



Método Combinado na Previsão de Séries Temporais de Consumo de Energia Elétrica

 Maria Clara Palermo de Souza Carvalho Caria¹ e  Anna Regina Corbo Costa²

Versão do autor aceita publicada online: 30 abril 2024

Publicado online: 09 maio 2024

¹ Bacharel em Engenharia de Produção pelo Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ (2022). Rio de Janeiro, RJ. Brasil

² Doutora em Engenharia Civil (UFRJ, 2017). Professora Associada do Departamento de Matemática do CEFET/RJ. Área de atuação: Matemática aplicada e Métodos estatísticos. Rio de Janeiro, RJ. Brasil

Nota dos Autores

Os autores declaram que não há conflitos de interesses.

Como citar esse artigo - American Psychological Association (APA):

Caria, M. C. P. S. C., & Costa, A. R. C. (2024, artigo aceito online). Método Combinado na Previsão de Séries Temporais de Consumo de Energia Elétrica. *Exacta*, artigo aceito online.

<https://doi.org/10.5585/2024.25125>

RESUMO

A produção de energia elétrica deve ser planejada de modo a otimizar seus processos e minimizar possíveis falhas, o que pode ser auxiliado pela análise de séries históricas de consumo. Este artigo propõe a construção de um modelo combinado de previsão de séries temporais, a fim de prever o consumo de energia elétrica por consumidor para todos os estados brasileiros. Este consiste na combinação linear dos modelos TSLMS, TSLMTS e SNAIVE utilizando três e cinco anos de histórico. Os pesos atribuídos a cada modelo são funções dos erros calculados pelo desvio médio absoluto das previsões individuais. O modelo combinado apresentou erro quadrático médio de 5,7 kWh por consumidor e Theil's U de 0,76, ilustrando resultado mais acurado para três anos de histórico e uma performance equivalente aos outros modelos individuais para um histórico maior. Assim, a presente proposta é aplicável para séries com poucos dados históricos disponíveis, apresentando resultados promissores para um ano de previsão.

Palavras-chave: séries temporais, energia elétrica, modelos de previsão, modelo combinado

Combined Method in Forecasting Time Series of Electric Power Consumption

Abstract: The production of electrical energy must be planned in order to optimize its processes and minimize possible failures, which can be helped by analyzing historical consumption series.

This article proposes the construction of a combined time series forecasting model, in order to predict electricity consumption per consumer for all Brazilian states. This consists of the linear combination of the TSLMS, TSLMTS and SNAIVE models using three and five years of history. The

weights assigned to each model are functions of the errors calculated by the mean absolute deviation of the individual predictions. The combined model presented a mean squared error of 5.7 kWh per consumer and Theil's U of 0.76, illustrating a more accurate result for three years of history and a performance equivalent to the other individual models for a longer history. Thus, the present proposal is applicable to series with little historical data available, presenting promising results for a forecast year.

Keywords: time series, electrical energy, forecast, combined model

1 Introdução

Desde a sua descoberta, a energia elétrica tornou-se uma das principais fontes de energia utilizadas pela sociedade. A evolução contada pela história da civilização fundamentava-se, principalmente por proporcionar iluminação, manutenção de alimentos em freezers e geladeiras, banhos climatizados, uso de aparelhos eletrodomésticos e entre outros. Entretanto, em uma sociedade altamente conectada, a falta da energia elétrica acarretaria a inutilização de seus protagonistas: os eletrônicos, apresentando uma sociedade e um mundo extremamente menos interligado (Farias & Sellitto, 2013).

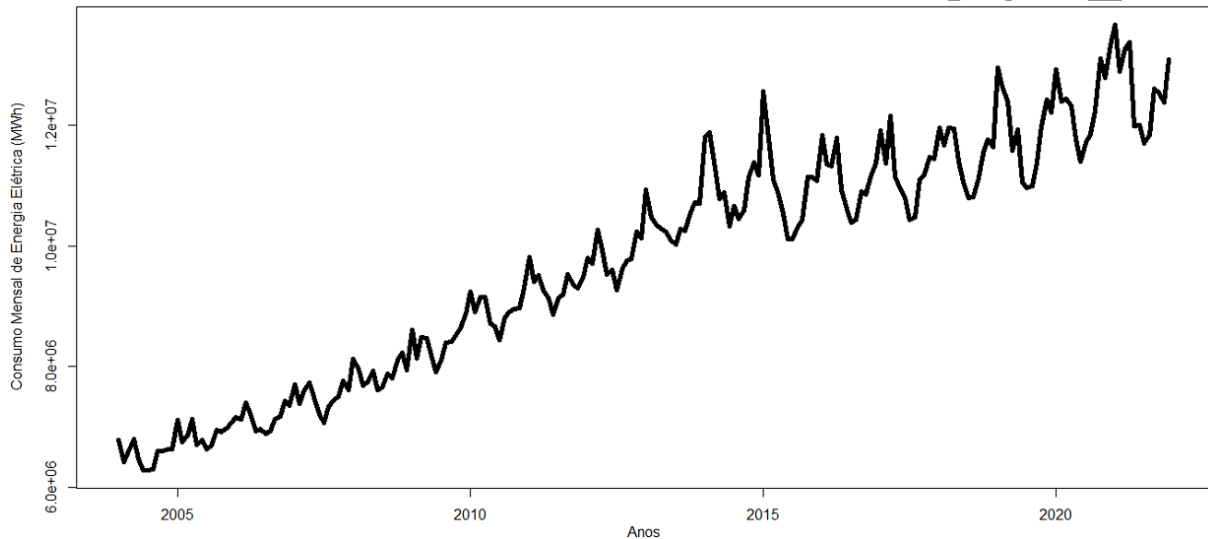
A energia elétrica pode ser gerada a partir de usinas hidrelétricas, termoeletricas, eólicas e solares, por exemplo; sendo esta primeira a principal fonte de produção no Brasil. Neste caso, deve-se comentar os impactos negativos associados à sua produção, como o alagamento e o aumento da poluição atmosférica. Dessa forma, entende-se a importância de estudar e prever o consumo de energia elétrica, de modo a evitar sua produção exacerbada e apontar o mínimo necessário para que essa supra a demanda (Reis, 2011).

O Brasil, como outros países em desenvolvimento, passa por um momento de aumento na demanda por energia, como é possível observar na Figura 1. Segundo Neto *et al* (2016), o consumo de energia per capita variou de 1,485 kWh, em 1990, para 2,225 kWh em 2009. Para o autor, para

que haja um desenvolvimento sustentável, é preciso que uma determinada região possua um planejamento de como o fornecimento desses recursos será feito no longo prazo para que não haja excessos ou faltas.

Figura 1

Consumo Total de Energia Elétrica (MWh) no Brasil entre 2004 e 2021.



Fonte: Elaboração própria do autor a partir de dados disponibilizados pela EPE (2022).

Dado a importância e a relevância de sua aplicação, a rede de distribuição de energia elétrica passou a ser considerada um mercado. No Brasil, tal responsabilidade dividiu-se entre o governo e o setor privado, onde este primeiro encarrega-se de gerar a energia e a regulamentar, sendo esta última ação sob a responsabilidade da ANEEL (Agência Nacional de Energia Elétrica); enquanto o setor privado incumbe-se da missão de gerir de forma eficiente tal recurso (Lorenzo, 2001).

Assim sendo, a linha de partida para construção deste mercado consiste nas usinas geradoras de energia. Segundo Carneiro (2014), uma concessionária de energia elétrica pode ser vista como um sistema produtivo, onde usinas, turbinas e geradores combinados com combustível ou água produzem energia, que será distribuída ao cliente.

Desse modo, entendido as particularidades deste mercado, recorda-se a principal premissa para que o funcionamento de um sistema siga como esperado: este deve ter sua produção planejada e controlada, de modo a otimizar seus processos e diminuir e/ou evitar possíveis falhas. Posto isto, fundamenta-se a relevância em, a partir de um histórico apresentado por uma série temporal, ser capaz de prever ou estimar seu comportamento futuro.

A partir desta proposta, avalia-se a abordagem de três principais cenários, baseados nos objetivos determinados para cada intervalo de tempo, sendo estes: a curto prazo, a médio prazo e a longo prazo. De acordo com a janela temporal escolhida, são avaliados, sobretudo: a capacidade produtiva da linha, onde pode cogitar-se mudanças de layout ou instalação de novas plantas; o suprimento da demanda, visando a máxima redução de desperdício; a necessidade de compra de matérias primas e o estudo de otimização das máquinas e equipamentos.

Já o planejamento da produção busca precaver-se de situações ou imprevistos que ocasionem desvios nos objetivos financeiros determinados em primeiro plano. Incorporado ao contexto de mercado de energia elétrica, estes baseiam-se no equilíbrio dos custos, principalmente aos que se referem a excessos ou faltas. No caso de uma produção excessiva, as multas aplicadas por agências controladoras ou a inviabilidade de estocar tal tipo de energia justificam um dos lados da balança; enquanto, a tentativa de inibir a baixa produção, de modo a atender sua demanda, compensa o outro.

A aplicação de modelos focada na geração de uma previsão é possível a partir da escolha prévia de uma série temporal. De acordo com Morettin e Tolo (2018), denomina-se “série temporal” a sequência de observações do comportamento de uma variável ao longo do tempo. Os principais objetivos de sua análise consistem em compreender o mecanismo gerador da série e prever o comportamento futuro da série.

A combinação de mais de um modelo temporal de previsão, em grande parte das vezes, resulta em previsões superiores em termos de acurácia quando comparados a técnicas individuais (Hibon & Evgeniou, 2005). Diversos trabalhos apontam resultados onde diferentes métodos de combinações obtiveram resultados mais acurados em relação a previsões individuais (Clemem,

1989; Makridakis & Hibon, 2000; Stock & Watson, 2004; Amendola & Storti, 2008; Jeong & Kim, 2009; Costantini & Pappalardo, 2010; Wallis, 2011).

Mais recentemente, diversos trabalhos aplicaram a técnica de combinação de modelos de previsão a dados observados em certos casos de estudo como para fluxo de tráfego (Liu *et al.*, 2020), vendas no varejo (Aras *et al.*, 2017), agroindústria (Ueda *et al.*, 2018), logística portuária (Agostino *et al.*, 2019) e para variáveis ambientais (Camelo *et al.*, 2018; Hu *et al.*, 2020). Já para dados de consumo de energia elétrica, trabalhos recentes vêm utilizando modelos complexos, como os de aprendizado de máquinas e redes neurais, para serem combinados (Chou & Tran, 2018; Fan *et al.*, 2020). No entanto, a utilização de combinação de modelos não é usual no tratamento de dados de consumo de energia elétrica no Brasil.

A previsão acurada de carga é um tema de grande interesse para o setor elétrico uma vez que erros de previsão acabam por gerar custo financeiro elevado, seja ele por erro de previsão por excesso (overforecast) ou erro de previsão por falta (underforecast) (Serrão, 2003). O modelo SARIMA (AutoRegressivos Integrados de Médias Móveis com Sazonalidade) tem sido bastante utilizado na literatura para o tratamento de séries de consumo de energia elétrica, em grande parte por acoplarem a parte sazonal, tão característica deste tipo de série, dentro do modelo (Santos & Chaucoski, 2020; Araújo *et al.*, 2023; Marcos & Pontes Júnior, 2021; Bilgili & Pinar, 2023; Zhu *et al.*, 2022; Mohammed *et al.*, 2024).

Para o caso específico de previsão de demanda para o consumo de energia elétrica industrial no Brasil, alguns trabalhos recentes utilizam-se de análises comparativas de métodos de previsão para séries temporais de modo a encontrar um modelo mais adequado ao problema. Gontijo *et al.* (2017), para o caso nacional, concluiu que, o método ARMA foi o mais adequado no ponto de vista da análise do erro médio.

Já Marcos e Pontes Júnior (2021), apontaram um modelo baseado em rede neural convolucional como o mais indicado para o caso de demanda na região nordeste brasileira. Para a previsão do consumo de energia elétrica da região sudeste do Brasil, Santos e Chaucoski (2020), avaliaram modelos baseados em redes neurais LSTM e SARIMA e concluíram, para um horizonte

de curto prazo, que os modelos fornecem estimativas confiáveis para o consumo de energia elétrica da região.

No estudo de Taylor *et al.* (2006), os autores visam avaliar seis métodos de previsão de energia a um curtíssimo prazo, de apenas um dia. Dentre estes, ressaltou-se o ARIMA e o Holt-Winters, onde este último foi considerado mais simples de ser implementado, ainda apresentando resultados satisfatórios.

Deste modo, o presente trabalho tem por objetivo a construção e aplicação de um modelo combinado de série temporal, baseado na competição híbrida entre três modelos individuais de baixa complexidade, que busca prever o consumo médio de energia elétrica, em kWh (quilowatt-hora), por consumidor para o intervalo de um ano. Este tipo de abordagem combinada para dados de energia é escassa na literatura brasileira e, portanto, o estudo visa a disponibilização uma ferramenta eficiente que pode auxiliar no planejamento e tomada de decisões no setor elétrico.

2 Metodologia

O presente estudo é uma pesquisa quantitativa e experimental cuja questão de pesquisa proposta consiste em prever o consumo residencial de energia elétrica no Brasil em médio prazo (1 ano). O estudo será realizado utilizando dados obtidos na base pública de dados referente ao consumo de energia elétrica no Brasil disponibilizada pela Empresa de Pesquisa Energética [EPE](2022). Utilizou-se a base de dados de *Consumo Mensal de Energia Elétrica por classe*. Neste banco de dados, encontra-se o consumo de energia elétrica total em megawatt-hora (MWh), para todo o Brasil e rateado por cada região do país desde 2004 a 2021, dividido por classes, como também o número de consumidores totais e por classe. A fim de promover um estudo abrangente, optou-se por realizar a análise contemplando todos os estados do país, incluindo o Distrito Federal, e apenas os consumos da classe residencial.

Uma das premissas definidas neste estudo consiste em evitar na base de dados longas, de modo a reduzir o esforço computacional para sua manipulação e minimizar a presença de “falhas”. Ademais, apesar de obtida uma base de dados com muitos anos de histórico, sabe-se que, no

geral, este evento não é comum e frequentemente é marcado por bases incompletas ou com poucos anos passados.

Deste modo, visando a construir um modelo que agregue valor em uma aplicação real, optou-se pela previsão do modelo para o ano de 2021 e uma composição de até três anos para a base histórica, sendo estes: 2018, 2019 e 2020.

Ao fim do estudo, de modo a entender se a performance dos resultados gerados pelo novo modelo seria a mesma considerando mais anos de histórico, realizou-se um experimento adicional. Neste, utilizou-se cinco anos completos de histórico, neste caso, de 2016 a 2020, para gerar a previsão do ano seguinte, 2021.

Com as bases de Consumo Mensal e Quantidade de Consumidores Residenciais selecionadas, em kilowatt-hora por consumidor, são apresentadas Tabela 1 e na Figura 2.

Tabela 1

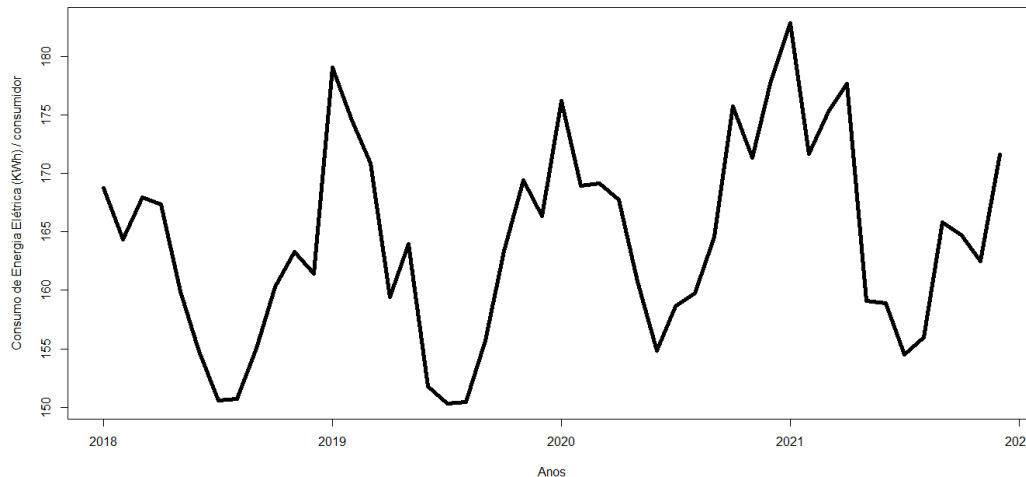
Consumo de Energia Elétrica (KWh) no Brasil, por consumidor por ano

Mês	2018	2019	2020	2021
Janeiro	168,778	179,073	176,213	182,878
Fevereiro	164,375	174,557	168,922	171,692
Março	167,922	170,894	169,162	175,290
Abril	167,369	159,420	167,745	177,671
Maió	159,867	163,954	160,781	159,079
Junho	154,793	151,791	154,851	158,901
Julho	150,587	150,310	158,626	154,526
Agosto	150,729	150,441	159,747	155,976
Setembro	155,006	155,713	164,562	165,815
Outubro	160,265	163,329	175,771	164,661
Novembro	163,264	169,424	171,363	162,494
Dezembro	161,400	166,375	177,958	171,615

Fonte: Elaboração própria do autor

Figura 2

Consumo de Energia Elétrica (kWh) no Brasil, por consumidor, de 2018 a 2021



Fonte: Elaboração própria do autor

3.1 Análise da série temporal

Matematicamente, uma série temporal unidimensional pode ser vista como um vetor $Z(t)$ de ordem $r \times 1$, dependente do também vetor $t = (t_1, t_2, \dots, t_r)$. O vetor t , usualmente, denota a discretização do tempo em r intervalos regulares e o vetor $Z(t)$ as r observações da variável em estudo em cada tempo t_i , onde $i=1, \dots, r$. Portanto, uma série temporal é uma sequência de números coletados em intervalos regulares durante um período de tempo.

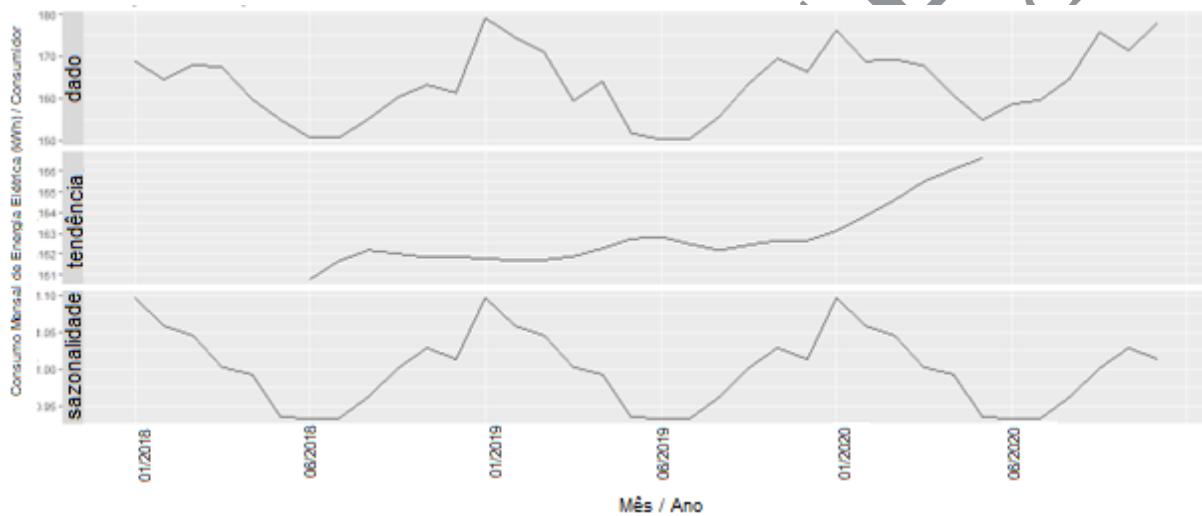
A realização de previsões, por si só, envolve muitas incertezas. Neste contexto, a combinação de previsões de modelos individuais em diversos casos se mostra eficaz pois possibilita captar um maior número possível de informações (Martins & Werner, 2014).

A fim de identificar possíveis comportamentos apresentados pela série escolhida, pesquisou-se as características de suas componentes. Em geral, as séries temporais podem ser decompostas em cinco componentes, que os modelos estatísticos buscam interpretar para gerar boas previsões para o futuro. São elas: nível, tendência, sazonalidade, ciclo e variáveis aleatórias.

A fim de entender seu comportamento no tempo e mapear quais componentes provavelmente poderiam ser mais privilegiadas na escolha dos modelos, obteve-se a decomposição de série temporal, apresentada na Figura 3.

Figura 3

Decomposição em componentes para a série de Consumo Mensal por Consumidor entre 2018 e 2020



Fonte: Elaboração própria do autor

É possível observar que na série temporal escolhida se destaca o fator de sazonalidade. A série apresenta menor consumo de energia elétrica nos meses de inverno no país e aumento significativo no consumo nos meses de verão, como já era esperado pela própria natureza dos dados.

3.2 Escolha dos Modelos Individuais

O objetivo desta etapa foi identificar os três modelos que melhor se adequassem à realidade, apresentando menor erro em suas previsões individuais para embasar o modelo

combinado. Escolheu-se 13 modelos a partir dos critérios: popularidade dentre o apresentado pelo levantamento bibliográfico; comportamento semelhante ao esperado para a série escolhida e originalidade em suas aplicações.

São eles: AED (Amortecimento Exponencial Duplo - Holt); AES (Amortecimento Exponencial Simples); AET (Amortecimento Exponencial Triplo - Holt Winters); BATS; DC (Decomposição Clássica STL); ETS (Suavização Exponencial); SARIMA; SARIMA com Fourier; SNAIVE; TBATS; TSLMS (Regressão Temporal com ênfase na componente sazonal); TSLMT (Regressão Temporal com ênfase na componente tendência) e TSLMTS (Regressão Temporal com ênfase em ambas as componentes sazonal e tendência).

Para a escolha dos três modelos que constituiriam a base de criação do novo modelo de previsão, realizou-se a aplicação de cada um, utilizando o histórico dos anos de 2018, 2019 e 2020, com previsão para 2021. Para determinar a acurácia dos modelos calculou-se as métricas de erros escolhidas pela relevância na literatura da área e estas serão descritas a seguir.

O RMSE, traduzido como ‘Raiz do Erro Quadrático Médio’, representa o desvio padrão dos erros de predição, sendo estes últimos referentes a distância existente entre os dados obtidos e a linha de regressão. Este apresenta duas particularidades: a aplicação da raiz quadrada, no intuito de deixar todos os resultados na mesma escala e a elevação ao quadrado, que, caso haja um erro muito grande, este se mostrará ainda mais significativo no cálculo. Assim, ressalta-se que quanto mais perto de 0 o RMSE se apresenta, menos erros são constatados pela amostra experimental de dados.

O MAD, traduzido como “Desvio Médio Absoluto”, é semelhante ao RMSE, entretanto, representa o valor absoluto, visto que se pode ter valores positivos e negativos, das diferenças entre a previsão e o dado real. A diferença prática entre estes consiste no fato do RMSE basear-se em uma ponderação que atribui mais importância a erros mais significativos, diferente do MAD, que não sofre com valores discrepantes. Entretanto, para erros pequenos, ambos entregam valores próximos e, portanto, conforme comentado anteriormente, deve atentar-se a valores menores (e mais próximos de 0), que representam modelos melhores.

O MAPE ou ‘Erro Percentual Absoluto Médio’ pela tradução é, por sua vez, uma métrica que apresenta a porcentagem da diferença entre a previsão e o real em relação aos valores constatados, ou seja, seu objetivo é mostrar quanto em média o modelo está errando com a sua previsão em relação aos dados reais. Vale ressaltar também que este indicador é bastante aplicado pelo fato deste considerar o valor absoluto, e, por isso, não compensar os erros positivos com os negativos. De acordo com Haluen (2020), o MAPE trata-se de um indicador de acerto de estimativa que varia entre 0% e 100%.

Theil’s U, também conhecido como “Coeficiente de Incerteza”, assim como o RMSE, atribui maior peso a erros maiores, o que auxilia na eliminação de métodos que os compõe. Diferente das outras métricas apresentadas, este possui uma interpretação diferente de seus resultados. Quando seu resultado é maior que 1, o uso daquele modelo não possui tanto valor, enquanto, quando menor que 1, possui. Assim, valores menores para melhores modelos são esperados para validação.

Para escolher os três modelos que iriam compor o novo modelo, definiu-se que o MAD seria o indicador a ser aplicado para todos os modelos, visto sua representatividade na literatura e por punir erros muito grandes. Assim, para cada previsão gerada, calculou-se o MAD para cada modelo. De acordo com essa proposta, os três modelos que apresentaram menor erro em relação a sua previsão, considerando 3 anos de histórico, foram: TSLMS, TSLMTS e SNAIVE.

Os modelos TSLMTS e TSLMS pertencem a família de modelos TSLM (*Time series linear model*) que consiste em um modelo de regressão linear adequado para séries temporais, oferecendo recursos adicionais. O TSLMTS é o modelo mais complexo, pois considera adicionalmente tanto a sazonalidade quanto a tendência. Por outro lado, o TSLMS é mais simples, focando apenas na sazonalidade. Já o último modelo, SNAIVE, se baseia na repetição dos valores do ano anterior para o mesmo mês, enfatizando a sazonalidade ao invés da tendência. Assim, a previsão desse modelo é igual ao valor real de 2020, refletindo as oscilações sazonais ao longo do ano. Ou seja, a previsão do modelo SNAIVE para março de 2021 será exatamente o mesmo valor observado em março de 2020, e a lógica se repete para todos os meses do ano.

Conforme esperado, os três modelos que mais se aproximaram dos valores reais registrados para o ano de 2021 e, portanto, foram escolhidos como base de construção do modelo combinado, assemelham-se por considerar a componente de sazonalidade, apontado anteriormente como uma forte característica da série estudada.

3.3 Cálculo dos Pesos do Modelo Proposto

O modelo proposto consiste na combinação linear convexa das previsões individuais obtidas pelos modelos TSLMTS, TSLMS e SNAIVE. Para isto, determinou-se que o respectivo peso de cada um seria estabelecido com função de sua performance na previsão para o ano de 2020. Dessa forma, a fim de simular a aplicação do novo modelo quando se dispõe apenas três anos de histórico, primeiramente são obtidas as previsões individuais para o ano de 2020, com base nos anos de 2018 e 2019. De acordo com a métrica de erro (MAD), obtém-se um peso para cada um dos modelos individuais. Estes pesos serão os coeficientes no modelo combinado gerado para a previsão de consumo de energia para o ano de 2021.

Para calcular o peso referente a cada modelo i (onde $i=1, 2, 3$), que seria atribuído na construção do cálculo da previsão do novo modelo, definiu-se que este representaria a significância de cada erro (ε_i) em relação a soma de todos estes. Nesta relação, modelos com menor erro ε_i teriam maior peso W_i . Deste modo, considerando que três erros seriam utilizados, os pesos W_i possuem a expressão dada pela Equação 1 de modo a garantir que a soma dos pesos resulte em 1.

$$W_i = 1 - \frac{2 \times \varepsilon_i}{\sum_1^3 \varepsilon_i} \quad (1)$$

3.4 Construção do Modelo Combinado

Para construir o modelo combinado, determinou-se que cada um dos três modelos teria sua participação ponderada de acordo com seus erros na previsão para o ano de 2020. Assim, determinou-se que o mesmo erro associado à previsão de 2020 seria replicado na construção do novo modelo com foco na previsão do consumo de energia mensal (em KWh) de 2021. A repetição deste erro é questionável, visto que não se pode afirmar que as mesmas variáveis existentes no ano de 2020 se repetiriam no ano seguinte. Entretanto, como o SNAIVE foi um dos três modelos que apresentaram previsões mais assertivas em relação a 2021 e este apenas replicou o dado real de 2020, pode-se inferir que não é esperado uma grande variação de contexto entre dois anos consecutivos.

Assim sendo, para construir a previsão de consumo mensal de energia elétrica (em KWh) de 2021 gerado pelo novo modelo, a cada mês deve-se considerar o peso associado a cada um dos três modelos, encontrado a partir do erro da previsão em relação a 2020, multiplicado por sua respectiva previsão para 2021. A previsão gerada para cada mês pelo modelo novo, portanto, pode ser conferida na Equação 2 a seguir:

$$P = W_1 \times N_1 + W_2 \times N_2 + W_3 \times N_3 \quad (2)$$

onde $W_1 + W_2 + W_3 = 1$;

P = previsão mensal gerada pelo novo modelo;

W_i = peso associado ao modelo i ;

N_i = previsão para o mês associado ao modelo i

Para determinar a previsão do consumo mensal de energia elétrica para o ano de 2021, resgatou-se as previsões geradas por cada modelo referente ao consumo mensal de energia elétrica de 2021 e obteve-se a previsão do novo modelo para o ano de 2021 a partir da soma do

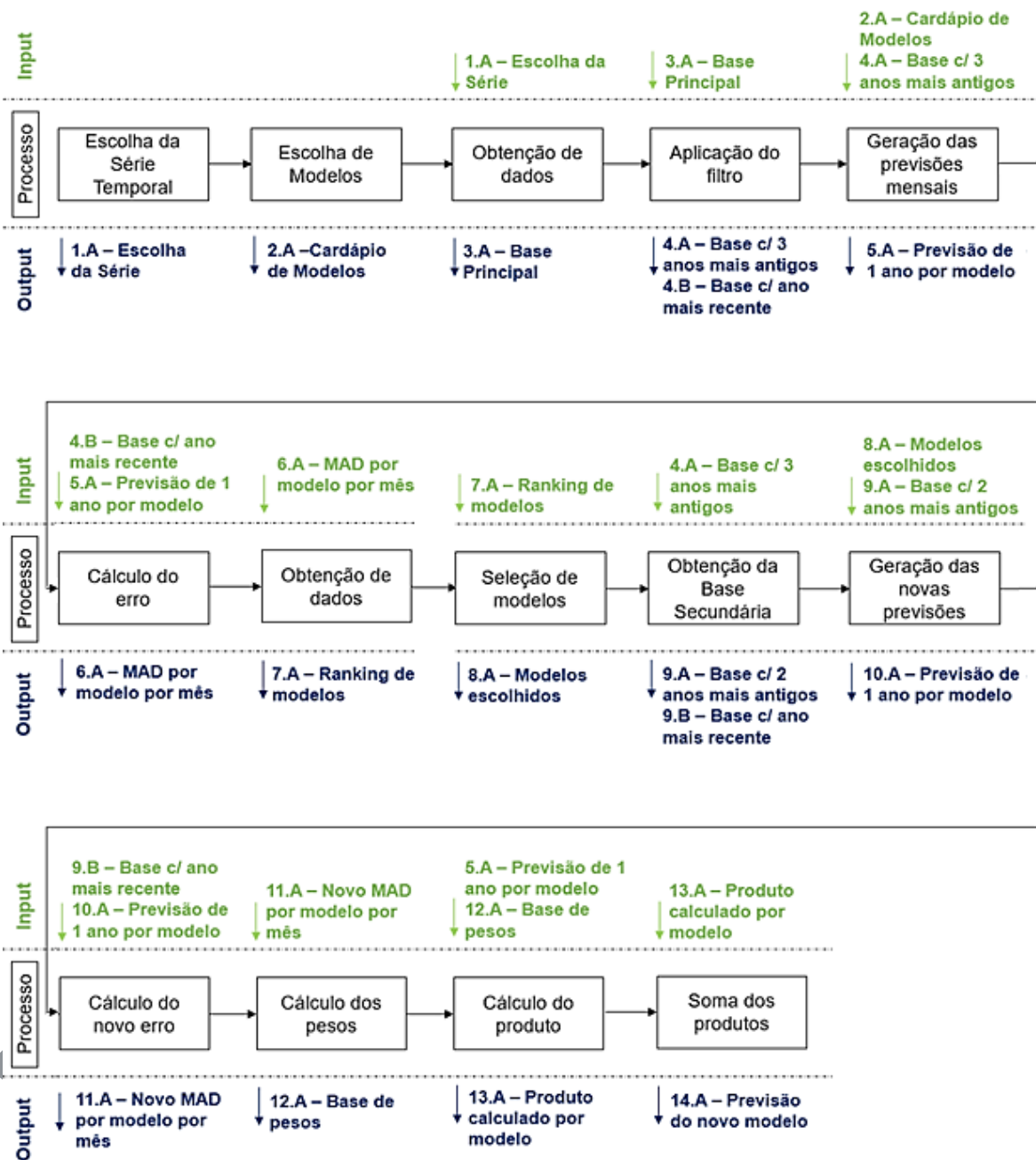
produto entre o peso do erro de previsão para 2020 e a previsão para 2021 do respectivo modelo associado a cada um dos meses de estudo.

De modo a ilustrar os passos apresentados, apresenta-se na Figura 4 um fluxograma com o resumo dos processos necessários, com seus devidos input e outputs, para a obtenção da previsão do novo modelo, a partir da premissa de utilizar três anos de histórico e três modelos para construir o novo:

ARTIGO ACEITO

Figura 4

Fluxograma da metodologia



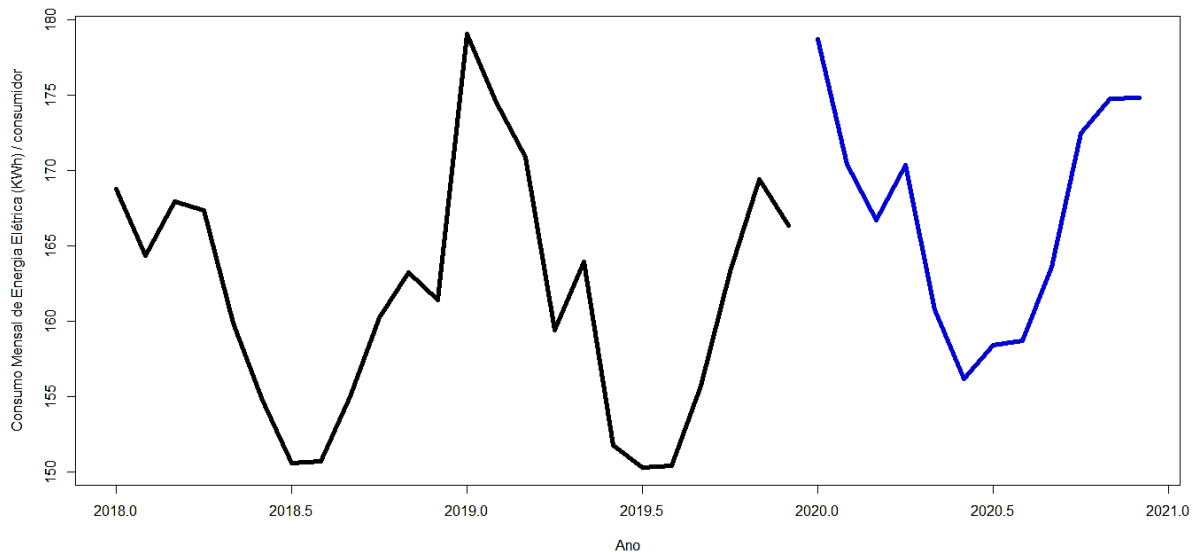
Fonte: Elaboração própria do autor

4 Resultados

A Figura 5 ilustra o histórico e a previsão gerada pelo modelo combinado considerando como histórico os anos de 2018, 2019 e 2020. Nela, é possível observar que, em uma análise comparativa com a previsão dos outros modelos, o modelo combinado mostra-se coerente e não destoante do esperado.

Figura 5

Previsão do modelo combinado para o ano de 2021 (em azul) e com o histórico base (em preto).



Fonte: Elaboração própria do autor

A fim de testar sua acurácia, calculou-se a previsão para os 12 meses do ano de 2021 e comparou-se o resultado obtido com a base de dados reais do referente ano. A fim de validar o modelo proposto e confrontar sua performance com os três modelos individuais, calculou-se os 4 indicadores apresentados na metodologia, a saber RMSE, MAE, MAPE e Theil's U, onde seus resultados são apresentados na Tabela 2.

Tabela 2

Indicadores de erro referentes a cada modelo em estudo

Modelo / Indicadores	RMSE	MAD	MAPE	Theil's U
TSLMTS	6,020	4,889	2,978	0,798
TSLMS	5,843	4,864	2,861	0,765
SNAIVE	6,340	5,557	3,307	0,839
Modelo Combinado	5,767	4,795	2,863	0,766

Fonte: Elaboração própria do autor

De acordo com a Tabela 2, nota-se que o modelo combinado apresentou o menor valor para os indicadores MAD e RMSE e comportamento similar ao modelo TSLMS nos indicadores MAPE e Theil's U, o que coloca o modelo proposto como protagonista nesta análise. Quando comparado com os outros dois modelos, TSLMTS e SNAIVE, o modelo combinado apresenta indicadores mais acurados e, portanto, mais assertivos. Assim, a partir dos indicadores calculados, o modelo aqui proposto mostrou-se como uma ferramenta eficaz para prever o consumo mensal de energia elétrica no curto prazo e com um histórico relativamente curto, considerando somente três anos.

A fim de evidenciar a hipótese de que mais anos de histórico impactariam na escolha dos três modelos iniciais e na performance do novo modelo construído, ajustou-se o histórico escolhido para cinco anos, gerando novos resultados descritos a seguir.

Para testar a hipótese em questão, repetiu-se todos os processos apresentados no fluxograma, ilustrado na Figura 4, alterando a premissa de utilizar três anos como histórico para cinco anos. Dessa forma, obteve-se a base de consumo mensal de energia elétrica em MWh para todo o Brasil e a quantidade de consumidores de 2016 a 2021, portanto adicionando os dados referentes aos anos 2016 e 2017 ao histórico base.

Para a escolha dos três modelos, agora com cinco anos de histórico, calculou-se a previsão gerada por cada um dos treze modelos selecionados para a competição e, mais uma vez, optou-se pela escolha dos três menores valores apresentados no cálculo da métrica MAD em relação a previsão gerada e o dado real.

Neste caso, notou-se a primeira discrepância ao utilizar um histórico com mais anos: o modelo SNAIVE, escolhido no cenário com três anos de histórico, é desbancado pelo modelo DC (Decomposição Clássica) e os outros dois modelos, TSLMSTS e TSLMS, permanecem como os dois modelos mais assertivos. Cabe comentar que o DC é um modelo que considera as quatro componentes apresentadas: nível, tendência, sazonalidade e ciclo; diferenciando-se por esta última, que, conforme indicado, é uma componente que pode ser identificada a partir de flutuações com duração superior a um ano.

Escolhido os três modelos e dispondo de suas previsões para 2021, novamente arquiva-se a base deste ano, a fim de construir o novo modelo apenas com dados que, teoricamente, estariam disponíveis, isto é, 2016, 2017, 2018, 2019 e 2020. Sendo assim, avança-se para o cálculo da performance apresentada por cada modelo em relação ao ano de 2020. Ou seja, utilizou-se os quatro anos restantes como histórico para geração da previsão e os dados reais de 2020 para calcular o MAD apresentado por cada modelo.

Em seguida, calculou-se o erro, que indica a participação de cada um dos três modelos em cada mês, a ser utilizado na construção do novo modelo e, seguindo a premissa de repetir a participação de cada um dos modelos no erro para 2020, constrói-se a previsão do novo modelo, indicada na Tabela 3.

Tabela 3

Previsão do consumo de energia elétrica (kWh) por consumidor para 2021

Meses	Novo Modelo
Janeiro	180,551
Fevereiro	176,124
Março	171,003
Abril	171,635
Maio	166,061
Junho	161,447
Julho	156,788
Agosto	157,068
Setembro	163,033
Outubro	167,591
Novembro	170,160
Dezembro	170,573

Fonte: Elaboração própria do autor

A fim de comparar a performance do novo modelo em comparação com os três outros modelos de forma individual, resgatou-se a base real referente ao consumo de energia elétrica (kWh) por consumidor em 2021 e calculou-se os 4 indicadores selecionados anteriormente, RMSE, MAD, MAPE e Theil's U, conforme apresentado na Tabela 4.

Tabela 4

1

Indicadores referentes a cada modelo

Modelo / Indicadores	RMSE	MAD	MAPE	Theil's U
TSLMTS	3,985	3,048	1,814	0,5191
TSLMS	5,917	5,048	2,969	0,7764
DC	6,026	5,147	3,150	0,7984
Modelo Combinado	4,258	3,699	2,220	0,5543

Fonte: Elaboração própria do autor

A partir dos resultados obtidos, o modelo combinado apresentou valores menos assertivos que o modelo TSLMTS, que prioriza as componentes de tendência e sazonalidade.

Ao comparar os resultados recém obtidos com os resultados calculados com três anos de histórico, apresentados anteriormente, entende-se que com um histórico com mais anos, o comportamento dos modelos, assim como suas previsões sofrem grandes alterações. Além disso, para o caso do modelo TSLMTS, por exemplo, a adição de mais dois anos no histórico fez com que este performasse melhor que o modelo construído a partir dos três selecionados, ou seja, ele sozinho apresenta previsões mais acuradas que quando apenas compõe um modelo combinado.

Estes resultados vão de encontro ao relatado na literatura. Segundo Camelo *et al* (2018), o uso de modelos combinados de previsão apresentou resultados mais acurados ao utilizar um histórico longo de 12 anos de dados. No entanto, o mesmo autor, apesar de também utilizar três modelos individuais como base, não realizou a combinação dos três diretamente e sim a análise de três casos distintos de combinação 2 a 2.

Em Hu *et al.* (2020), a combinação por média simples de modelos individuais complexos, resultou em um modelo comparativamente mais acurado que os demais para previsões a curto prazo (7 dias). Tais resultados diferem do presente estudo por apresentar uma combinação simples de modelos individuais complexos e com alto custo computacional, o que torna custosa própria combinação já para curto prazo.

Dessa forma, conclui-se que o modelo combinado é o mais indicado e apresenta uma boa performance nas previsões em séries com poucos anos de histórico. Tal fato é de grande valor para a pesquisa, visto que, conforme apontado previamente, a obtenção de históricos com muitos anos não é imediata, sendo muito comum a falta de disponibilidade de dados mais antigos. Além disso, buscava-se a apresentação de um modelo que fosse aplicável, de fácil implementação e tivesse bom resultado considerando que, na realidade, a disponibilização de dados para este tipo de abordagem é escassa.

Conclusão

Diante das inúmeras descobertas realizadas pelo homem, a energia elétrica é destacada por possibilitar diversas utilidades à sociedade, lhe proporcionando segurança, conforto e bem-estar. Assim, uma previsão assertiva mostra-se de extremo valor para guiar seu planejamento, prevenindo imprevistos e gerindo-a da melhor forma.

Este artigo apresenta a construção e aplicação de um modelo combinado de previsão para o consumo de energia elétrica (kWh) por consumidor, que ponderado com base nos modelos TSLMTS, TSLMS e SNAIVE, visa prever o comportamento da série de previsão para o intervalo de um ano. Para a construção do novo modelo, a principal premissa adotada consiste na repetição do mesmo erro associado a cada modelo para previsão do ano de 2020 com 2 anos de histórico, visto que, teoricamente, não se tem os dados reais daquilo que se tenta prever.

É possível concluir que o modelo proposto, apresenta uma melhor performance, quando comparado com os outros três escolhidos separadamente. Para o período de análise, os indicadores foram calculados a partir da comparação dos valores gerados com os dados reais registrados; onde o novo modelo, em dois dos quatro indicadores escolhidos (MAD e RMSE), apresentou valores menores em relação aos outros modelos de forma individual e empatou com os outros dois (MAPE e Theil's U), tornando-se o mais acurado e de melhor desempenho. Isso ressalta a importância de considerar uma abordagem combinada para previsão, ponderando o desempenho de diferentes modelos.

Adicionalmente, testou-se a mesma metodologia para um período de cinco de histórico. De acordo com os resultados, o modelo combinado ficou com o segundo melhor desempenho quando analisados os indicadores de erro, visto que o modelo TSLMTS, no histórico maior, conseguiu identificar as componentes e gerar previsões mais assertivas sozinho.

Assim, de modo geral, pode-se afirmar que o modelo proposto neste estudo pode ser considerado um bom modelo de previsão, em termos de acurácia, para prever o consumo mensal de energia elétrica (kWh) por consumidor para o curto prazo. Uma vez que é comum não dispor facilmente de dados históricos, acredita-se que o modelo combinado é uma ferramenta eficiente

e pode ser aplicado como suporte ao planejamento e a tomada de decisões no setor elétrico, já que apresenta previsões mais acuradas a curto prazo que os outros modelos individuais.

Quanto às contribuições para a teoria, o trabalho ressalta a importância da utilização de técnicas avançadas de análise de séries temporais e modelagem estatística na previsão de consumo de energia elétrica. Isso corrobora com a premissa de que a aplicação de métodos científicos rigorosos pode melhorar a compreensão e previsão de fenômenos complexos.

Como trabalho futuro, é possível sugerir a comparação de performances entre o modelo combinado aqui proposto e outros modelos combinados da literatura como, por exemplo, a combinação por variância mínima. Além disso, uma abordagem interessante seria a verificação da utilização de combinação não-linear entre os modelos individuais na construção do modelo combinado.

Referências

- Agostino, I., Araujo, R. S., de Oliveira Noronha, M., Fonseca, J. I. D. S., & Souza, A. M. (2019). Combinação de previsões aplicada à modelagem de operações: um estudo de caso em um terminal portuário. *Exacta*, 17(1), 99-110. <https://doi.org/10.5585/exactaep.v17n1.8327>
- Aras, S.; Deveci, İ. & Polat, C. (2017). Comparative study on retail sales forecasting between single and combination methods. *Journal of Business Economics and Management*, 18(5), 803-832. <https://doi.org/10.3846/16111699.2017.1367324>
- Araújo, E. G., Júnior, S. F. A. X., Barbosa, N. F. M., & Oliveira, T. A. (2023). Modelagem e previsão de modelos de séries temporais do consumo de energia elétrica na Região Nordeste do Brasil. *Sigmae*, 12(1), 10-28. <https://publicacoes.unifal-mg.edu.br/revistas/index.php/sigmae/article/view/2008>
- Bilgili, M., & Pinar, E. (2023). Gross electricity consumption forecasting using LSTM and SARIMA approaches: A case study of Türkiye. *Energy*, 284, 128575. <https://doi.org/10.1016/j.energy.2023.128575>
- Camelo, H., Lucio, P. S., Junior, J. B. V. L., de Carvalho, P. C. M., & dos Santos, D. V. G. (2018). Innovative hybrid models for forecasting time series applied in wind generation based on

the combination of time series models with artificial neural networks. *Energy*, 151, 347-357. <https://doi.org/10.1016/j.energy.2018.03.077>

Carneiro, A (2014). *Previsão do consumo de energia elétrica a curto prazo, usando combinações de métodos univariados* (Dissertação de Mestrado). Universidade Federal de Juiz de Fora, Minas Gerais, Brasil.

Chou, J., Tran, D. (2018). Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. *Energy*, 165(B), 709-726. <https://doi.org/10.1016/j.energy.2018.09.144>.

Empresa de Pesquisa Energética – EPE (2022). *Consumo Mensal de Energia Elétrica por Classe (regiões e subsistemas)*. (n.d.). Retrieved August 29, 2022, from <https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/consumo-de-energia-eletrica>

Fan, G. F., Wei, X., Li, Y. T., & Hong, W. C. (2020). Forecasting electricity consumption using a novel hybrid model. *Sustainable Cities and Society*, 61, 102320. <https://doi.org/10.1016/j.scs.2020.102320>

Farias, L. M., & Sellitto, M. A. (2013). Uso da energia ao longo da história: evolução e perspectivas futuras. *Revista Liberato*, 12(17), 07–16. <https://revista.liberato.com.br/index.php/revista/article/view/164>

Gontijo, T. S., Rodrigues, F. D. M., de Cássio Rodrigues, A., da Silva, S. A., & de Azevedo, A. A. (2017). Consumo industrial de energia elétrica: um estudo comparativo entre métodos preditivos. *Brazilian Journal of Production Engineering*, 3(3), 31-45. https://periodicos.ufes.br/bjpe/article/view/v3n3_03

Hibon, M., & Evgeniou, T. (2005). To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting*, 21(1), 15–24. <https://doi.org/10.1016/j.ijforecast.2004.05.002>

Hu, H., Zhang, J. & Li, T. (2020). A comparative study of VMD-based hybrid forecasting model for nonstationary daily streamflow time series. *Complexity*, 1-21. <https://doi.org/10.1155/2020/4064851>

Liu, B., Tang, X., Cheng, J., & Shi, P. (2020). Traffic flow combination forecasting method based on improved LSTM and ARIMA. *International Journal of Embedded Systems*, 12(1), 22-30. <http://arxiv.org/abs/1906.10407>

- Lorenzo, H. C. (2001). O setor elétrico brasileiro: passado e futuro. *Perspectivas: Revista de Ciências Sociais*. São Paulo, 24-25: 147-170.
<https://periodicos.fclar.unesp.br/perspectivas/article/view/406>
- Marcos, I. P., & Pontes Júnior, A. P. (2021). Previsão do Consumo de Energia Elétrica na Região Nordeste do Brasil. *Revista de Engenharia e Pesquisa Aplicada*, 6(3), 21–30.
<https://doi.org/10.25286/rep.v6i3.1684>
- Martins, V. L. M., & Werner, L. (2014). Comparação de previsões individuais e suas combinações: um estudo com séries industriais. *Production*.
<https://lume.ufrgs.br/handle/10183/115339?locale-attribute=es>
- Mohammed, H. H., Asem, A. & El-Bakry, H. (2024). Advanced Time Series Forecasting Models for Electricity Demand Prediction: A Comparative Study. *Fusion: Practice and Applications*, 15(1), 19-9. <https://doi.org/10.54216/FPA.150102>
- Morettin, P. A. & Tolo, C. M. D. C. (2018). *Análise de séries temporais*. 3a ed. São Paulo, Edgard Blucher. 474p.
- Neto, A. B., Corrêa, W. L. R., & Perobelli, F. S. (2016). Consumo de Energia e Crescimento Econômico: uma Análise do Brasil no Período 1970-2009. *Análise Econômica*, 34(65).
<https://doi.org/10.22456/2176-5456.44622>
- Reis, L. B. (2011). *Geração de energia elétrica*. 2a ed. São Paulo, Manole. 203p.
- Serrão, F. C. C. (2003). *Modelo de Previsão de Carga de Curto Prazo Utilizando Redes Neurais e Lógica Fuzzy* (Doctoral dissertation, PUC-Rio). https://www2.dbd.puc-rio.br/pergamum/tesesabertas/0116365_03_pretexto.pdf
- Santos, J.A.A. & Chaucoski, Y. (2020). Previsão do Consumo de Energia Elétrica na Região Sudeste: Um Estudo de Caso Usando SARIMA e LSTM. *Revista Cereus*, v. 12, n. 4, p. 93-104. <http://ojs.unirg.edu.br/index.php/1/article/view/3255>
- Taylor, J. W., de Menezes, L. M., & McSharry, P. E. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, 22(1), 1–16. <https://doi.org/10.1016/j.ijforecast.2005.06.006>
- Ueda, R. M., Agostino, Í., Marasca, L., da Silva, R. B., & Souza, A. M. (2018). Combinação de técnicas de previsão de demanda: aplicação em uma agroindústria de nutrição animal. *Exacta*, 16(2), 83-93. <https://doi.org/10.5585/exactaep.v16n2.7295>

Zhu, X., Song, Z., Sen, G., Tian, M., Zheng, Y. & Zhu, B. (2022). Prediction study of electric energy production in important power production base, China. Scientific Reports, 12(1), 21472. <https://doi.org/10.1038/s41598-022-25885-w>

ARTIGO ACEITO