

# Modelo SARIMA na Análise Temporal da Produção de Etanol em São Paulo

Isabela M. de Oliveira<sup>1</sup>

FEC/UNESP, Rosana, SP

Marilaine Colnago<sup>2</sup>

IBILCE/UNESP, São José do Rio Preto, SP

Lucas A. B. de Matos<sup>3</sup>

IQ/UNESP, Araraquara, SP

São Paulo é líder na produção nacional de etanol, com 171 usinas distribuídas em mais de 60% de seus municípios. Representando cerca de 44,66% da produção total de etanol na safra 2022/2023, a localização das usinas é amplamente dispersa, favorecida por uma combinação de fatores naturais, ambientais, de infraestrutura e políticas governamentais. Com mais de 50% da safra nacional de cana cultivada no estado, São Paulo contribui com 25,3% da oferta energética total, sendo o etanol responsável por aproximadamente metade da energia consumida por veículos leves [1].

Dada a relevância destacada, este estudo propõe o uso de métodos temporais para prever a produção futura de etanol em São Paulo. Essa abordagem visa fornecer *insights* decisivos para auxiliar na gestão estratégica e na tomada de decisões.

O ARIMA (do inglês *AutoRegressive Integrated Moving Average*) é um modelo estatístico amplamente utilizado na análise de séries temporais para compreender a estrutura dos dados e identificar padrões e tendências, que permite que valores futuros de uma série sejam previstos considerando apenas seus valores passados e presentes. Se no modelo são considerados valores sazonais, temos um modelo SARIMA (do inglês *Seasonal ARIMA*).

Segundo [2], um modelo SARIMA( $p, d, q$ )( $P, D, Q$ )<sub>s</sub> pode ser representado da seguinte forma:

$$\phi(L)\Phi(L)\Delta^d\Delta^D y_t = \theta(L)\Theta(L)\epsilon_t \quad (1)$$

onde  $p$  e  $q$  são, respectivamente, as ordens do polinômio autorregressivo e de médias móveis não sazonais  $\phi(L)$  e  $\theta(L)$ ,  $P$  e  $Q$  são, respectivamente, as ordens do polinômio autorregressivo e de médias móveis sazonal  $\Phi(L)$  e  $\Theta(L)$ , e  $d$  e  $D$  são, respectivamente as ordens de diferença não sazonal e sazonal.

Para analisar a série temporal, foram utilizados os dados mensais de produção de etanol de janeiro de 2012 a junho de 2023, com aproximadamente 80% dos dados de treinamento (01/2012 a 02/2021) e os 20% restantes para teste (03/2021 a 06/2023). Para compreender melhor o padrão da série temporal, realizou-se a decomposição nos componentes de tendência, sazonalidade e residual, revelando uma tendência de crescimento e decrescimento periódica, além de sazonalidade nos dados. Todos os testes computacionais foram realizados a partir da linguagem Python, utilizando o *statsmodels*, que fornece classes e funções para a estimativa de diversos modelos estatísticos, incluindo o SARIMAX.

O modelo compilado foi o SARIMA(1, 1, 1)(0, 1, 2)<sub>12</sub>, determinado a partir da função *auto\_arima*, da biblioteca *pmdarima*, que busca o melhor conjunto de parâmetros, e o tempo de processamento foi de 13.68 segundos.

<sup>1</sup>isabela.mendes-oliveira@unesp.br

<sup>2</sup>marilaine.colnago@unesp.br

<sup>3</sup>lucas.b.matos@unesp.br

Para validar os resultados, foram calculadas as seguintes métricas estatísticas: MAE, MAPE e  $R^2$  [3]. Os resultados quantitativos indicaram que o modelo teve um bom desempenho:  $MAE = 223332.17$ ,  $R^2 = 0.8707$  e  $MAPE = 4.87\%$ . Porém, a partir da Figura 1 pode-se observar que o método previu valores negativos, o que para o contexto apresentado não seria satisfatório.

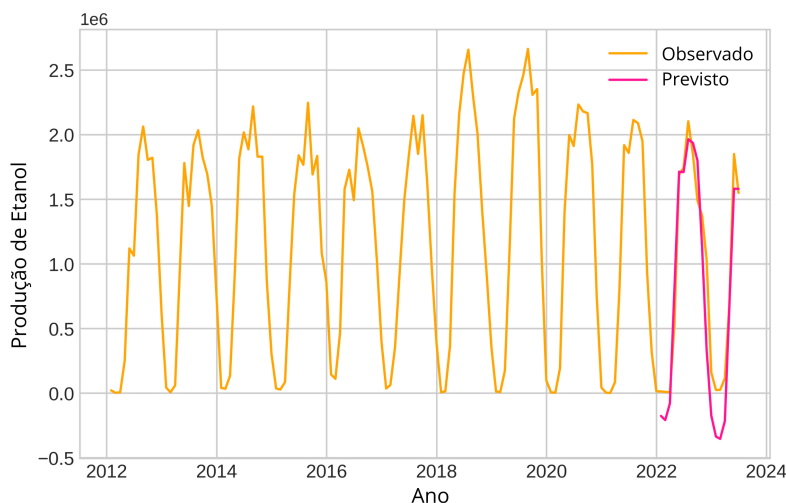


Figura 1: Gráfico de previsão de produção de etanol realizada pelo modelo SARIMA. Fonte: dos autores.

Existem algumas abordagens que podem ser empregadas para corrigir os valores negativos, sendo uma delas o ajuste dos valores para zero, o que se mostra pertinente no contexto deste trabalho, uma vez que pode-se observar pela Figura 1, que os valores negativos estão concentrados onde deveriam ser próximos de zero. Após este ajuste, obteve-se novos valores para as métricas, sendo:  $MAE = 137588.03$ ,  $R^2 = 0.92$  e  $MAPE = 0.49\%$ .

Os resultados evidenciam que o modelo preditivo SARIMA foi eficaz na estimativa da produção de etanol no estado de São Paulo, principalmente após a etapa de pós-processamento dos dados. Esse desempenho evidencia o potencial do modelo para orientar decisões estratégicas de agências governamentais, atender às demandas dos consumidores locais e fornecer *insights* valiosos para setores específicos, como a indústria de energia agrícola ou o mercado de subprodutos da cana-de-açúcar.

## Agradecimentos

Agradecemos à FAPESP pelo fomento à pesquisa (Processo: 2022/15191-5).

## Referências

- [1] UNICA - União da Agroindústria Canavieira do Estado de São Paulo. **Setor Sucreenergético**. Online. Disponível em: <https://www.unica.com.br>. Acessado em: 22/03/2024. 2023.
- [2] A. C. Barros et al. **Análise de Séries Temporais em R: Curso Introductório**. Grupo GEN, 2017. ISBN: 978-8-535-29087-5.
- [3] M. Paula et al. "Predicting Energy Generation in Large Wind Farms: A Data-Driven Study with Open Data and Machine Learning". Em: **Inventions** 8.5 (2023), p. 126.