial no R é implementar as fórmulas das medidas de erro definidas nas Eqs. (2.15), (2.16) e (2.17). Se for utilizada a função `ets()` é possível obter esses valores do erro por meio da função `accuracy()` do pacote `forecast`. Como exemplo de uso tem-se:

```
accuracy(ajuste)
```

No caso de uma validação cruzada quando o desempenho do modelo é avaliado no conjunto de teste, tem-se o seguinte uso da função:

```
accuracy(ajuste.teste, teste)
```

sendo *ajuste.teste* o conjunto de valores do modelo aplicado ao conjunto de teste.

3.2 Ajuste e Validação de um Modelo ARIMA em R

A função `auto.arima()` do R tentará encontrar o melhor modelo ARIMA com base no conjunto de treinamento. A sintaxe dessa função é:

```
library(forecast)

ajuste <- auto.arima(treino_data)
```

sendo *treino_data* o conjunto de treino estabelecido anteriormente. Para fazer previsões no conjunto de teste, usando o modelo ajustado *ajuste* utiliza-se, por exemplo, a seguinte sintaxe:

```
forecast <- forecast(ajuste, h = length(test_data))
```

onde o argumento *h* especifica o número de períodos à frente para os quais deseja-se fazer previsões.

A avaliação do modelo, como no caso da suavização exponencial, também pode ser realizada com a ferramenta `accuracy()`.

3.3 Conjunto de Treino e Conjunto de Teste

Para criar um modelo de previsão de séries temporais é conveniente dividir os dados em um conjunto de treino, cujos dados constituirão a base principal para estimar os parâmetros do modelo e em um conjunto de teste cujos dados serão usados para testar o modelo proposto e comparar com as previsões obtidas. Essa abordagem é essencial para avaliar a capacidade do modelo em prever valores futuros com base em informações passadas. A parte de treinamento é usada para treinar o modelo, enquanto a parte de teste é usada para avaliar o desempenho do modelo em

dados que ele não viu durante o treinamento. (HYNDMAN, 2014 [29]) (RASCHKA, 2017 [30])

A divisão entre treino e teste ajuda a verificar se o modelo é capaz de generalizar bem para novos dados e se está apto para fazer previsões precisas. É importante que a parte de teste contenha dados temporais que vêm após os dados de treinamento, para simular um cenário real em que o modelo deve fazer previsões para o futuro com base no que aprendeu com o passado. Um valor de proporção comum utilizado é de 80% para treino e 20% para teste, mas esses valores podem variar de acordo com o contexto.

O conjunto de dados a ser considerado neste será a série histórica, de janeiro de 2004 a novembro de 2022, do consumo de energia elétrica residencial na região Sudeste, disponibilizada no site da Empresa de Pesquisa Energética (EPE) [31]. Para o conjunto de treino será utilizado o período de 2004 até 2021, sendo a série de 2022 em diante utilizada como conjunto de teste.

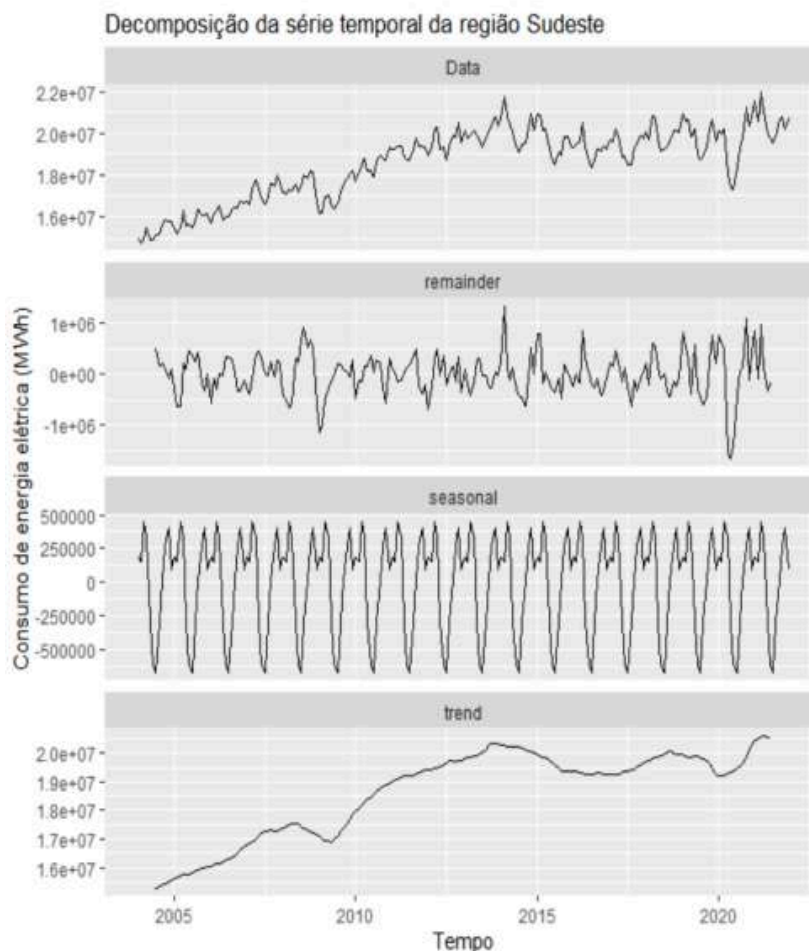
4. RESULTADOS E DISCUSSÃO

4.1 Decomposição da Série

A decomposição de uma série temporal é um processo que separa os componentes de tendência, sazonalidade e aleatoriedade dos dados. No R, isto pode ser realizado usando a função `decompose()`. A decomposição de uma série temporal permite observar separadamente a tendência, sazonalidade e a aleatoriedade dos dados.

Na Figura 4.1 são mostradas a série temporal (*data*), a parte residual da série após a decomposição (*remainder*), a sazonalidade (*seasonal*) e a tendência (*trend*). A parte residual é a diferença entre os dados observados e a soma das componentes de tendência, sazonalidade e aleatoriedade. O painel *trend* no gráfico mostra uma linha crescente, o que indica a série temporal aumentando ao longo do tempo. O painel *seasonal* mostra que existem padrões ou flutuações regulares que ocorrem em intervalos fixos, no caso a cada 12 meses (sazonalidade anual). Também pode ser observado que não há padrão ou estrutura na parte residual que possa indicar uma decomposição incompleta ou que há algum outro componente na série que não foi considerado.

Figura 4.1- Decomposição da série histórica de 2002 a 2021 do consumo residencial de energia elétrica da região Sudeste (conjunto treino).



Fonte: Próprio autor, 2023

4.2 Suavização Exponencial

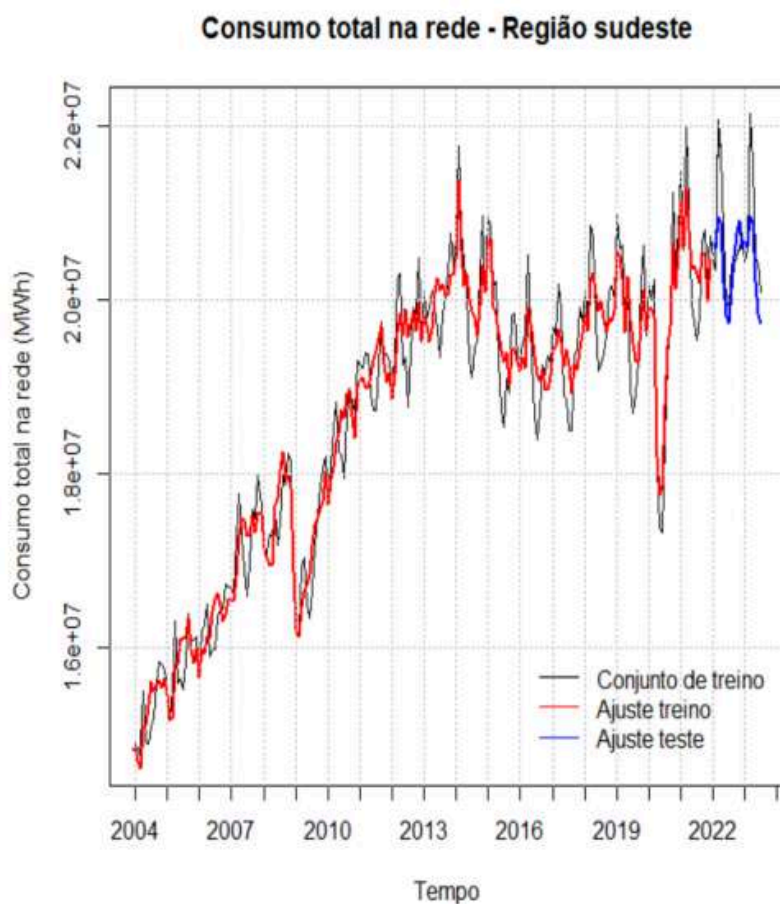
Nesta seção será utilizada a função `ets()` do pacote *forecast* do R. De acordo com a seção anterior, para a série temporal deste trabalho deverá ser escolhido um modelo ETS que considere erro, tendência e sazonalidade. Foi utilizado código `ets(dados, "ZZZ")`, válido no R para ajustar um modelo de suavização exponencial a uma série temporal com componentes de nível, tendência e sazonalidade. A função `ets()` usa o critério de informação de Akaike (AIC) para selecionar o melhor modelo entre os modelos ajustados aos dados. Em outras

palavras, ajusta vários modelos com diferentes combinações de componentes e seleciona o modelo com o menor valor de AIC. (HYNDMAN, 2018 [26]) (HYNDMAN, 2008 [27]) (HYNDMAN, 2020 [28])

A função `forecast()` do pacote `forecast` do R foi utilizada para gerar previsões de valores futuros. A função `accuracy()` foi usada para avaliar a acurácia das previsões em relação aos valores reais do conjunto de teste. (HYNDMAN, 2018 [26]) (HYNDMAN, 2008 [27]) (HYNDMAN, 2020 [28])

A Figura 4.2 mostra o resultado do ajuste do modelo de suavização exponencial da série temporal de treino de 01/2004 a 12/2021 e da série de teste de 01/2022 a 07/2023. O modelo que obteve os melhores resultados foi o de Holt-Winters considerando-se a presença de aleatoriedade, tendência e sazonalidade.

Figura 4.2: Ajuste do modelo de suavização exponencial da série temporal de treino de 01/2004 a 12/2021 e da série de teste de 01/2022 a 07/2023 do consumo residencial de energia elétrica da região Sudeste.



Fonte: Próprio autor, 2023

As estimativas dos parâmetros de suavização e as métricas de acurácia do modelo de suavização exponencial obtidos com a aplicação da função `ets()` do R estão na Tabela 4.1. Os parâmetros α e γ têm o mesmo significado discutido no modelo de Holt-Winters, isto é, estão relacionados respectivamente com a aleatoriedade e sazonalidade. Os parâmetros β e ϕ estão associados à taxa de variação da tendência ao longo do tempo, sendo que o parâmetro ϕ é empregado para os casos de tendência sazonal. Se o valor de ϕ for próximo de 1, a tendência sazonal será mais sensível às mudanças recentes nos dados, enquanto um valor próximo de 0 fará com que a tendência sazonal varie mais lentamente (HYNDMAN, 2018 [26]). Os parâmetros de acurácia são a RMSE (*Root Mean Square Error*), o MAPE (*Mean Absolute Percentage Error*), o AIC (*Akaike Information Criterion*) e o BIC (*Bayesian Information Criterion*) discutidos no capítulo sobre métodos de previsão. Estes últimos são critérios de seleção de modelo que levam em consideração a complexidade do modelo e o ajuste aos dados. O AIC é usado para comparar modelos que diferem em número de parâmetros, enquanto o BIC é usado para comparar modelos que diferem em tamanho da amostra.

Tabela 4.1: Parâmetros de suavização, acurácia e variância residual usando o modelo ETS

α	β	ϕ	γ	RMSE	MAPE	AIC	BIC	s^2
0,767	1×10^{-4}	0,9794	2×10^{-4}	524073	1.937	6789	6850	$5,6 \times 10^{-4}$

Fonte: Próprio autor, 2023

A Tabela 4.2 apresenta os valores da série temporal de teste no período de janeiro de 2022 a julho de 2023, juntamente com os valores previstos pela suavização exponencial e os respectivos limites inferiores (LI-95%) e superiores (LS-95%) de um intervalo de 95% de confiança para os valores previstos. Observa-se que os valores reais pertencem ao intervalo de confiança dos valores previstos, ou seja as previsões podem ser consideradas confiáveis e precisas. Em outras palavras, há uma alta probabilidade de que o valor real do parâmetro populacional esteja dentro do intervalo de confiança.

Tabela 4.2: Valores previstos e limites de 95% de confiança usando Suavização Exponencial (modelo ETS)

Data	Valor real	Previsão	L.Inf. 95%	L.Sup. 95%
jan/22	20474247	20653632	19699057	21608207
fev/22	20328123	20586882	19387850	21785913
mar/22	22068130	20944808	19516732	22372883
abr/22	21718100	20889010	19283646	22494373
mai/22	20788847	20197938	18488348	21907527
jun/22	19924575	19821213	18002098	21640327
jul/22	19948621	19728834	17787559	21670109
ago/22	20409125	20114267	18010231	22218304
set/22	20410074	20494229	18230329	22758130
out/22	20496295	20758939	18350210	23167668
nov/22	20581356	20906285	18369332	23443238
dez/22	20763406	20585126	17982271	23187981
jan/23	20422809	20662568	17948634	23376501
fev/23	20584858	20595605	17793143	23398066
mar/23	22140643	20953499	18006603	23900396
abr/23	21551007	20897499	17865961	23929038
mai/23	20490741	20205977	17187908	23224047
jun/23	20429945	19828940	16784304	22873575
jul/23	20090219	19736367	16625599	22847134

Fonte: Próprio autor

Os resultados obtidos indicam que a série temporal exhibe flutuações imprevisíveis ao longo do tempo (aleatoriedade) e um comportamento que inclui uma direção geral (tendência) consistente dos dados ao longo do tempo, juntamente com padrões regulares que se repetem em intervalos fixos, geralmente relacionados a ciclos sazonais.

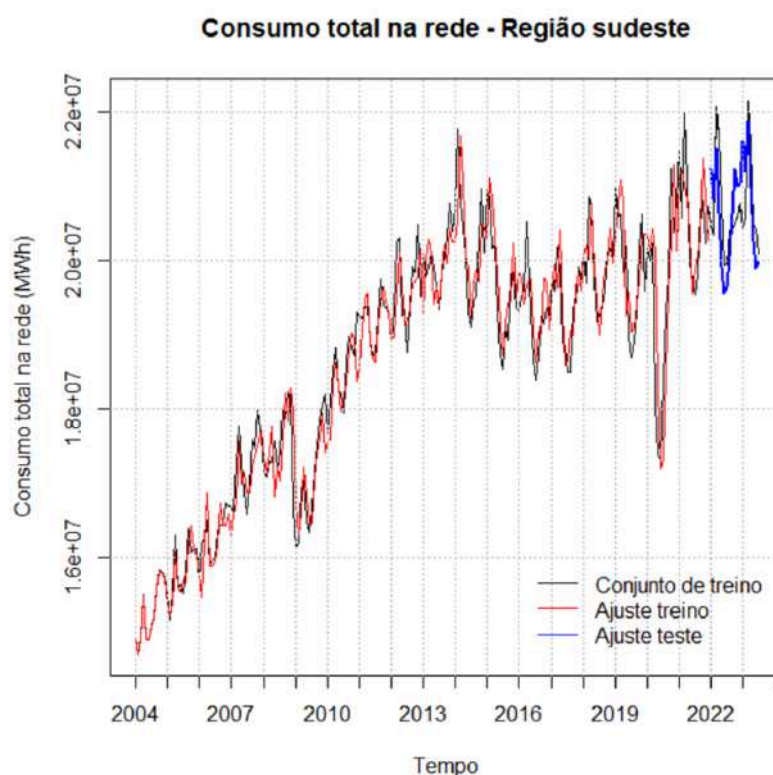
4.3 Ajuste de um Modelo ARIMA

Nesta seção será utilizada a função `auto.arima()` do pacote `forecast` do R. A função `auto.arima()` do R usa uma variação do algoritmo desenvolvido por Hyndman e Khandakar que combina testes de raiz unitária, minimização de AICc e MFV para obter um modelo ARIMA.. (HYNDMAN, 2018 [26]) (HYNDMAN, 2008 [27]) (HYNDMAN, 2020 [28])

Novamente, a função `forecast()` do pacote `forecast` do R foi utilizada para gerar previsões de valores futuros. A função `accuracy()` foi usada para avaliar a acurácia das previsões em relação aos valores reais do conjunto de teste. (HYNDMAN, 2018 [26]) (HYNDMAN, 2008 [27]) (HYNDMAN, 2020 [28])

A Figura 4.3 mostra o resultado do ajuste do modelo de suavização exponencial da série temporal de treino de 01/2004 a 12/2021 e da série de teste de 01/2022 a 07/2023.

Figura 4.3: Ajuste do modelo ARIMA da série temporal de treino de 01/2004 a 12/2021 e da série de teste de 01/2021 a 07/2023.



Fonte: Próprio autor, 2023.

As estimativas dos parâmetros de suavização e as métricas de acurácia do modelo ARIMA obtidos com a aplicação da função `auto.arima()` do R estão na Tabela 4.3. Os parâmetros são os mesmos discutidos na seção anterior, quando foi exposto na Tabela 4.1 os parâmetros para a suavização exponencial.

Tabela 4.3: Parâmetros, acurácia e variância residual usando o modelo ARIMA(p,d,q)

α	β	ϕ	γ	RMSE	AIC	MAPE	AIC
1	0	1	596249	2,519	5901	5917	2×10^{11}

Fonte: Próprio autor, 2023.

No modelo ARIMA ajustado o parâmetro p é igual a 1, o parâmetro d é 0 e o parâmetro q é 1. Isso indica a presença de um termo auto regressivo de ordem 1, um termo de média móvel de ordem 1. A variância dos erros de previsão do modelo também é mostrada na tabela.

A Tabela 4.4 apresenta os valores da série temporal de teste no período de janeiro de 2022 a julho de 2023, juntamente com os valores previstos pelo modelo ARIMA e os respectivos limites inferiores (LI-95%) e superiores (LS-95%) de um intervalo de 95% de confiança para os valores previstos.

Podemos analisar que os valores reais estejam dentro do intervalo de confiança das previsões, o que indica que as previsões são confiáveis e precisas. Ou seja, há uma alta probabilidade de que o valor real do parâmetro populacional esteja contido no intervalo de confiança.

Tabela 4.4: Valores previstos e limites de 95% de confiança usando ARIMA (modelo auto arima)

Data	Valor real	Previsão	L.Inf. 95%	L.Sup. 95%
jan/22	20474247	21223176	20346734	22099617
fev/22	20328123	20822738	19760274	21885202
mar/22	22068130	21513353	20319635	22707071
abr/22	21718100	20632787	19341238	21924336
mai/22	20788847	19910453	18543806	21277100
jun/22	19924575	19545869	18120518	20971220
jul/22	19948621	19636084	18164285	21107884
ago/22	20409125	20048733	18539866	21557600
set/22	20410074	20621078	19082444	22159712
out/22	20496295	21222145	19659495	22784794
nov/22	20581356	20992577	19410482	22574672
dez/22	20763406	21035429	19437545	22633312
jan/23	20422809	21599682	19926294	23273070
fev/23	20584858	21190649	19473454	22907844
mar/23	22140643	21873475	20121131	23625819
abr/23	21551007	20985851	19205167	22766535
mai/23	20490741	20257122	18453503	22060742
jun/23	20429945	19886744	18064509	21708979
jul/23	20090219	19971709	18134331	21809086

Fonte: Próprio autor, 2023.

Os valores de acurácia e variância residual (Tabelas 4.1 e 4.3) mostram que o modelo ETS de suavização exponencial apresentou melhores resultados de previsão. Isso também pode ser observado ao comparar os valores previstos com seus

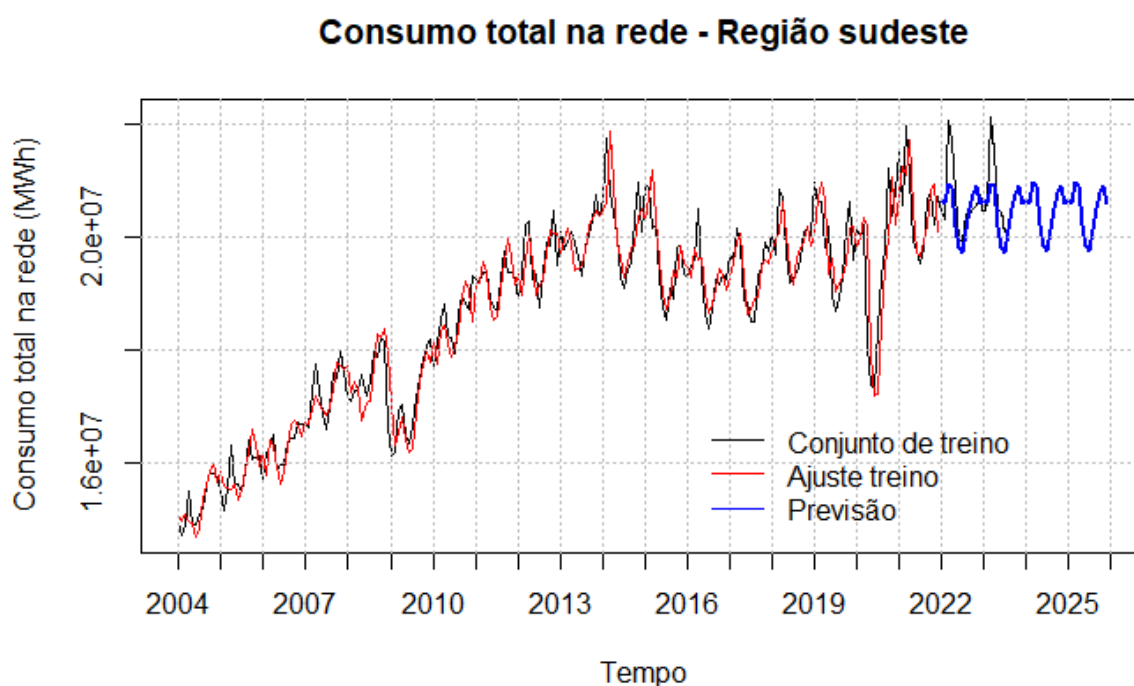
respectivos intervalos de confiança nas Tabelas 4.2 e 4.4, bem como nas Figuras 4.2 e 4.3.

4.4 Previsão pelo Melhor Ajuste

Novamente, função `forecast()` do pacote `forecast` do R foi utilizada para gerar previsões de valores futuros, desta vez até dezembro de 2025. Foi utilizado o modelo `ets()` para realizar esta previsão, já que ele se mostrou mais eficiente para esta série temporal do que o modelo `auto.arima()`.

A Figura 4.4 mostra o resultado do ajuste do modelo de suavização exponencial da série temporal de treino de 01/2004 a 12/2021 e da série de previsão de 01/2022 a 12/2025.

Figura 4.4 Previsão utilizando o ajuste do modelo ETS da série temporal até 2025.



A Tabela 4.5 apresenta os valores da série temporal de previsão no período de janeiro de 2022 a dezembro de 2025, juntamente com os valores previstos pela suavização exponencial e os respectivos limites inferiores (LI-95%) e superiores (LS-95%) de um intervalo de 95% de confiança para os valores previstos.

Tabela 4.5: Valores previstos até 2025 e limites de 95% de confiança usando modelo ETS

Data	Previsão	L.Inf. 95%	L.Sup. 95%
ago/23	20121789	16870383	23373195
set/23	20501735	17109452	23894018
out/23	20766385	17251609	24281160
nov/23	20913629	17296372	24530887
dez/23	20592208	16955744	24228672
jan/24	20669530	16945774	24393286
fev/24	20602401	16818766	24386036
mar/24	20960272	17039025	24881519
abr/24	20904115	16922936	24885293
mai/24	20212242	16295877	24128606
jun/24	19834961	15927071	23742850
jul/24	19742236	15789307	23695165
ago/24	20127649	16033995	24221304
set/24	20507583	16272862	24742304
out/24	20772186	16419140	25125233
nov/24	20919352	16472225	25366479
dez/24	20597727	16157549	25037904
jan/25	20674955	16157261	25192649
fev/25	20607697	16044839	25170556
mar/25	20965549	16263318	25667779
abr/25	20909269	16160416	25658122
mai/25	20217123	15568864	24865382
jun/25	19839652	15223278	24456026
jul/25	19746809	15098012	24395606
ago/25	20132216	15338215	24926217
set/25	20512140	15572769	25451511
out/25	20776707	15718608	25834807
nov/25	20923811	15775061	26072560
dez/25	20602027	15479012	25725042

Fonte: Próprio autor, 2023.

6. CONCLUSÃO E SUGESTÕES PARA TRABALHOS FUTUROS

Com a realização do presente estudo, pudemos observar a eficiência da análise de séries temporais para aplicação de métodos de previsão. Ambos os modelos testados, suavização exponencial e ARIMA, se mostraram ferramentas confiáveis para prever dados futuros para o conjunto de dados estudado.

Mesmo com a validação de ambos os modelos para o propósito de previsão, observou-se que o modelo de suavização exponencial se mostrou ligeiramente mais eficiente e confiável que o modelo ARIMA para o conjunto de dados, após análise dos parâmetros de validação dos métodos de previsão.

Mais adiante, podemos comprovar que a utilização do R como ferramenta para estudo de séries temporais e previsão de valores futuros é bastante interessante, posto que a linguagem oferece mecanismos didáticos e funções simplificadas de aplicação dos métodos temporais em comparação com a aplicação manual/analítica deste conceitos.

Com os valores previstos pelo modelo de suavização exponencial, foi possível observar que os valores de consumo na região sudeste vão aumentar gradativamente ao longo dos anos, com relação aos mesmos meses de anos anteriores.

Para evitar novos problemas no SIN, especialmente na região sudeste, algumas medidas podem ser tomadas, tais quais a aplicação de políticas governamentais que influenciam diretamente no consumo em meses críticos (como a restauração do horário de verão) ou ainda a utilização de novas tecnologias na geração de energia, como painéis fotovoltaicos e usinas virtuais.

Por fim, a fim de complementar a análise realizada pelo estudo, seria interessante a aplicação dos mesmos métodos para outros estados e subsistemas de geração de energia do país, como o Centro-Oeste, que faz parte do Subsistema SE/CO.

Uma análise mais robusta deste nível contribuiria para entender melhor os desafios naturais de cada região, levando em conta aspectos socioeconômicos, políticos e ambientais, para entender comportamentos e traçar planos de ações mais assertivos para cada localidade.

7. REFERÊNCIAS BIBLIOGRÁFICAS

[1] ZUMBA, Francisco Daniel da Silva et al. Adequação de instalação elétrica residencial às normas de segurança e serviço. 2022.

[2] Censo 2022 já contou quase 60 milhões de pessoas no país. Disponível em: <<https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/34728-censo-2022-ja-contou-quase-60-milhoes-de-pessoas-no-pais>>. Acesso em: 18 jun. 2023.

[3] UNIVERSIDADE FEDERAL DE MINAS GERAIS. Em meio ao risco de novo apagão, 'Outra estação' aborda fontes alternativas de energia. Disponível em: <<https://ufmg.br/comunicacao/noticias/outra-estacao-aborda-fontes-alternativas-de-energia-eletrica-para-evitar-novo-apagao-no-brasil>>. Acesso em: 18 jun. 2023.

[4] O que é o SIN. Disponível em: <<https://www.ons.org.br/paginas/sobre-o-sin/o-que-e-o-sin>>. Acesso em: 21 jun. 2023.

[5] SANTOS, Bernardo Pereira Salotto dos. Reprodução de perturbações no Sistema Interligado Nacional a partir de dados de medição fasorial sincronizada. 2013.

[6] CAVALCANTI, José Costa. A Itaipu Binacional - um exemplo de cooperação internacional na América Latina. Revista de Administração Pública, v. 10, n. 1, p. 19 a 68-19 a 68, 1976.

[7] BERMANN, Célio. O projeto da Usina Hidrelétrica Belo Monte: a autocracia energética como paradigma. Novos Cadernos NAEA, v. 15, n. 1, 2012.

[8] KAGAN, Nelson; DE OLIVEIRA, Carlos César Barioni; ROBBIA, Ernesto João. Introdução aos sistemas de distribuição de energia elétrica. Editora Blucher, 2021.

[9] CASTRO, Cristiano Souza. Análise da série consumo de energia elétrica no subsistema sudeste/centro-oeste e no estado de São Paulo. 2021.

[10] WIENER, Norbert. Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications. The MIT press, 1949.

[11] DEB, C. et al. A review on time series forecasting techniques for building energy consumption. v. 74, p. 902–924, 2017. ISSN 13640321.

[12] ANTUNES, José Leopoldo Ferreira; CARDOSO, Maria Regina Alves. Uso da análise de séries temporais em estudos epidemiológicos. *Epidemiologia e Serviços de Saúde*, v. 24, p. 565-576, 2015.

[13] RODRIGUES, Julia Naves. *Inflação no Brasil: uma aplicação de Séries Temporais e Redes Neurais Recorrentes*. 2021.

[14] BASSETO, Ana Laura Canassa. *Previsão de demanda em uma empresa de produção de peças para implementos agrícolas*. 2015. Trabalho de Conclusão de Curso. Universidade Tecnológica Federal do Paraná.

[15] PONTES, Antonio Carlos Fonseca. *Obtenção dos níveis de significância para os testes de Kruskal-Wallis, Friedman e comparações múltiplas não-paramétricas*. 2000. Tese de Doutorado. Universidade de São Paulo.

[16] DE REZENDE FARIA, Ruan et al. ANÁLISE DE DESEMPENHO DO FILTRO DE PARTÍCULAS SIR USANDO A AUTOCORRELAÇÃO DO RESÍDUO. XXI Encontro Nacional de Modelagem Computacional e IX Encontro de Ciência e Tecnologia de Materiais, 2018.

[17] MÜLLER, Ingrid Illich; KRÜGER, Cláudio Marchand; KAVISKI, Eloy. Análise de estacionariedade de séries hidrológicas na bacia incremental de Itaipu. *Revista Brasileira de Recursos Hídricos*, v. 3, n. 4, p. 51-71, 1998.

[18] DE FIGUEIREDO SOEIRO, Joelma et al. Modelos de séries temporais aplicados ao consumo de energia elétrica do estado do Amapá (pp. 127). *Revista da Estatística da Universidade Federal de Ouro Preto*, v. 2, 2012.

[19] MORETTIN, P. A.; TOLOI, C. M. C. *Análise de séries temporais*. São Paulo: Editora Blucher, 2018.

[20] MONTGOMERY, Douglas C.; JENNINGS, Cheryl L.; KULAHCI, Murat. *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.

[21] MAKRIDAKIS, Spyros; WHEELWRIGHT, Steven C.; HYNDMAN, Rob J. *Forecasting methods and applications*. John Wiley & Sons, 2008.

[22] PELLEGRINI, Fernando R.; FOGLIATTO, Flávio S. Passos para implantação de sistemas de previsão de demanda: técnicas e estudo de caso. *Production*, v. 11, p.43-64, 2001.

[23] WINTERS, Peter R. Forecasting sales by exponentially weighted moving averages. *Management science*, v. 6, n. 3, p. 324-342, 1960.

[24] FERREIRA, Pedro Guilherme Costa et al. *Análise de séries temporais em R: Curso introdutório*. São Paulo, SP: GEN Atlas, 2018.

[25] MCLEOD, A. Ian; YU, Hao; MAHDI, Esam. Time series analysis with R. In: *Handbook of statistics*. Elsevier, 2012. p. 661-712.

[26] HYNDMAN, Rob J.; ATHANASOPOULOS, George. *Forecasting: principles and practice*. OTexts, 2018.

[27] HYNDMAN, Rob J.; KHANDAKAR, Yeasmin. Automatic time series forecasting: the forecast package for R. *Journal of statistical software*, v. 27, p. 1-22, 2008.

[28] HYNDMAN, Rob J. et al. Package `forecast'. Online] <https://cran.r-project.org/web/packages/forecast/forecast.pdf>, 2020.

[29] HYNDMAN, Rob J. Measuring forecast accuracy. *Business forecasting: Practical problems and solutions*, p. 177-183, 2014.

[30] RASCHKA, Sebastian; MIRJALILI, Vahid. *Python machine learning second edition*. 2017.

[31] EMPRESA DE PESQUISA ENERGÉTICA. Consumo Mensal de Energia Elétrica por Classe (regiões e subsistemas). Disponível em: <<https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/consumo-de-energia-eletrica>>. Acesso em: 1 mar. 2023.

APÊNDICES

Modelo ETS

```
library(forecast)

library(ggplot2)

library(ggfortify)

dados<- read.table("Dados.txt",header=F,sep="\t",dec=",")

teste <-
read.table("Teste.txt",header=F,sep="\t",dec=",")

dados.full <-
read.table("Dados_full.txt",header=F,sep="\t",dec=",")

dados.fullts<-ts(dados.full,start=c(2004,1),end=c(2023,7)
,frequency=12)

dados.ts<-ts(dados,start=c(2004,1),end=c(2021,12),frequen
cy=12)

teste.ts<-ts(teste,start=c(2022,1),end=c(2023,7),frequenc
y=12)

se<-dados.ts

ajuste<-ets(se,"ZZZ")

ajuste_prev<-forecast(ajuste,h=36,level=95)

print(ajuste_prev)

acuracia<-accuracy(ajuste_prev$mean,teste.ts)

print(acuracia)
```

```

plot(dados.fullts,main="Consumo total na rede - Região
sudeste",xlab="Tempo",ylab="Consumo total na rede
(MWh)",xaxt="n", xlim=c(2004,2025))

lines(teste.ts,lwd=1,lty=2,col="black")

axis(1, at = seq(from = 2004, to = 2026, by = 1), labels
= seq(from = 2004, to = 2026, by = 1), las = 1)

lines(fitted(ajuste),lwd=1,lty=1,col="red")

lines(ajuste_prev$mean,lwd=2,lty=1,col="blue")

abline(v=seq(2004,2026,1),h=seq(1.6e+07,2.2e+07,0.2e+7),l
ty=3,col="gray")

legend(2016,1.75e+07,c("Conjunto de treino ","Ajuste
treino","Previsão"),lwd=c(1,1,1),lty=c(1,1,1),col=c("black","r
ed","blue"),bty='n')

```

Modelo ARIMA

```

library(forecast)

library(ggplot2)

library(ggfortify)

dados<- read.table("Dados.txt",header=F,sep="\t",dec=",")

teste <-
read.table("Teste.txt",header=F,sep="\t",dec=",")

dados.full <-
read.table("Dados_full.txt",header=F,sep="\t",dec=",")

dados.fullts<-ts(dados.full,start=c(2004,1),end=c(2023,7)
,frequency=12)

```

```

dados.ts<-ts(dados,start=c(2004,1),end=c(2021,12),frequen
cy=12)

teste.ts<-ts(teste,start=c(2022,1),end=c(2023,7),frequenc
y=12)

se<-dados.ts

ajuste <- auto.arima(dados.ts, stepwise = FALSE,
approximation = FALSE, trace = TRUE)

ajuste_prev<-forecast(ajuste,h=36,level=95)
print(ajuste_prev)

acuracia<-accuracy(ajuste_prev$mean,teste.ts)
print(acuracia)

plot(dados.fullts,main="Consumo total na rede - Região
sudeste",xlab="Tempo",ylab="Consumo total na rede
(MWh)",xaxt="n", xlim=c(2004,2025))

lines(teste.ts,lwd=1,lty=2,col="black")

axis(1, at = seq(from = 2004, to = 2026, by = 1), labels
= seq(from = 2004, to = 2026, by = 1), las = 1)

lines(fitted(ajuste),lwd=1,lty=1,col="red")

lines(ajuste_prev$mean,lwd=2,lty=1,col="blue")

abline(v=seq(2004,2026,1),h=seq(1.6e+07,2.2e+07,0.2e+7),l
ty=3,col="gray")

legend(2016,1.75e+07,c("Conjunto de treino ","Ajuste
treino","Previsão"),lwd=c(1,1,1),lty=c(1,1,1),col=c("black","r
ed","blue"),bty='n')

```

