57º SBPO
SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL
Gramado • RS
DE 5 A 9 DE OUTUBRO DE 2025

GERAÇÃO DE CENÁRIOS PARA SÉRIES TEMPORAIS DE DEMANDA DE ETANOL: UMA ANÁLISE A PARTIR DE MÉTODOS CLÁSSICOS E TÉCNICAS DE *ENSEMBLE*

Luiza Magalhães Henriques, Pedro Paixão Borges, Carolina Monteiro Carneiro da Silva, Fernando Luiz Cyrino Oliveira
Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)
Rua Marquês de São Vicente, 225 − Gávea, Rio de Janeiro, RJ − Brasil − 22451-000
luizampbraga@gmail.com, pedropaixaob@gmail.com, monteirocarneirosc@gmail.com, cyrino@puc-rio.br

RESUMO

O etanol desempenha papel fundamental na descarbonização do setor energético, ganhando destaque na transição para fontes renováveis. Este estudo explora o uso de métodos de séries temporais para gerar cenários de demanda desse biocombustível nos Estados Unidos, com dados de 1981 a 2024. Para mitigar os impactos da pandemia de Covid-19, aplicou-se o método *Seasonal Naive* para remover *outliers*. Foram gerados 100 cenários de demanda usando decomposição STL, transformação Box-Cox e *Moving-Block Bootstrap*. Três abordagens preditivas - *Seasonal Naive*, ETS e ARIMA - foram avaliadas com e sem *bagging* para capturar dinâmicas temporais de consumo. A avaliação, baseada nas métricas RMSE, MAPE e DTW, destaca a robustez dos cenários com *bagging* e o potencial do ARIMA, combinado com essa técnica, para preservar padrões estruturais da série nas previsões futuras. Os resultados sugerem a utilidade dos métodos para planejamento sob incerteza no setor de biocombustíveis.

PALAVRAS CHAVE. Demanda de etanol, Séries temporais, Cenários de consumo, Bootstrap.

Tópicos (EN&OG, EST&AM)

ABSTRACT

Ethanol plays a key role in the decarbonization of the energy sector, gaining prominence in the transition to renewable energy sources. This study explores the use of time series methods to generate demand scenarios for this biofuel in the United States, using data from 1981 to 2024. To mitigate the impacts of the COVID-19 pandemic, the Seasonal Naive method was applied to remove outliers. A total of 100 demand scenarios were generated using STL decomposition, Box-Cox transformation, and Moving-Block Bootstrap. Three predictive approaches-Seasonal Naive, ETS, and ARIMA-were evaluated with and without bagging techniques to capture temporal dynamics of consumption. The evaluation, based on RMSE, MAPE, and DTW metrics, highlights the robustness of the scenarios generated with bagging and the potential of ARIMA combined with this technique to preserve structural patterns of the series in future forecasts. The results suggest the usefulness of these methods for planning under uncertainty in the biofuel sector.

KEYWORDS. Ethanol demand. Time series. Consumption scenarios. Bootstrap.

Paper topics (EN&OG, EST&AM)

DOI: 10.59254/sbpo-2025-212434
https://proceedings.science/p/212434?lang=pt-br

1. Introdução

Nos últimos anos, a preocupação global com os impactos negativos do consumo de combustíveis fósseis tem aumentado, evidenciando a necessidade urgente de uma transição para fontes de energia mais sustentáveis e de baixo carbono. Diversas iniciativas têm sido implementadas para promover a descarbonização do setor energético, com destaque para fontes renováveis como energia solar e eólica. Nesse contexto, a biomassa — derivada de materiais orgânicos como plantas, madeira e resíduos agrícolas — surge como uma alternativa significativa, ainda pouco explorada.

Os combustíveis derivados da biomassa possuem um enorme potencial para substituir fontes fósseis em diversas aplicações, como geração de energia elétrica, siderurgia, produção de cimento e aviação. Entre eles, o etanol destaca-se como uma solução consolidada e versátil, amplamente utilizada no setor de transportes. Produzido principalmente a partir de cana-de-açúcar e milho, o etanol contribui para a redução das emissões de gases de efeito estufa e apresenta vantagens como a compatibilidade com motores existentes e a possibilidade de ser misturado a combustíveis fósseis, facilitando a transição energética.

Nos Estados Unidos, mais de 98% da gasolina comercializada contém algum percentual de etanol, sendo a mistura mais comum o E10 (10% etanol e 90% gasolina). O E85 (51% a 83% de etanol) é destinado a veículos *flex-fuel* e integra os esforços do país para ampliar a substituição dos combustíveis fósseis. Atualmente, cerca de 93% do etanol combustível norte-americano é produzido a partir do milho [U.S. Department of Energy, 2024].

A promulgação da EISA [2007] estabeleceu padrões obrigatórios para a incorporação de combustíveis renováveis no setor de transporte americano, transformando rapidamente a demanda pelo etanol no país. Como resposta, entre 2006 e 2008, a produção local desse combustível cresceu cerca de 90%, passando de aproximadamente 4,88 bilhões para 9,31 bilhões de galões. Mais recentemente, fatores como o aumento da pressão internacional por práticas sustentáveis e a pandemia de Covid-19 afetaram a demanda, disponibilidade e os preços de etanol a nível global [Mizik et al., 2020].

Compreender a dinâmica da demanda por etanol nos Estados Unidos é essencial para impulsionar a expansão de negócios sustentáveis no setor energético no país. A identificação de padrões de consumo é fundamental para empresas, investidores e formuladores de políticas, pois permite o planejamento eficiente das cadeias produtivas e dos investimentos.

Nesse contexto, considerando a incerteza associada à evolução dessa demanda, a análise de múltiplos cenários torna-se crucial. Essa abordagem possibilita a aplicação de modelos de otimização estocástica, que auxiliam na tomada de decisões mais robustas, considerando tanto o cenário médio quanto cenários extremos. Além disso, tais estratégias de modelagem permitem a inclusão de diferentes níveis de aversão ao risco, promovendo a resiliência do setor.

Com base nessa motivação, este artigo propõe a geração de cenários para a demanda de etanol combustível, oferecendo uma ferramenta robusta para planejar e dimensionar negócios diante das incertezas. Para isso, utiliza-se a base de dados da *United States Department of Agriculture*, que disponibiliza uma série histórica com dados mensais sobre o consumo de etanol combustível nos Estados Unidos, abrangendo o período de janeiro de 1981 a março de 2024 [USDA, 2024].

O artigo está estruturado da seguinte forma. Iniciamos com uma breve revisão da literatura sobre biocombustíveis e sua modelagem por meio de séries temporais. Em seguida, apresentamos a metodologia do estudo, abordando os métodos utilizados, as etapas de tratamento dos dados, a geração de cenários e a realização das previsões. A seção seguinte apresenta os resultados dos modelos, ilustrados por gráficos e tabelas, seguidos de uma análise detalhada no capítulo de discussões. Por fim, concluímos com considerações finais e sugestões para futuras pesquisas.

DOI: 10.59254/sbpo-2025-212434
https://proceedings.science/p/212434?lang=pt-br

2. Revisão de literatura

A adoção em larga escala de biocombustíveis enfrenta desafios significativos, como altos custos de produção e barreiras financeiras [Gironès et al., 2017]. Compreender as dinâmicas de oferta e demanda de biocombustíveis é crucial para a gestão eficiente de recursos e o planejamento de políticas públicas que promovam o uso de fontes renováveis em cada país [Alizadeh et al., 2020]. Neste contexto, previsões associadas ao consumo dessas energias alternativas desempenham um papel estratégico, tendo valor fundamental na transição rumo a matrizes energéticas mais sustentáveis.

Diversos estudos têm se dedicado à modelagem da demanda de biocombustíveis. Luchansky e Monks [2009], por exemplo, reconheceram a importância de projetar a curva de demanda por etanol e propuseram um modelo de regressão para analisar a sensibilidade da demanda frente às variações de preço. Entretanto, para avaliações prospectivas de demanda, seria necessário dispor de projeções consistentes para a série de preços. Outros trabalhos, como o de Dey et al. [2023], exploraram a previsão de consumo de etanol em países que buscam ampliar o seu uso, como a Índia. Nesse estudo, a demanda de gasolina foi utilizada como direcionadora para a previsão de consumo de etanol, sendo testados modelos de regressão linear, não linear e ARIMA, com destaque para o desempenho superior deste último.

Mais recentemente, Baratsas et al. [2024] desenvolveram um framework híbrido para a seleção de modelos de previsão de preços de produtos energéticos, incluindo o etanol combustível. O framework utilizou técnicas avançadas, como redes neurais profundas, métodos de *bootstrapping* e agregação. Apesar da robustez metodológica, o foco do trabalho permaneceu restrito à previsão de preços.

Entre os métodos amplamente utilizados para previsão de séries temporais, destacam-se os modelos ETS e ARIMA utilizados conjuntamente com técnicas de *Bagging* para consumo de gás natural na Europa [Meira et al., 2022]. Adicionalmente, métodos mais recentes, como os *Grey models*, mostraram bons resultados na modelagem de demanda de etanol nos Estados Unidos, Índia e China, ampliando as opções disponíveis para previsões nesse setor, mas sem contemplar a geração de cenários [Javed et al., 2020].

Além das questões relativas à escolha do modelo, um dos maiores desafios na previsão de séries temporais de etanol está na pluralidade de fatores que influenciam simultaneamente a oferta e a demanda. Nesse contexto, abordagens baseadas na geração de cenários surgem como uma solução promissora, pois incorporam a estocasticidade inerente aos dados. Estudos recentes em outros setores demonstram que técnicas de *bootstrap* e *bagging* podem melhorar a qualidade das previsões, ao capturar incertezas e variações nos dados [Dantas et al., 2017; Petropoulos et al., 2018]. Essas metodologias têm se mostrado particularmente úteis no setor de energia, ganhando relevância na maioria dos trabalhos acadêmicos e competições recentes [Meira et al., 2022].

Apesar dos potenciais ganhos a partir desse tipo de abordagem, não foram encontrados na literatura trabalhos que discutam a construção de cenários de demanda para o mercado de etanol. Nesse sentido, o presente trabalho busca explorar esse *gap* ao adotar estratégias de modelagem baseadas em *Bootstrap*, tendo como pano de fundo o contexto americano. A metodologia detalhada para o desenvolvimento dessas estratégias é apresentada a seguir.

3. Métodos

A metodologia deste trabalho foi estruturada como uma sequência de etapas interdependentes, esquematizadas na Figura 1 e detalhadas nas subseções a seguir. Cada passo foi planejado para abordar aspectos específicos do problema de geração de cenários de consumo do etanol, passando desde o tratamento dos dados até a avaliação dos modelos.

DOI: 10.59254/sbpo-2025-212434
https://proceedings.science/p/212434?lang=pt-br

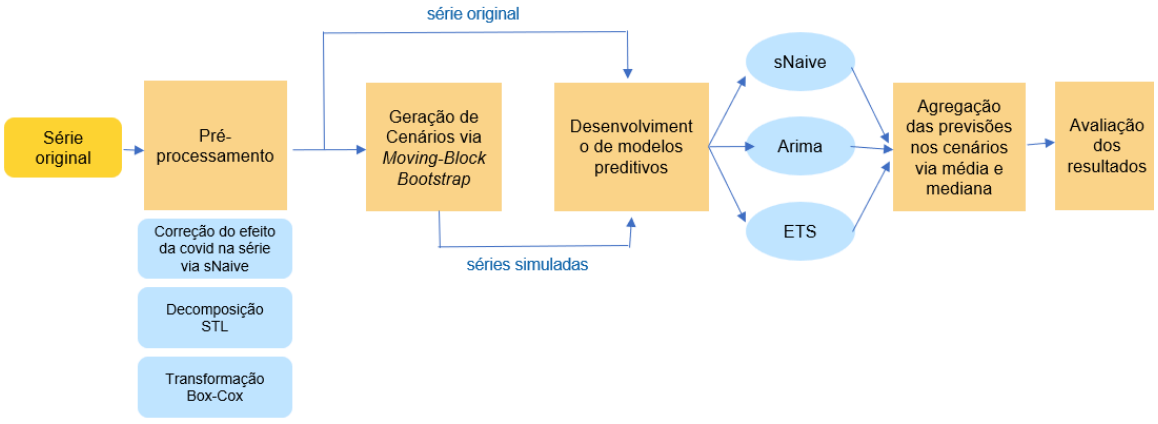


Figura 1: Esquema representando a metodologia de pesquisa

3.1. Análise exploratória de dados

O desenvolvimento desta pesquisa se iniciou com uma análise exploratória das séries temporais de consumo de etanol, com o objetivo de compreender suas principais características. Os dados utilizados possuem granularidade mensal, abrangendo inicialmente o período de 1981 a 2024. No entanto, para este estudo, optou-se por delimitar o intervalo de análise a partir de 2013, com base na observação de um comportamento mais estável nas curvas de consumo a partir deste período.

Como parte da análise, a técnica *Seasonal-Trend decomposition using LOESS* (STL) foi aplicada para decompor a série em componentes de tendência, sazonalidade e resíduos [Cleveland et al., 1990]. Complementarmente, as funções de autocorrelação (ACF) e autocorrelação parcial (PACF) da série original e de suas diferenciações foram analisadas para capturar relações autorregressivas e dependências temporais. Finalmente, os testes estatísticos Dickey-Fuller aumentado (ADF) e KPSS foram utilizados para avaliar a estacionariedade da série temporal e assegurar que as condições necessárias para a aplicação de modelos de previsão nas etapas posteriores do estudo sejam atendidas. A fase de análise exploratória de dados termina com a divisão da série original em conjuntos de teste e treino, tomando como limiar o mês de Mar/2023. Como resultado, o *set* de treino é definido no período entre Abr/2023 e Mar/2024.

3.2. Pré-processamento e tratamentos iniciais

Assim como ocorre em diversas séries temporais relacionadas ao consumo, o mercado de etanol foi significativamente impactado pela pandemia de Covid-19 (Ramsey et al. [2023]). Esse efeito é evidente em uma análise visual preliminar, que revela uma disrupção nas características da série temporal entre Jan/2020 e Mar/2021. Antes de realizar qualquer etapa de geração de cenários ou previsão, torna-se indispensável o tratamento adequado da série, a fim de mitigar o impacto dessa sequência de *outliers*. Esse ajuste é fundamental para evitar que distorções influenciem negativamente a inferência dos parâmetros dos modelos, comprometendo a confiabilidade das previsões futuras.

Para este trabalho, os dados do período da Covid-19 foram tratados por meio de substituição utilizando o método *Seasonal Naive*. Essa abordagem baseia-se na premissa de que o valor de uma observação em um determinado período pode ser estimado diretamente pelo valor correspondente na mesma posição sazonal do ciclo anterior. A escolha deste método se justifica por sua simplicidade e eficácia em cenários onde a série apresenta alta sazonalidade e pouca tendência, o que é o caso da série temporal nos anos anteriores e posteriores à pandemia.

DOI: 10.59254/sbpo-2025-212434
https://proceedings.science/p/212434?lang=pt-br

Somando-se ao tratamento dos dados para correção do consumo durante a pandemia, foi aplicada a transformação de Box-Cox (Box e Cox [1964], Bickel e Doksum [1981]). Trata-se de uma técnica com o objetivo de estabilizar a variância da série temporal e remover possíveis características heterocedásticas, sendo definida pela função:

$$w_t = \begin{cases} \frac{sign(y_t)|y_t|^{\lambda}-1}{\lambda}, & \text{se } \lambda \neq 0, \\ \ln(y_t), & \text{se } \lambda = 0, \end{cases}$$

O parâmetro λ pode variar de acordo com a série temporal e o tipo de transformação desejada. No presente trabalho, adotamos a estimação de [Guerrero, 1993], que propõe um algoritmo para cálculo automatizado do valor deste parâmetro. As transformações realizadas são essenciais para restabelecer a consistência dos dados, permitindo que as etapas subsequentes de modelagem e previsão sejam realizadas de maneira robusta e alinhada com as condições originais do mercado de etanol, antes da mudança de comportamento causada pela pandemia.

3.3. Geração de Cenários com Bootstrap

Com o objetivo de refletir as incertezas inerentes ao processo gerador de dados e aos possíveis métodos de previsão posteriores, parte fundamental do presente trabalho é a construção de cenários de consumo de etanol. Para garantir a aderência destes cenários às características da série original, o método *Moving-Block Bootstrap* (MBB) foi adotado. O método, proposto originalmente por Künsch [1989], se inicia por meio da divisão da série em blocos de tamanho fixo. Para uma série de tamanho n e um bloco de tamanho l, temos $n - l + 1$ possíveis blocos, incluindo sobreposições. Por fim, os blocos são amostrados aleatoriamente com reposição e reordenados de diferentes formas, permitindo a geração de novas séries temporais.

Tendo em conta as permutações necessárias para construir as novas séries a partir dos blocos, no entanto, notamos que a técnica de MBB só pode ser aplicada em séries estacionárias. Este requisito pode ser atendido ao aplicar o MBB somente no componente residual (*remainder*) da decomposição STL da série. Os resíduos *bootstrapeados*, então, são posteriormente somados aos componentes de tendência e sazonalidade originais para gerar as séries simuladas. Utilizando esta técnica, foi possível gerar 100 cenários alternativos de consumo de etanol, que foram avaliados por meio da análise dos seguintes elementos:

- **Medidas de tendência central:** Verificação da proximidade dos valores centrais das séries simuladas, como média e mediana, com relação à série original;
- **Função de Autocorrelação (ACF)** Comparação da ACF nos dois casos, objetivando compreender se as dependências temporais observadas nos cenários refletem as dependências nos dados reais;
- **Distribuição dos dados mensais:** Comparação da distribuição dos valores das séries em meses pré-selecionados (ex. Janeiro), para avaliar a aderência dos cenários simulados à distribuição sazonal dos dados originais.

3.4. Modelagem e previsão

Após o processamento dos dados e geração de cenários a partir do conjunto de treino, o próximo passo consiste na escolha dos métodos de previsão. Com base na literatura, foram escolhidos três tipos de modelos, cuja descrição resumida pode ser encontrada a seguir:

DOI: 10.59254/sbpo-2025-212434
https://proceedings.science/p/212434?lang=pt-br

- **Seasonal Naive (SNaive):** Como já apresentado anteriormente, as previsões no método *Seasonal Naive* se dão a partir da replicação dos últimos valores sazonais conhecidos. No caso deste trabalho, isto representa a previsão dos valores futuros de demanda de etanol a partir das estatísticas de consumo observadas exatamente 12 meses antes.

- **ETS (Error, Trend, Seasonality):** De forma simplificada, a família de modelos ETS tem como base a modelagem dos componentes da série temporal por meio de equações recursivas de caráter exponencial. No total, temos 30 variações possíveis para os modelos ETS, diferindo de acordo com o tipo de especificação adotada para as equações de tendência, sazonalidade e erro. A tendência pode ser inexistente, aditiva, multiplicativa, aditiva amortecida ou multiplicativa amortecida; a sazonalidade pode ser inexistente, aditiva ou multiplicativa; e o erro pode ser modelado de forma aditiva ou multiplicativa [Hyndman et al., 2008].

- **ARIMA (AutoRegressive Integrated Moving Average):** A família de modelos ARIMA(p,d,q) é formada por três componentes: autoregressivo (AR), integração (I) e médias móveis ou *moving average* (MA). A parte autoregressiva é responsável por capturar a relação linear entre uma observação e seus valores passados, sendo especificada pelo parâmetro p. A componente de integração (I) se refere ao número de diferenciações (d) aplicadas à série para torná-la estacionária. Finalmente, a equação de médias móveis representa a dependência linear entre uma observação e os erros residuais passados, especificada pelo parâmetro q. A família de modelos ARIMA sazonal (SARIMA) é uma extensão do modelo ARIMA que incorpora a capacidade de modelar padrões sazonais a partir de parâmetros adicionais P, D, Q. Estes representam respectivamente as componentes autorregressiva, integrada e de média móvel sazonais. Nesse caso, definimos também um parâmetro s representando o período de sazonalidade da série, sendo igual a 12 neste trabalho.

A implementação dos modelos e escolha dos hiperparâmetros foram feitas de maneira automática utilizando o pacote *forecast*, presente na linguagem R (Hyndman e Khandakar [2008]). Além dos tipos de modelos distintos, as estratégias para modelagem também variaram com relação à inclusão ou não de estratégias de *bagging*. Temos dois casos:

1. **Previsão a partir de série original (sem *bagging*):** Nessa abordagem, os modelos SNaive, ETS e ARIMA foram aplicados diretamente aos dados originais, sem qualquer manipulação adicional.

2. **Previsão a partir de cenários e agregação (com *bagging*):** Nesta estratégia, os modelos foram aplicados aos cenários gerados com o *Moving-Block Bootstrap* e as previsões finais foram construídas a partir de agregação via mediana e média.

A variação nas estratégias de modelagem teve como objetivo avaliar a capacidade preditiva dos modelos sob diferentes condições de incerteza e para comparar o impacto das técnicas de *bagging* na precisão das previsões geradas. Dessa forma, tomando como base o conjunto de treino entre Janeiro/2013 e Março/2023, foram realizadas previsões para o período entre Abril/2023 e Março/2024.

3.5. Avaliação

Para avaliar a qualidade das previsões geradas pelos modelos, foram utilizadas as métricas RMSE (*Root Mean Squared Error*), MAPE (*Mean Absolute Percentage Error*) e DTW *(Dynamic Time Warping)*.

DOI: 10.59254/sbpo-2025-212434
https://proceedings.science/p/212434?lang=pt-br

O RMSE mede o erro absoluto médio das previsões em relação aos valores reais do conjunto de teste. Temos, assim, a formulação:

$$\text{RMSE}(Y, \hat{Y}) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2},$$

onde Y e $\hat{Y}$ são as séries real e prevista, y_t representa os valores pontuais reais, $\hat{y}_t$ as previsões pontuais, e N o número total de observações. O RMSE é sensível a desvios grandes, atribuindo maior peso a erros maiores devido à elevação ao quadrado. Isso o torna uma métrica importante para identificar inconsistências significativas no desempenho do modelo de previsão.

Complementando o valor absoluto obtido pelo RMSE, o MAPE mede o erro percentual médio das previsões em relação aos valores reais, sendo calculado como:

$$\text{MAPE}(Y, \hat{Y}) = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100,$$

no qual N representa o número total de observações. Essa métrica fornece uma medida percentual do erro médio, permitindo interpretar a precisão do modelo em relação à magnitude dos dados reais. O MAPE é especialmente útil em contextos onde a escala dos dados varia significativamente, facilitando a comparação entre diferentes modelos ou séries temporais.

Somando-se às outras avaliações, a métrica DTW pode ser utilizada para capturar a similaridade entre duas séries temporais. Uma característica distintiva dessa métrica é sua capacidade de comparar a forma geral das séries, sendo menos afetada por possíveis deslocamentos ou ”mudanças de velocidade”na evolução temporal dos dados. Isso ocorre pois, diferentemente do RMSE e do MAPE, que calculam os erros exclusivamente pela comparação ponto a ponto no mesmo instante temporal, a DTW encontra o alinhamento ótimo entre diferentes pontos das séries. Dessa forma, a distância entre duas séries temporais usando a métrica DTW pode ser expressa como:

$$\text{DTW}(Y, \hat{Y}) = \min \sum_{(i,j)\in P} d(y_i, \hat{y}_j),$$

onde $d(y_i, \hat{y}_j)$ é a distância entre os pontos y_i e $\hat{y}_j$, e P é o caminho de alinhamento ótimo. No contexto deste trabalho, a DTW foi utilizada para verificar a semelhança estrutural das previsões em relação aos padrões sazonais e às tendências da série real, complementando a análise de erro absoluto fornecida pelo RMSE e a de erro relativo pelo MAPE.

4. Resultados Empíricos

Nesta seção, apresentamos e validamos os resultados da geração de cenários utilizando o método *Moving-Block Bootstrap*, detalhando cada etapa do processo. Em seguida, abordamos a etapa de agregação dos cenários gerados (*bagging*), com o objetivo de apoiar a seleção do modelo de previsão mais adequado. A agregação, realizada por meio do cálculo da média ou mediana, permite a comparação direta com um segmento de teste da série histórica, possibilitando a identificação do modelo com melhor desempenho, com base nas métricas de acurácia previamente discutidas no item 3.5.

Os dados históricos foram previamente ajustados para considerar os tratamentos relativos ao período da pandemia, que compreende de janeiro de 2020 a fevereiro de 2021. A Figura 2 apresenta a série real com identificação do tratamento realizado. Na figura, o particionamento destacado em vermelho corresponde às substituições realizadas na série original, com o objetivo de

DOI: 10.59254/sbpo-2025-212434
https://proceedings.science/p/212434?lang=pt-br

mitigar os efeitos da pandemia nos cenários gerados. Adicionalmente, foi realizada a redução de variância da série após aplicação do método Box-Cox, gerando a série que pode ser observada na Figura 3.

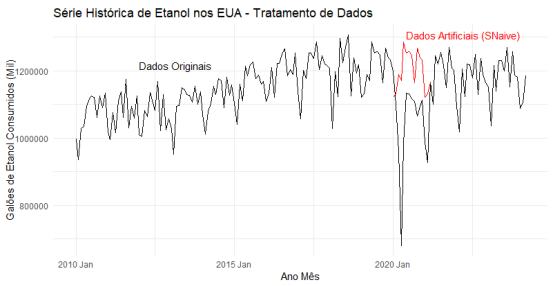


Figura 2: Série histórica de consumo de etanol com dados artificiais via *Seasonal Naive*.

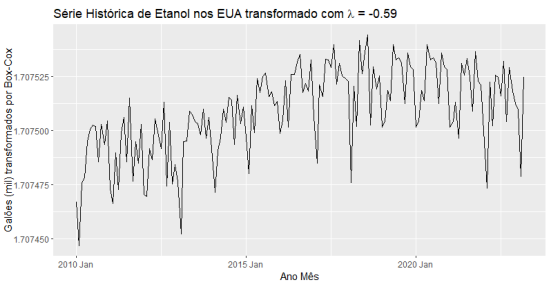


Figura 3: Série histórica de consumo de etanol com redução de variância via aplicação do método Box-Cox.

De posse da série exposta na Figura 3, foi realizada a eliminação das componentes de tendência e sazonalidade via decomposição STL para que o *Moving Block Bootstrap* possa ser realizado somente nos resíduos (Figura 4).

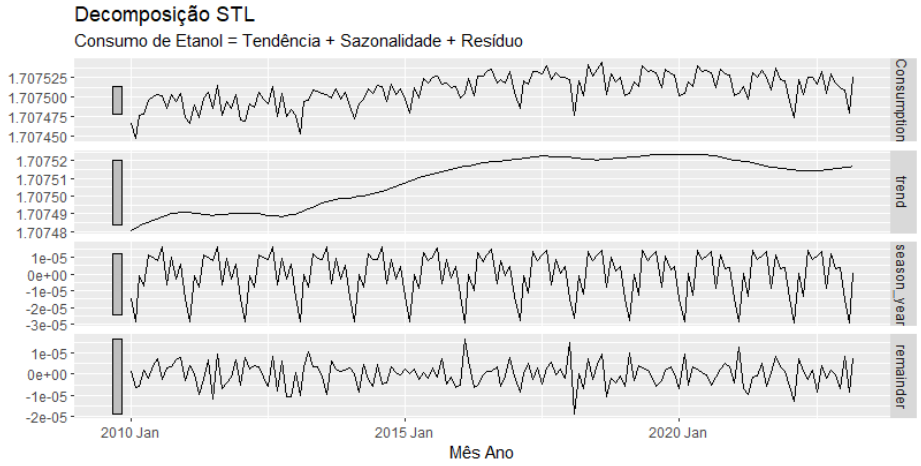


Figura 4: Componentes obtidas após Decomposição STL nos dados históricos com variância reduzida

Com a aplicação do MBB com blocos de tamanho 8, foram gerados 100 cenários da série histórica já tratada. A Figura 5 mostra os cenários já tendo sido reincorporadas as componentes de tendência e sazonalidade, assim como aplicada a função inversa do método Box-Cox.

DOI: 10.59254/sbpo-2025-212434
https://proceedings.science/p/212434?lang=pt-br

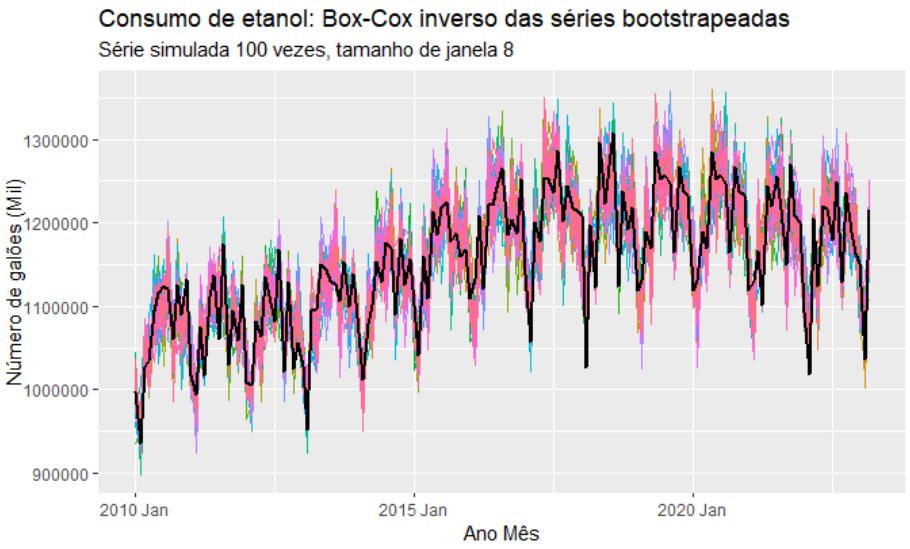


Figura 5: Série histórica de consumo de etanol com 100 cenários gerados via *Moving Block Bootstrap*

A validação dos cenários gerados por *bootstrap* foi realizada avaliando as medidas de tendência central das séries simuladas e verificando se os padrões de dependência temporal foram preservados em relação à série original. O erro médio das 100 séries simuladas em comparação com a série original foi calculado e apresentado no histograma da Figura 6, evidenciando a tendência central dessa medida. De forma análoga, também foram realizadas análises com os valores de Dynamic Time Warping (DTW), conforme ilustrado na Figura 7.

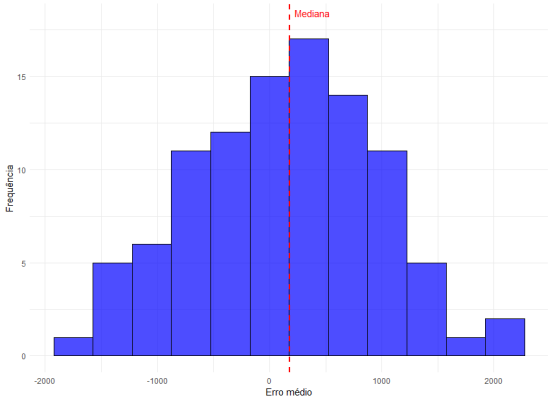


Figura 6: Histograma com a distribuição dos erros médios das séries *bootstrappeadas*.

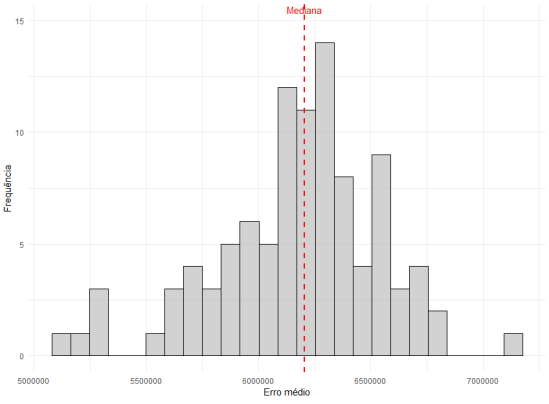


Figura 7: Histograma com a distribuição dos valores de DTW das séries *bootstrappeadas*.

Para avaliar a preservação dos padrões sazonais, utilizamos a Função de Autocorrelação (ACF). Comparamos os valores dessas funções nos cenários gerados com os valores obtidos na série original. Os resultados, apresentados na Figura 8, mostram que as séries geradas reproduzem de forma consistente os padrões de dependência entre os *lags* observados na série original, confirmando a manutenção das características sazonais.

DOI: 10.59254/sbpo-2025-212434
https://proceedings.science/p/212434?lang=pt-br

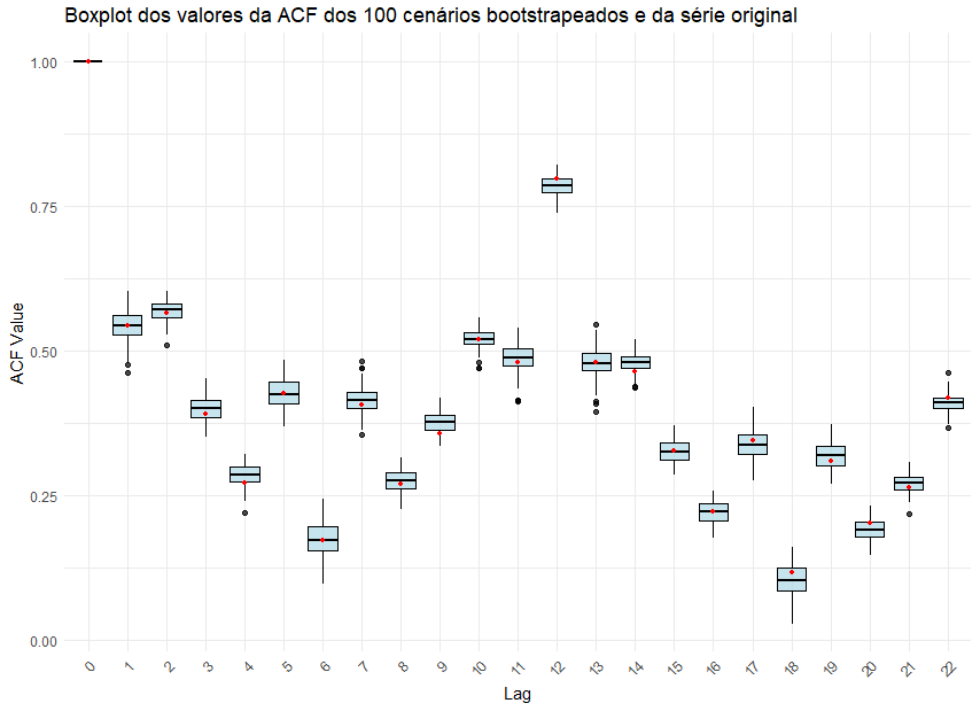


Figura 8: Avaliação dos valores da Função de Autocorrelação (ACF) obtidos a partir dos 100 cenários gerados. Os pontos vermelhos indicam os valores correspondentes à série original.

Os resultados empíricos indicam que as abordagens por *bootstrap* são ferramentas eficientes para a geração de cenários futuros da demanda de etanol nos Estados Unidos. Conforme evidenciado pela Tabela 2, todas as métricas de acurácia mostram que o ARIMA aplicado a séries *bootstrappeadas* é uma abordagem promissora para prever esta série temporal. Os valores futuros gerados por este modelo, considerando tanto a média quanto a mediana das previsões, demonstraram-se consistentemente próximos aos melhores resultados em todas as métricas consideradas, reforçando sua confiabilidade.

Método de Previsão	**Estatística**	**MAPE (%)**	**RMSE (Mil Galões)**	**DTW**
MBB-ARIMA	Média	1,5503	21 664	356 691
	Mediana	1,5491	21 699	**356 578**
MBB-ETS	Média	1,6468	22 977	366 050
	Mediana	1,7147	23 709	383 036
MBB-SNaive	Média	1,7692	23 878	383 025
	Mediana	1,8465	24 589	380 523
Trad. ARIMA	Média	**1,5323**	**20 952**	370 150
Trad. ETS	Média	1,6362	22 198	370 150
Trad. SNaive	Média	2,2197	32 095	510 468

Tabela 1: Resultados dos métodos de previsão com diferentes estatísticas e métricas

DOI: 10.59254/sbpo-2025-212434
https://proceedings.science/p/212434?lang=pt-br

Método de Previsão	Estatística	MAPE (%)	RMSE (Mil Galões)	DTW
MBB-ARIMA	Média	1,5503	21 664	351 009
	Mediana	1,5491	21 699	**352 578**
MBB-ETS	Média	1,6468	22 977	384 153
	Mediana	1,7147	23 709	379 607
MBB-SNaive	Média	1,7692	23 878	383 025
	Mediana	1,8465	24 589	380 523
Trad. ARIMA	Média	**1,5323**	**20 952**	370 150
Trad. ETS	Média	1,6362	22 198	370 150
Trad. SNaive	Média	2,2197	32 095	510 468

Tabela 2: Resultados dos métodos de previsão com diferentes estatísticas e métricas

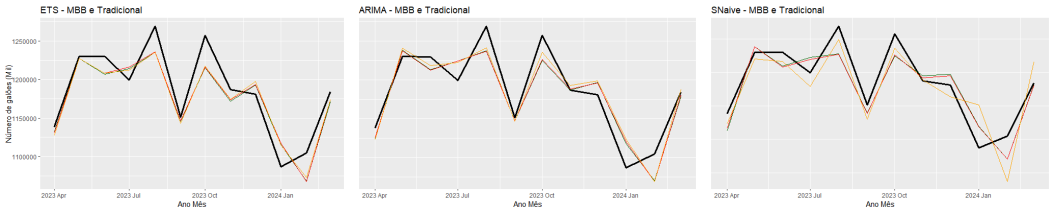


Figura 9: Gráficos com as previsões obtidas por diferentes métodos. A linha preta indica a série real, enquanto as linhas vermelha, verde e amarela representam as previsões geradas por distintas abordagens.

No entanto, ao analisar as métricas baseadas em comparação pontual, como o RMSE e o MAPE, o desempenho do ARIMA via *bagging* foi ligeiramente inferior ao dos métodos tradicionais (sem *bagging*). No caso específico do MAPE, a diferença foi mínima, com apenas $0{,}90\%$ em relação ao melhor resultado. Esses valores indicam que, em termos de precisão pontual, os métodos tradicionais ainda apresentam uma ligeira vantagem em relação ao *bagging*.

Por outro lado, ao avaliarmos a métrica DTW, que considera alinhamentos mais complexos entre as séries temporais, os métodos com *bagging* se destacaram. Em particular, o ARIMA aplicado às séries *bootstrappeadas* apresentou um desempenho superior, sugerindo que esta abordagem capturou melhor as dinâmicas da série original. Essa capacidade de modelar padrões temporais mais amplos reflete o potencial do *bagging* para gerar previsões que vão além da precisão pontual, oferecendo *insights* importantes sobre a estrutura da série.

5. Considerações finais

Este trabalho apresentou uma metodologia para geração de cenários e previsão da demanda de etanol, tomando como plano de fundo a análise da evolução do consumo recente deste biocombustível nos Estados Unidos. Incluímos nesse processo a realização de tratamentos na estrutura da série temporal para suavizar as disrupções causadas pela pandemia de Covid-19, utilização de técnicas de *bootstrap* e *bagging* para gerar cenários, além da avaliação de diferentes tipos de modelos de previsão no período de um ano.

A qualidade dos cenários gerados foi validada por meio da preservação das principais características da série original, como padrões de sazonalidade e dependências temporais, evidenciados pela análise da Função de Autocorrelação (ACF). Além disso, a variabilidade das séries simuladas foi bem controlada, permitindo a criação de séries alternativas que capturaram de forma adequada as incertezas inerentes ao consumo de etanol. Essa abordagem demonstrou-se eficaz para fornecer informações mais completas sobre o comportamento do mercado, indo além de previsões pontuais e ampliando o escopo da análise para cenários prospectivos.

DOI: 10.59254/sbpo-2025-212434
https://proceedings.science/p/212434?lang=pt-br

Com relação aos métodos de previsão, os resultados destacaram a superioridade do modelo ARIMA na preservação de dinâmicas complexas da série temporal. Apesar de apresentar *scores* ligeiramente piores nas métricas tradicionais, o uso de *bagging* mostrou-se uma estratégia poderosa para capturar padrões estruturais da série, como indicado a partir da métrica DTW.

Como sugestões de prosseguimento desta pesquisa, a robustez das conclusões poderá ser ampliada ao testar a metodologia em outros mercados de etanol, avaliando sua generalização e aplicabilidade em contextos distintos. Além disso, abordagens baseadas em outros tipos de modelos preditivos, como as baseadas aprendizado de máquina, redes neurais ou métodos híbridos podem complementar as estratégias atuais e aumentar a acurácia das previsões, sendo uma possibilidade a ser explorada. Por fim, pesquisas adicionais sobre o uso do *bootstrap* a partir de diferentes formas de pré-processamento dos dados podem refinar a geração de cenários. Com isso, buscamos compreender ainda mais as dinâmicas e variações da demanda por biocombustíveis, fortalecendo o uso dessas ferramentas no planejamento estratégico do setor de energias renováveis.

Referências

Alizadeh, R., Lund, P. D., e Soltanisehat, L. (2020). Outlook on biofuels in future studies: A systematic literature review. *Renewable and Sustainable Energy Reviews*, 134:110326.

Baratsas, S., Iseri, F., e Pistikopoulos, E. N. (2024). A hybrid statistical and machine learning based forecasting framework for the energy sector. *Computers & Chemical Engineering*, 188:108740. ISSN 0098-1354. URL https://www.sciencedirect.com/science/article/pii/S0098135424001583.

Bickel, P. J. e Doksum, K. A. (1981). An analysis of transformations revisited. *Journal of the American Statistical Association*, 76(374):296–311.

Box, G. E. P. e Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 26(2):211–252.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., e Terpenning, I. (1990). STL: A seasonal-trend decomposition. *Journal of Official Statistics*, 6(1):3–73.

Dantas, T. M., Oliveira, F. L. C., e Repolho, H. M. V. (2017). Air transportation demand forecast through Bagging Holt Winters methods. *Journal of Air Transport Management*, 59:116–123.

Dey, B., Roy, B., Datta, S., e Ustun, T. S. (2023). Forecasting ethanol demand in India to meet future blending targets: A comparison of ARIMA and various regression models. *Energy Reports*, 9: 411–418.

EISA (2007). Energy Independence and Security Act of 2007. Public Law 110–140, USA. URL https://www.govinfo.gov/content/pkg/PLAW-110publ140/pdf/PLAW-110publ140.pdf. Accessed: 2024-04-28.

Gironès, V. C., Moret, S., Peduzzi, E., Nasato, M., e MarÃ©chal, F. (2017). Optimal use of biomass in large-scale energy systems: Insights for energy policy. *Energy*, 137:789–797.

Guerrero, V. M. (1993). Time-series analysis supported by power transformations. *Journal of Forecasting*, 12(1):37–48.

DOI: 10.59254/sbpo-2025-212434
https://proceedings.science/p/212434?lang=pt-br

Hyndman, R. J., Koehler, A. B., Ord, J. K., e Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach*. Springer-Verlag. URL http://www.exponentialsmoothing.net.

Javed, S. A., Zhu, B., e Liu, S. (2020). Forecast of biofuel production and consumption in top CO2 emitting countries using a novel grey model. *Journal of Cleaner Production*, 276:123997.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, p. 1217–1241.

Luchansky, M. S. e Monks, J. (2009). Supply and demand elasticities in the U.S. ethanol fuel market. *Energy Economics*, 31(3):403–410. ISSN 0140-9883. URL https://www.sciencedirect.com/science/article/pii/S0140988308001904.

Meira, E., Oliveira, F. L. C., e Menezes, L. M. d. (2022). Forecasting natural gas consumption using bagging and modified regularization techniques. *Energy Economics*, 106:105760.

Mizik, T., Nagy, L., Gabnai, Z., e Bai, A. (2020). The Major Driving Forces of the EU and US Ethanol Markets with Special Attention Paid to the COVID-19 Pandemic. *Energies*, 13(21): 5614. URL https://doi.org/10.3390/en13215614.

Petropoulos, F., Hyndman, R. J., e Bergmeir, C. (2018). Exploring the sources of uncertainty: why does bagging for time series forecasting work? *European Journal of Operational Research*, 268 (3):960–971.

Ramsey, S., Williams, B., Jarrell, P., e Hubbs, T. (2023). Global Demand for Fuel Ethanol Through 2030. *Amber Waves: The Economics of Food, Farming, Natural Resources, and Rural America*, 2023(Bioenergy Number (BIO-05)).

U.S. Department of Energy (2024). Ethanol Fuel Basics. https://afdc.energy.gov/fuels/ethanol. Accessed: 2025-04-28.

USDA (2024). U.S. Bioenergy Statistics. URL https://www.ers.usda.gov/data-products/u-s-bioenergy-statistics/. Accessed: 2024-12-03.

DOI: 10.59254/sbpo-2025-212434
https://proceedings.science/p/212434?lang=pt-br