**Análise e Previsão de Carga Elétrica na Argentina: Modelos Estatísticos de Séries Temporais e Regressões**

Pedro de Faria Neves Duarte; Gustavo Lobo

**Análise e Previsão de Carga Elétrica na Argentina: Modelos Estatísticos de Séries Temporais e Regressões**

Sumário

Este estudo se dedica à análise e previsão da carga elétrica na Argentina, um tema de extrema relevância para a gestão eficaz da infraestrutura elétrica no país, bem como para operações entre o Brasil e a Argentina, incluindo a comercialização de energia elétrica entre ambos os países. Neste estudo, a exploração abrange a utilização de modelos de séries temporais, incluindo ARIMA (Auto Regressive Integrated Moving Average) e ETS (Error Trend and Seasonality, ou alisamento exponencial), além da aplicação de modelos de regressão para analisar a relação entre as variáveis carga e temperatura. O trabalho se inicia com uma introdução que apresenta o problema e a relevância da previsão de carga no contexto argentino. Em seguida, são detalhados os materiais e métodos utilizados, incluindo a extração, coleta, tratamento e aplicação dos modelos mencionados a esses dados. Os resultados obtidos são apresentados e discutidos em profundidade, com ênfase em estatísticas descritivas, modelos de séries temporais, análise de correlação e insights derivados dessas análises. Por fim, são resumidas as conclusões essenciais do estudo, enfatizando o impacto potencial da previsão de carga na comercialização de energia entre o Brasil e a Argentina, além de compartilhar as lições aprendidas ao longo da pesquisa.

**Palavras-Chave:** Previsão de Carga, Séries Temporais, ARIMA, ETS, Regressão, Argentina, Energia Elétrica

1. Introdução

A previsão de carga desempenha um papel fundamental na eficiência e na estabilidade do sistema elétrico de um país. Na Argentina, um país com uma economia diversificada e uma ampla demanda por energia, essa previsão torna-se ainda mais crítica. A economia argentina é altamente dependente do fornecimento confiável de energia elétrica, que sustenta uma variedade de setores, desde a indústria até o comércio e a agricultura (Fofana et al., 2019).

A Argentina, devido à sua vasta extensão geográfica e diversidade climática, experimenta uma ampla variação sazonal nas temperaturas. As mudanças nas condições climáticas ao longo do ano desempenham um papel significativo na demanda de energia elétrica, já que as necessidades de aquecimento e refrigeração variam drasticamente. Durante os meses de inverno, as temperaturas podem cair acentuadamente em muitas regiões, levando a um aumento na demanda por eletricidade para aquecimento. Por outro lado, no verão, o uso intensificado de sistemas de resfriamento devido ao calor resulta em uma demanda adicional de energia. Portanto, a temperatura é um fator crítico que influencia diretamente a carga elétrica do país, e compreender essa dinâmica sazonal é essencial para uma previsão precisa e eficaz da carga (ISAAC et al., 2009; DAVIS et al., 2015).

Além disso, a Argentina tem importado uma parte significativa de seu gás durante os meses de inverno, principalmente por meio de contratos de GNL (Gás Natural Liquefeito). A previsão de carga elétrica desempenha um papel crucial na determinação da quantidade de energia necessária para a região, incluindo o dimensionamento adequado das importações de GNL. A falta de previsão precisa pode resultar em ineficiências operacionais e custos elevados (Vasquez et al., 2020).

O contexto do sistema elétrico argentino é complexo, com desafios específicos relacionados à sua estrutura e à dinâmica de mercado. A Argentina possui uma matriz energética diversificada, com destaque para a geração hidrelétrica, térmica e eólica. A geração de energia elétrica ocorre em diferentes regiões do país, cada uma contribuindo de forma distinta para o suprimento nacional. Isso inclui o sistema de transmissão interligado, que desempenha um papel vital na distribuição de eletricidade por todo o território argentino. É fundamental compreender essas complexidades para desenvolver uma previsão de carga precisa e adaptada às particularidades de cada região (Zeballos et al., 2018).

No âmbito do sistema elétrico argentino, a regionalização da geração e da carga elétrica é um fator determinante. Diferentes regiões do país contribuem com a geração de energia de maneiras distintas, dependendo de seus recursos naturais e infraestrutura. Essa heterogeneidade na geração precisa ser levada em consideração ao prever a carga, uma vez que a demanda de energia varia de região para região. É crucial otimizar a alocação de recursos e garantir a confiabilidade do fornecimento de energia em todas as partes do país, mantendo um equilíbrio entre a oferta e a demanda (González et al., 2021).

O sistema de transmissão interligado é uma espinha dorsal do sistema elétrico argentino, permitindo a distribuição eficiente de eletricidade em todo o país. Ele conecta as diversas regiões de geração e consumo, desempenhando um papel fundamental na manutenção da estabilidade e na prevenção de falhas no fornecimento de energia. A previsão de carga precisa levar em consideração a capacidade e a confiabilidade desse sistema, garantindo que a eletricidade seja entregue de forma eficaz e contínua a todas as áreas do país (Blanco et al., 2020).

A exportação de energia elétrica do Brasil e a Argentina é um elemento-chave na dinâmica energética entre esses dois países vizinhos. A exportação por fonte termoelétrica brasileira foi regulamentada pela portaria Portaria Normativa MME nº 418/2019 e aprimorado Portaria Normativa MME nº 49/2022, que estabelece diretrizes para a exportação de energia elétrica, em regime comercial, destinada à República Argentina ou à República Oriental do Uruguai, proveniente de excedente de geração de energia elétrica de usinas hidrelétricas disponíveis para atendimento ao Sistema Interligado Nacional (SIN) do Brasil. (Ministério de Minas e Energia, 2022).

Nos últimos anos, essa troca de energia tem assumido um papel cada vez mais significativo, com implicações não apenas para a estabilidade do fornecimento de energia, mas também para as operações comerciais entre as nações. A partir do ano de 2022, observou-se um aumento substancial na exportação de energia elétrica do Brasil para a Argentina. Esse aumento significativo pode ser atribuído, em parte, à situação de armazenamento elevado de energia no Brasil, que permitiu ao país expandir sua capacidade de exportação e contribuir para o atendimento das demandas energéticas na região, tanto por fonte térmica, como hídrica. (Ministério de Minas e Energia, 2022).

Este trabalho tem como objetivo abordar essas questões relacionadas a carga da Argentina por meio da aplicação de modelos avançados de previsão. Foram utilizados modelos de séries temporais, como ARIMA (Modelo Autorregressivo Integrado de Médias Móveis) e ETS (Erro, Tendência, Sazonalidade), para capturar padrões históricos e tendências de carga elétrica na Argentina.

Além disso, foram exploradas as relações entre a carga elétrica e a temperatura, aplicando técnicas de análise de regressão simples e quadrática para entender como as variações sazonais afetam a demanda de energia. A combinação dessas técnicas proporcionou uma visão abrangente e precisa da previsão de carga elétrica na Argentina.

Em resumo, este trabalho visa contribuir para a compreensão e melhoria da previsão de carga elétrica na Argentina para a tomada de decisões estratégicas.

1. Materiais e Métodos

Para a realização deste estudo, a coleta de dados foi conduzida de maneira rigorosa e abrangente. Duas fontes principais foram utilizadas para adquirir os dados necessários: a API (Interface de Programação de Aplicativos) fornecida pela CAMMESA (Compañía Administradora del Mercado Mayorista Eléctrico S.A.) e um relatório síntese em formato Excel fornecido pela CAMMESA. A API da CAMMESA permitiu a obtenção de dados diários, incluindo informações cruciais, como a carga e a geração de energia elétrica por fonte, ambas segmentadas por regiões elétricas, além da temperatura de Buenos Aires e o custo marginal de operação (R$/MWh). Após a aquisição, os dados brutos passaram por um processo de tratamento e estruturação usando a linguagem de programação Python. Posteriormente, esses dados foram armazenados com segurança em um banco de dados em nuvem baseado em PostgreSQL, garantindo a disponibilidade para análises futuras.

Os dados coletados foram organizados em dataframes, um formato de estrutura de dados tabular, para facilitar a manipulação e a análise. Os dataframes incluíam informações cruciais, como carga e geração de energia elétrica, ambas categorizadas por região elétrica, bem como a temperatura de Buenos Aires e o custo marginal de operação (R$/MWh). Essa organização permitiu a realização de análises detalhadas e a criação de séries temporais para cada variável de interesse. Além disso, para garantir a segurança dos dados, foram criados backups em formato CSV, além dos dados serem armazenados no banco de dados PostgreSQL.

Nossa abordagem envolveu os seguintes passos:

Diagrama

Descrição gerada automaticamente

Figura x. Fluxograma dos processos de base diária.

* Acesso à API da CAMMESA: Inicialmente, acessamos a API da CAMMESA por meio do produto 'PROGRAMACION DIARIA', baixando os arquivos ZIP correspondentes para máquina local.
* Extração dos Dados: Em seguida, realizou-se a extração dos dados contidos nos respectivos arquivos ZIPs.
* Leitura dos Arquivos de Texto: Os dados foram extraídos de três arquivos de texto fornecidos: o balanço energético, o preço horário da energia e o custo variável por máquina despachada. Esses arquivos estavam formatados como "arquivos de texto de largura fixa". Realizou-se os tratamentos necessários para transformá-los em dataframes legíveis.
* Criação de Função Automática: Uma função automatizada foi desenvolvida para salvar os dados do dataframe, de modo que pudesse ser aplicada de maneira uniforme aos três conjuntos de dados, tanto no banco de dados quanto em arquivos CSV.
* Geração de CSVs: Como resultado desse processo, foram criados três arquivos CSV:
* (CSV #1) balanco.csv (com 29 colunas) contendo informações sobre a geração de energia horária por tipo, área, data e o timestamp da extração.
* (CSV #2) precios.csv (com 27 colunas) contendo informações sobre as máquinas mais caras, os preços associados a elas (custo marginal de operação) em intervalos horários e o timestamp da extração.
* (CSV #3) cvp.csv (com 5 colunas) contendo informações sobre o preço de cada máquina gerada no dia, sua ordem de prioridade definida pela CAMMESA e o timestamp da extração.
* Captura dos dados de carga e temperatura mensal diretamente da sessão ‘Síntese Mensual’ do site da CAMMESA e convertidos em dois CSVs:
* (CSV #4) carga\_mensal.csv.
* (CSV #5) temperatura\_mensal.csv

É importante observar que os arquivos ‘precios.csv’ e ‘cvp.csv’ não foram relevantes para a análise desse trabalho, e sua utilização será adiada para outra etapa.

A análise e previsão dos dados foram conduzidas utilizando a linguagem de programação Python. Foram aplicados modelos de séries temporais para prever tendências e padrões nos dados de carga elétrica considerando que os dados passados têm a capacidade de prever o futuro. Além disso, realizou-se análises de regressões para entender as relações entre as variáveis, com foco especial na correlação entre carga elétrica e temperatura de Buenos Aires. A comparação desses métodos permitiu avaliar a eficácia de cada um na previsão e modelagem da carga elétrica na Argentina. Uma explicação mais detalhada desses métodos será apresentada na sessão de “Resultados e Discussões”, especificamente a partir do item ‘3.2’.

Como parte do rigoroso processo de análise e monitoramento das variáveis de interesse, foi elaborado um abrangente relatório gerencial (em anexo), fundamentado nos dados coletados, com o propósito de oferecer uma visão minuciosa das flutuações da carga elétrica ao longo do tempo. Adicionalmente, desenvolveu-se uma análise e previsão da carga de energia em um contexto mensal, destinada a resumir as tendências de carga em uma perspectiva de período mais amplo. Estes relatórios desempenham um papel de grande importância no contexto deste estudo, proporcionando informações críticas para a gestão eficaz da infraestrutura elétrica na Argentina e, de igual forma, apoiando as decisões estratégicas no setor de energia. O cerne deste trabalho está direcionado à análise mensal das variáveis em questão.

1. Resultados e Discussões
   1. **Estatística Descritiva**

A partir dos conjuntos de dados coletados, foram adquiridos dados de temperatura (temperatura\_mensal.csv) e carga elétrica (carga\_mensal.csv) com o propósito de gerar representações gráficas que permitissem a visualização dos dados. Além disso, foram aplicados testes de normalidade e análises estatísticas descritivas a fim de aprofundar a compreensão dos dados e extrair informações relevantes. A utilização dessas ferramentas estatísticas se fundamenta na necessidade de analisar o conteúdo estatístico dos gráficos e dados, proporcionando uma avaliação mais robusta e precisa das tendências e padrões presentes (Bussab, W. O., & Morettin, P. A., Estatística Básica, Editora Saraiva).

Gráfico, Histograma

Descrição gerada automaticamente

Figura x. Dados mensais de carga. Série temporal mensal da carga (MWmed) [1]. Carga mensal para cada ano (MWmed) [2]. Histograma com uma curva normal ajustada aos dados [3].

Fonte: Resultados originais da pesquisa.

Gráfico, Histograma

Descrição gerada automaticamente

Figura x. Dados mensais de temperatura. Série temporal mensal de temperatura (ºC) [1]. Temperatura mensal para cada ano (ºC) [2]. Histograma com uma curva normal ajustada aos dados [3].

Fonte: Resultados originais da pesquisa.

Um aspecto essencial para a condução deste estudo é a minuciosa decomposição das séries temporais, um procedimento que envolve a representação gráfica dos dados de carga e temperatura, segmentados mês a mês, de modo a identificar a tendência das séries, as variações sazonais previsíveis e os resíduos, que são os padrões imprevisíveis remanescentes. Essa abordagem oferece uma compreensão aprofundada das flutuações temporais nos dados e é fundamental para a análise precisa e a modelagem das séries temporais.

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

Figura x. Decomposição dos dados de carga. Série temporal mensal (MWmed) [1]. Tendência [2]. Sazonalidade [3]. Resíduos [4].

Fonte: Resultados originais da pesquisa.

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

Figura x. Decomposição dos dados de temperatura. Série temporal mensal (ºC) [1]. Tendência [2]. Sazonalidade [3]. Resíduos [4].

Fonte: Resultados originais da pesquisa.

Tabela 1. Estatística Descritiva para os dados de carga (MWmed) e temperatura (ºC)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | estatística descritiva carga | valor | | Contagem | 90.0 | | Média | 15917.8 | | Desvio Padrão | 1391.3 | | Mínimo | 12612.0 | | 1º Quartil | 14859.8 | | Mediana | 15771.5 | | 3º Quartil | 17042.5 | | Máximo | 19651.0 | | |  |  | | --- | --- | | estatística descritiva temperatura | valor | | Contagem | 90.0 | | Média | 18.8 | | Desvio Padrão | 5.1 | | Mínimo | 10.6 | | 1º Quartil | 14.2 | | Mediana | 18.2 | | 3º Quartil | 23.4 | | Máximo | 30.3 | |

Fonte: Resultados originais da pesquisa.

Ao analisar os gráficos e as figuras apresentadas anteriormente, é possível identificar algumas tendências e padrões importantes relacionados ao consumo de energia na Argentina. Primeiramente, nota-se uma notável tendência de crescimento na carga elétrica ao longo do período analisado. Além disso, as séries temporais exibem sazonalidades bem definidas, sugerindo a presença de padrões recorrentes que se repetem em intervalos regulares.

É importante destacar que ambas as séries de dados, referentes à energia (MWmed) e temperatura, foram submetidas ao teste de normalidade de Kolmogorov-Smirnov, com um nível de significância de 5%. Esse teste tem o propósito de verificar se os valores em ambas as séries seguem uma distribuição normal, o que é um pressuposto importante para muitas análises estatísticas. Os resultados desses testes indicaram que ambos os conjuntos de dados não apresentaram desvios significativos da normalidade, com p-valores de 0.4927 para a série de energia (MWmed) e 0.3158 para a série de temperatura. Esses p-valores, atendem ao nível de significância de 5%, sugerem que não há evidências estatísticas para rejeitar a hipótese nula de normalidade. Isso é fundamental para garantir a validade das análises subsequentes realizadas neste estudo ("Métodos Estatísticos" de George W. Snedecor e William G. Cochran).

Adicionalmente, as estatísticas descritivas fornecem uma visão abrangente das características centrais das séries temporais. Os valores médios, medianos, desvios padrão e outros parâmetros estatísticos estão detalhados na tabela 1. Essas estatísticas oferecem informações essenciais sobre a tendência central, a dispersão e a variabilidade dos dados.

Essas observações e conclusões têm implicações significativas para o entendimento do consumo de energia na Argentina e podem orientar futuras investigações e estratégias de gerenciamento de energia no país.

* 1. **Séries Temporais**

No contexto do estudo de séries temporais, é essencial compreender as principais componentes que contribuem para a evolução dos dados ao longo do tempo. Uma série temporal pode ser desagregada em três componentes principais: tendência, sazonalidade e componente residual. A tendência representa a direção geral dos dados ao longo do tempo, revelando se eles estão aumentando, diminuindo ou permanecendo estáveis. A sazonalidade refere-se a padrões repetitivos que ocorrem em intervalos fixos de tempo, como sazonalidade anual, mensal, semanal etc. Esses padrões podem ser influenciados por eventos sazonais, como feriados, estações do ano ou ciclos de negócios. Por fim, a componente residual engloba variações aleatórias e não explicadas pelas tendências ou sazonalidades, representando o "ruído" nos dados.

A identificação e modelagem adequada dessas componentes são cruciais para a análise de séries temporais. Isso é feito por meio de técnicas estatísticas, como decomposição, suavização e modelos autorregressivos. A decomposição de séries temporais permite separar as tendências, sazonalidades e componentes residuais, tornando mais fácil analisar cada aspecto separadamente. A suavização exponencial, por outro lado, é uma técnica que visa capturar as tendências subjacentes, enquanto os modelos autoregressivos, como o ARIMA, abordam a dependência temporal nos dados (Shumway & Stoffer, 2017).

Além disso, é importante destacar a necessidade de testar a estacionariedade das séries temporais, pois muitas técnicas estatísticas assumem que os dados são estacionários. A estacionariedade implica que as propriedades estatísticas da série não mudam com o tempo, tornando mais viável a aplicação de modelos. A análise de autocorrelação e normalidade dos resíduos são frequentemente usados para entender a dependência serial dos dados (Shumway & Stoffer, 2017).

Para os próximos dois métodos, o conjunto de dado de carga da argentina foi separado em treino e teste. Dado que o intuito é fazer a previsão para o período de um ciclo sazonal de um ano, os dados de treino foram estabelecidos de janeiro/2016 até junho/2022 e os dados previstos de julho/22 até junho/23.

* + 1. **Séries Temporais - Suavização Exponencial**

A técnica de Suavização Exponencial é uma abordagem amplamente utilizada para previsão em séries temporais. Ela se baseia em métodos que consideram a média ponderada dos valores passados para estimar os valores futuros. Nesse contexto, a Suavização Exponencial é especialmente útil quando lidamos com séries temporais que exibem tendências e sazonalidades.

O objetivo de utilização da Suavização Exponencial nesse contexto é criar previsões estatísticas, em bases mensais, para determinar a carga de energia em séries temporais. Para isso, foram aplicadas diferentes variantes dessa técnica para modelar os dados e fazer previsões, considerando os seguintes métodos: Suavização Exponencial Simples (SES) é a forma mais básica, adequada para séries sem tendência ou sazonalidade. O Suavização Exponencial de Holt (SEH) estende o SES com uma tendência linear, útil quando há uma inclinação na série. O Suavização Exponencial de Holt-Winters (HW) adiciona sazonalidade; o modelo pode ser aditivo, quando a sazonalidade é constante, ou multiplicativo, quando varia com a série. Além disso, aplicamos o método ETS (Erro, Tendência, Sazonalidade), uma generalização flexível que combina diferentes componentes para lidar com uma ampla gama de padrões em séries temporais, permitindo adaptação precisa aos dados de carga de energia (Forecasting: Principles and Practice de Rob J. Hyndman e George Athanasopoulos).

Na avaliação de modelos de previsão, três métricas desempenham um papel fundamental. O Erro Médio Quadrático (RMSE - Root Mean Squared Error) é uma medida que calcula a raiz do erro médio quadrático entre as previsões e os valores reais. Ele avalia a dispersão dos erros e penaliza desvios significativos, tornando-o sensível a valores discrepantes. Quanto menor o valor de RMSE, melhor o ajuste do modelo aos dados observados. Em contraste, o Erro Médio Absoluto (MAE - Mean Absolute Error) é a média dos valores absolutos dos erros, proporcionando uma medida direta da magnitude média dos erros. O objetivo, novamente, é minimizar o MAE, e um valor menor indica um ajuste superior do modelo. Finalmente, o Erro Médio Percentual Absoluto (MAPE - Mean Absolute Percentage Error) calcula a média das porcentagens dos erros absolutos em relação aos valores reais, expressando o erro em termos percentuais. Isso facilita a compreensão do erro relativo, e um MAPE menor reflete uma previsão mais precisa em relação aos valores observados

Após a análise dos resultados, configurou-se o modelo ETS de acordo com a melhor parametrização indicada pelos outros métodos, conforme os seguintes passos:

* Tendência: 'add' indica que foi inserida uma componente de tendência aditiva no modelo, o que significa que estamos considerando uma tendência linear ao longo do tempo.
* Sazonalidade: ‘mul’ indica que estamos incluindo uma componente sazonal multiplicativa no modelo, o que implica que há variações sazonais que são somadas à série.
* Períodos Sazonais: ‘12’ define o período sazonal como 12 meses, sugerindo que estamos lidando com dados mensais e que a sazonalidade ocorre nesses intervalos.
* Em seguida, ajustamos o modelo aos dados de treinamento usando ets\_model.fit(). Isso calcula os parâmetros do modelo ETS com base nos dados de treinamento.
* Finalmente, fazemos previsões para os dados de teste usando ets\_fit.predict().
* Todas as previsões foram plotadas conforme a figura a seguir.

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

Figura x. Resultados dos métodos de suavização exponencial.

Fonte: Resultados originais da pesquisa.

Tabela 2. Métricas para análise de desempenho dos modelos de suavização exponencial

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Métrica | SES | SEH | HW - Aditivo | HW - Multiplicativo | ETS |
| RMSE | 2737.4 | 2913.4 | 1464.3 | 1271.8 | 1271.8 |
| MAE | 2221.7 | 2379.2 | 1281.0 | 1081.3 | 1081.3 |
| MAPE (%) | 14.3 | 15.3 | 7.9 | 6.6 | 6.6 |

Fonte: Resultados originais da pesquisa.

Para verificar a qualidade de uma previsão em um modelo de séries temporais, realizou-se uma série de testes de diagnóstico nos resíduos do modelo: Ljung-Box, Teste de Shapiro-Wilk e Teste ARCH, são comumente usados para avaliar diferentes aspectos da qualidade do modelo.

O teste de Ljung-Box verifica se há autocorrelação serial significativa nos resíduos. Se os resíduos forem independentes, não deve se encontrar autocorrelação significativa. A hipótese nula (H0) é que não há autocorrelação nos resíduos. Portanto, você deseja que o p-valor associado ao teste seja maior que um nível de significância (geralmente 0,05 ou 0,01) para aceitar a hipótese nula. Utilizou-se a função acorr\_ljungbox do statsmodels para realizar este teste.

O Teste de Shapiro-Wilk verifica se os dados seguem uma distribuição normal. A hipótese nula (H0) é que os resíduos seguem uma distribuição normal. É desejado que o p-valor associado ao teste seja maior que um nível de significância (geralmente 0,05 ou 0,01) para aceitar a hipótese nula. Utilizou-se a função shapiro do scipy.stats para realizar este teste.

O Teste ARCH é empregado para detectar a presença de heteroscedasticidade condicional nos resíduos, o que sugere que a variância dos resíduos pode variar com o tempo ou com os valores anteriores. Nesse procedimento, é possível ajustar um modelo ARCH (AutoRegressive Conditional Heteroskedasticity) aos resíduos e avaliar a significância dos coeficientes ARCH resultantes. A hipótese nula (H0) é que não há efeitos de heteroscedasticidade condicional nos resíduos. Para conduzir essa análise, utilizamos a biblioteca ARCH.

Em resumo, a análise dos resíduos revelou resultados significativos. O teste de Ljung-Box não encontrou evidências de autocorrelação serial nos resíduos (LAG 6 = 0.4899), indicando independência. No entanto, o teste de Shapiro-Wilk apontou desvios da normalidade, com um p-valor de 0.0205 para um intervalo de confiança de 5%, embora tenha passado no teste com um intervalo de confiança de 10%. Não há evidência de efeitos ARCH nos resíduos. Essas descobertas têm implicações importantes para a validade das suposições em análises subsequentes.

* + 1. **Séries Temporais - SARIMA**

Compreendendo a importância da previsão em séries temporais, uma técnica crucial que exploramos é o ARIMA (AutoRegressive Integrated Moving Average). Ao contrário da Suavização Exponencial, o ARIMA se baseia em uma abordagem diferente para modelar e prever séries temporais. Esta metodologia é particularmente eficaz quando se trata de séries que não exibem uma estrutura clara de tendência ou sazonalidade.

O ARIMA envolve três componentes principais: o modelo autorregressivo (AR), o modelo de média móvel (MA) e a diferenciação integrada (I). A componente autorregressiva (AR) considera as relações lineares entre as observações passadas e a observação atual, permitindo capturar padrões de autocorrelação. O modelo de média móvel (MA) se concentra nos erros passados e os utiliza para modelar a série. A diferenciação integrada (I) envolve a transformação da série em uma série estacionária por meio da subtração das observações anteriores das observações atuais, o que ajuda a remover tendências (Shumway & Stoffer, 2017).

Ademais, utilizou-se o SARIMA (Seasonal ARIMA) para lidar com séries temporais que exibem sazonalidade, adicionando componentes sazonais ao ARIMA tradicional. Isso é particularmente relevante nesse contexto, uma vez que lidamos com dados mensais de carga de energia, onde sazonalidades podem ser proeminentes. O SARIMA inclui componentes sazonais AR, MA e I, permitindo a captura de padrões sazonais complexos (Shumway & Stoffer, 2017).

Portanto, ao empregar o SARIMA nesse estudo, é possível modelar e prever a carga de energia de forma eficaz, considerando as características específicas das séries temporais em questão. O trabalho seguiu os seguintes passos:

* Especificação dos Componentes: Define-se os componentes ARIMA (AutoRegressive Integrated Moving Average) e SARIMA (Seasonal ARIMA) para modelar a série temporal
* Para a estimativa dos parâmetros do modelo SARIMA, a função 'auto\_arima' foi utilizada, permitindo a determinação automática dos melhores valores de p, d, q, P, D e Q com base nos dados de treinamento. Os resultados obtidos foram comparados com um modelo de desenvolvimento interno que avalia diversas combinações de parâmetros e os registra em um banco de dados, levando em consideração métricas como RMSE, MAE e MAPE, além dos resultados dos testes de Ljung-Box, Shapiro-Wilk e ARCH. Importante mencionar que os resultados detalhados do modelo interno estão disponíveis exclusivamente na base de dados dos códigos, sendo que o foco deste trabalho está na função 'auto\_arima', o modelo interno tem apenas um viés de aprovar o ‘auto-arima’.
* Previsões e Avaliação: Após a estimação dos parâmetros, realiza-se previsões para os dados de teste. As previsões são comparadas com os valores reais para calcular métricas de avaliação, como o Erro Quadrático Médio (MSE), o Erro Absoluto Médio (MAE) e o Erro Médio Percentual Absoluto (MAPE). Essas métricas serão fundamentais para comparar os métodos de series temporais.
* Análise dos Resíduos: Além das métricas de avaliação, foram avaliados os resíduos gerados pelo modelo SARIMA. Os resíduos são as diferenças entre os valores reais e os valores previstos. Foram aplicados testes estatísticos, como autocorrelação dos resíduos, normalidade e teste ARCH, para avaliar se os resíduos são adequados para previsões.

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

Figura x. Resultados dos métodos SARIMA. Comparação entre os resultados do teste SARIMA e os dados reais [1]. Gráfico dos resíduos [2]. Histograma dos resíduos [3].

Fonte: Resultados originais da pesquisa.

O modelo SARIMA (2, 0, 4)x(1, 0, [], 12) representa uma especificação do modelo SARIMA que foi identificada pela função ‘auto\_arima’. Segue uma explicação mais eficiente:

* ARIMA(2, 0, 4): Isso se refere à parte não sazonal do modelo. O primeiro número (2) representa a ordem do termo autorregressivo (AR). O segundo número (0) refere-se à ordem da diferenciação não sazonal (I). O terceiro número (4) representa a ordem do termo de média móvel (MA).
* x: Isso indica a parte sazonal do modelo.
* (1, 0, [], 12): Isso se refere à parte sazonal do modelo. O primeiro número (1) representa a ordem do termo autorregressivo sazonal (SAR), que é semelhante ao AR, mas aplicado à sazonalidade. O segundo número (0) novamente indica que nenhuma diferenciação sazonal foi aplicada. O terceiro valor ([]) significa que não foram incluídos termos de média móvel sazonal (SMA). O último número (12) representa a periodicidade sazonal, que neste caso é 12 meses, indicando sazonalidade anual.

Tabela 3. Métricas para análise de desempenho do modelo SARIMA

|  |  |
| --- | --- |
| Métrica | SARIMA |
| RMSE | 3636873.0 |
| MAE | 1213.7 |
| MAPE (%) | 6.6 |

Fonte: Resultados originais da pesquisa.

Em resumo, a análise dos resíduos revelou resultados significativos. O teste de Ljung-Box não encontrou evidências de autocorrelação serial nos resíduos (LAG 1 = 0.1197), indicando independência. No entanto, o teste de Shapiro-Wilk apontou desvios da normalidade, com um p-valor de 0.0315 para um intervalo de confiança de 5%, embora tenha passado no teste com um intervalo de confiança de 10%. Não há evidência de efeitos ARCH nos resíduos. Essas descobertas têm implicações importantes para a validade das suposições em análises subsequentes.

* 1. **Regressões**

Na análise de regressões de dados, o principal objetivo é compreender o relacionamento entre variáveis. Diferentes técnicas estatísticas podem ser aplicadas para modelar esse relacionamento e identificar possíveis padrões nos dados. Neste trabalho, a abordagem utilizada será a comparação entre a temperatura de Buenos Aires (a mesma temperatura de acompanhamento utilizada pela CAMMESA) e carga de energia na Argentina (Brockwell & Davis, 2016).

No contexto deste estudo, aplicou-se um tratamento de anomalias nos dados com o objetivo de aprimorar a qualidade das regressões. As margens de corte foram definidas com base no teste Z-Score, seguindo as recomendações da literatura que consideram resíduos maiores do que 2, 2,5 ou 3 desvios padrão como anômalos. Para a identificação de anomalias nos dados, foi estabelecido um limiar de 2,5, permitindo a detecção de pontos de dados que se afastaram significativamente da variação esperada. Como exemplo, observou-se uma anomalia notável na variável de energia, na qual um ponto de dados apresentou um Z-Score excepcionalmente alto de 2,68, indicando uma ocorrência fora do padrão. Como parte do processo de pré-processamento, essa entrada, que excedeu o limiar estabelecido, foi removida do conjunto de dados. Essa abordagem de detecção de anomalias pôde ser eficaz devido à tendência de normalidade observada nos dados, permitindo a identificação eficiente de valores discrepantes (Krawiec et al., 2021).

A primeira técnica utilizada é a regressão linear por Mínimos Quadrados Ordinários (OLS). Essa técnica visa compreender como uma variável, como a temperatura, se relaciona com a carga de energia. O modelo OLS linear assume uma relação linear entre as variáveis e procura encontrar uma linha reta que melhor se ajuste aos dados, representando a tendência geral.

Foi ajustado um modelo OLS linear entre a temperatura e a carga de energia, fornecendo informações sobre a direção dessa relação. O resumo do modelo apresenta se a relação é positiva (aumento da temperatura resulta em aumento na carga) ou negativa (aumento da temperatura resulta em diminuição na carga). Além disso, é possível avaliar a qualidade do ajuste e sua significância estatística.

Entretanto, em alguns casos, os dados não seguem uma relação linear perfeita. Pode ocorrer que as relações sejam não-lineares ou que os dados não atendam aos pressupostos de normalidade. Nessas situações, a transformação de Box-Cox pode ser aplicada para melhorar a adequação do modelo.

A transformação de Box-Cox é uma técnica que ajusta os dados por meio de um parâmetro lambda (λ). Essa transformação pode aproximar os dados de uma distribuição normal, tornando-a útil para diversas técnicas estatísticas. Ao aplicar a transformação de Box-Cox à série temporal, busca-se melhorar a linearidade da relação entre temperatura e carga de energia.

Porém, em alguns cenários, a relação entre variáveis pode ser mais bem modelada como uma curva quadrática em vez de uma linha reta. Para abordar essa situação, a técnica de OLS quadrático é aplicada. Isso permite que a relação entre temperatura e carga de energia seja capturada de forma mais flexível, considerando não apenas a direção da mudança, mas também a curvatura.

Foi ajustado um modelo OLS quadrático entre temperatura e carga de energia, possibilitando a identificação de relações mais complexas do que uma simples tendência linear. O resumo do modelo inclui informações sobre coeficientes quadráticos e lineares, proporcionando uma compreensão mais completa da relação.

Os modelos de regressão foram parametrizados e estabelecidos, conforme os seguintes passos:

* Criação de um dataframe com com CSVs #4 e #5 excluindo a coluna de data, por se tratar de uma análise atemporal.
* Utilizar a função ‘OLS’ da biblioteca statsmodel para realizar as modelagens de correlação linear e quadrática entre carga e temperatura.
* Utilizar a função ‘box-cox’ da biblioteca ‘scipy.stats’ para realizar a transformação na variável dependente carga de energia (Y), com a obtenção dos parâmetros lambda e a carga de energia transformada (Y\*).
* Utilizar a função ‘OLS’ da biblioteca statsmodel para realizar as modelagens de correlação linear com a carga de energia transformada (Y\*).
* Para as 3 modelagens, a eficiência do modelo foi analisada pelo coeficiente de determinação (R²) e a sua validação estatística foi inferida pelo teste Shapiro-Francia.

Gráfico, Gráfico de dispersão

Descrição gerada automaticamente

Figura x Correlações simples e quadráticas: Energia (MWmed) x Temperatura.

Fonte: Resultados originais da pesquisa.

Tabela 4. Métricas para análise de desempenho do modelo de correlação quadrática

|  |  |  |  |
| --- | --- | --- | --- |
| Métrica | Correlação Quadrática | Transformada Box-Cox | Correlação Quadrática |
| RMSE | - | - | 1739035.2 |
| MAE | - | - | 1114.5 |
| MAPE (%) | - | - | 3.2 |
| R² (%) | 2.1 | 2 | 77.8 |

Fonte: Resultados originais da pesquisa.

Os resultados dos testes de normalidade dos resíduos, com p-valor correspondentes a 0.3332, 0.3288 e 0.2950 para os modelos OLS linear, OLS linear com transformação Box-Cox e OLS quadrático, respectivamente, indicam que, em todos os casos, não há evidências estatísticas para rejeitar a hipótese nula de que os resíduos seguem uma distribuição normal. Essa conformidade com a normalidade dos resíduos é um aspecto relevante para a validade dos modelos e análises subsequentes.

Os valores de R², uma métrica essencial na avaliação de modelos estatísticos, foram obtidos neste estudo para entender a relação entre temperatura e carga de energia. Observou-se que os modelos de regressão linear e transformação de Box-Cox tiveram baixos R², indicando que apenas cerca de 2,1% e 2% da variância na carga de energia pode ser explicada pela temperatura, respectivamente. No entanto, o modelo de regressão quadrática obteve um R² significativamente mais alto, aproximadamente 77,8%, sugerindo que a relação entre essas variáveis é melhor capturada por um modelo não linear. Esses resultados fornecem insights valiosos, mas não concluem definitivamente sobre a relação subjacente.

1. Conclusões

Neste estudo, foi conduzida uma análise abrangente das séries temporais de carga de energia na Argentina, empregando técnicas de Suavização Exponencial, Modelos ARIMA e análises de regressão. O objetivo central foi compreender as tendências, sazonalidades e relacionamentos entre variáveis para fornecer insights críticos para prever a variável carga.

Na primeira parte, foram aplicados os métodos de Suavização Exponencial, incluindo Suavização Exponencial Simples (SES), Suavização Exponencial de Holt (SEH), Suavização Exponencial de Holt-Winters (HW) e o método ETS (Erro, Tendência, Sazonalidade). Os resultados destacaram o ETS como o modelo mais preciso, com um MAPE (Erro Médio Percentual Absoluto) de apenas 6,6%. Esse modelo demonstrou uma notável capacidade de se adaptar às complexas tendências e sazonalidades nos dados de carga de energia.

Na segunda parte, os Modelos ARIMA foram explorados, uma abordagem essencial e avançada para a análise de séries temporais. Foi utilizado o modelo SARIMA (ARIMA Sazonal) para lidar com dados que exibiam padrões sazonais claros. Após a estimativa de parâmetros e previsão, foi obtido um modelo SARIMA (2, 0, 4)x(1, 0, [], 12) que se destacou com um MAPE de 6,6%. Isso demonstra a eficácia do SARIMA na modelagem e previsão de séries temporais com sazonalidade.

Na terceira parte, foram investigadas as relações entre a temperatura e a carga de energia por meio de análises de regressão linear e quadrática. O modelo de regressão quadrática capturou melhor a complexidade dessa relação, com um R² (coeficiente de determinação) de aproximadamente 77,8%. Isso sugere que a relação entre essas variáveis não é puramente linear e que a temperatura de Buenos Aires desempenha um papel significativo na demanda de energia.

Em resumo, este estudo forneceu uma visão abrangente das técnicas de análise de séries temporais, modelos ARIMA e análises de regressão, oferecendo uma compreensão aprofundada da dinâmica da carga de energia na Argentina. As métricas de desempenho, como MAPE, RMSE e R², foram cruciais para avaliar a precisão dos modelos e as relações entre variáveis.

Os modelos de séries temporais são de sumo interesse na análise puramente com base no histórico, para capturar suas características e realizar as previsões. Já os modelos de correlação tem uma característica interessante: A temperatura de Buenos Aires para os meses seguintes são fornecidas por instituições de respeito, como por centros Norte Americanos e Europeus, fornecendo uma ideia do futuro.

À medida que continuamos a coletar e analisar dados, as técnicas de séries temporais, modelos ARIMA e análises de regressão permanecerão ferramentas vitais para prever e entender as complexas dinâmicas das séries temporais em diversos setores.

1. **Reconhecimento**

Agradeço imensamente ao meu orientador pelo suporte crucial durante este trabalho. Também sou grato ao Msc. Raimundo Otávio por suas valiosas orientações, que enriqueceram significativamente a pesquisa, guiando-a com maestria pelo campo da ciência de dados.

1. Referências

Blanco, R., Santos, F. M., & Madureira, A. G. (2020). A review of electric power grid interconnection in South America. Energies, 13(17), 4374.

Brockwell, Peter J. and Davis, Richard A. (2002). Introduction to Time Series and Forecasting, 2nd. ed., Springer-Verlang.

Fofana, A., Wang, J., & Du, J. (2019). Short-Term Load Forecasting Based on a Hybrid Model Combining Singular Spectrum Analysis and ARIMA. Energies, 12(14), 2788.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. (2nd ed.).

González, P. J., Romero, R., & Prevé, S. (2021). A review of electrical load forecasting models. Energies, 14(12), 3472.

Krawiec, P., Junge, M. e Hesselbach, J. Comparison and Adaptation of Two Strategies for Anomaly Detection in Load Profiles Based on Methods from the Fields of Machine Learning and Statistics. Open Journal of Energy Efficiency, 10, 37-49.

Snedecor, G.W. and Cochran, W.G. (1989) Statistical Methods. 8th Edition, Iowa State University Press, Ames.

Ministério de Minas e Energia. Portaria normativa MME nº 418/2019, Brasília, DF.

Disponível em, 2019.

Ministério de Minas e Energia. Portaria normativa MME nº 49/2022, Brasília, DF.

Disponível em, 2022.

SHUMWAY, R. H.; STOFFER, D. S. Time Series Analysis and Its Applications: With R Examples. Springer, 2017.

Vasquez, C., & Rivadeneira, J. (2020). Short-term load forecasting using artificial intelligence techniques: A review. Energies, 13(1), 168.

Zeballos, D. C., Gerbex, S., & Cherkaoui, R. (2018). A review of probabilistic forecasting techniques for wind energy generation.