

## Análise da Evolução da Concentração Atmosférica de CO<sub>2</sub> utilizando Média Móvel Integrada Auto-Regressiva

Carolina C. Simões<sup>1</sup>, Henrique A. M. Faria<sup>2</sup>, Marisa V. Capela<sup>3</sup>, Jorge M.V. Capela<sup>4</sup>  
 DEFM/IQ-UNESP, Araraquara, SP

Os métodos de previsão em séries temporais fundamentam-se na premissa de que há um padrão ou estrutura na série histórica que pode ser identificado e projetado para o futuro. Neste estudo, adotaram-se modelos de média móvel integrada auto-regressiva, ARIMA (*Autoregressive Integrated Moving Average*), na análise da série temporal da concentração global de CO<sub>2</sub>, conforme disponibilizada pelo Global Monitoring Laboratory [5]. Os dados compreendendo o período de janeiro de 1979 a dezembro de 2019 foram designados como conjunto de treinamento, enquanto os dados a partir de 2020 foram reservados para o conjunto de teste. O conjunto de treinamento expõe o modelo a dados históricos, permitindo ajustes para capturar características relevantes da série temporal. Após um treinamento bem-sucedido, o modelo deve generalizar padrões para prever o comportamento futuro da série. Posteriormente, o conjunto de teste é utilizado para avaliar o desempenho do modelo [2]. Além dos componentes autoregressivos, de média móvel e de diferenciação, o modelo ARIMA pode incorporar componentes sazonais para o caso de séries temporais que exibem padrões sazonais e tendências [1, 4].

Utilizou-se a função *auto.arima()* do R, disponível no pacote *forecast*, para automatizar o processo de seleção do modelo mais apropriado para a série temporal em questão. Essa função produz um modelo na notação ARIMA(p,d,q)(P,D,Q), em que, para os dados não sazonais, “p” representa o número de termos autoregressivos (AR), “d” indica o número de diferenciações necessárias para tornar a série estacionária, e “q” denota o número de termos de média móvel (MA). As letras “P”, “D” e “Q” têm significados semelhantes, mas aplicados aos dados sazonais [3].

A acurácia dos modelos foi avaliada ajustando o modelo ao conjunto de treinamento e avaliando o seu desempenho no conjunto de teste. As medidas de erro são métricas que medem o quão distantes os valores previstos estão dos valores reais. Se a parte de treino da série temporal for  $Z_1, Z_2, \dots, Z_M$  e a parte de teste  $Z_{M+1}, Z_{M+2}, \dots, Z_N$ , então tem-se a raiz quadrada média da soma dos quadrados dos erros, RMSE (*Root Mean Square Error*) e o erro percentual médio absoluto, MAPE (*Mean Absolute Percentage Error*), definidos respectivamente por

$$\text{RMSE} = \sqrt{\frac{1}{N - M} \sum_{i=M+1}^N (Z_t - \hat{Z}_t)^2} \quad (1)$$

e

$$\text{MAPE} = \frac{1}{N - M} \sum_{i=M+1}^N \left| \frac{Z_t - \hat{Z}_t}{Z_i} \right|, \quad (2)$$

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

<sup>1</sup>cc.simoes@unesp.br

<sup>2</sup>henrique.faria@unesp.br

<sup>3</sup>marisa.capela@unesp.br

<sup>4</sup>jorge.capela@unesp.br

Os resultados revelaram que o  $ARIMA(0,1,3)(0,1,1)$  foi o modelo mais bem ajustado, caracterizado pela ausência de termos autoregressivos sazonais e não sazonais, uma diferenciação não sazonal e uma diferenciação sazonal, três termos de média móvel não sazonais e um termo de média móvel sazonal. O valor do RMSE foi de 0,126, enquanto o MAPE foi de 0,025%. A Figura 1 ilustra o ajuste do modelo à série temporal, onde a linha sólida em preto representa o ajuste global, a linha vermelha indica o ajuste ao conjunto de treino, e a linha azul representa o ajuste ao conjunto de teste.

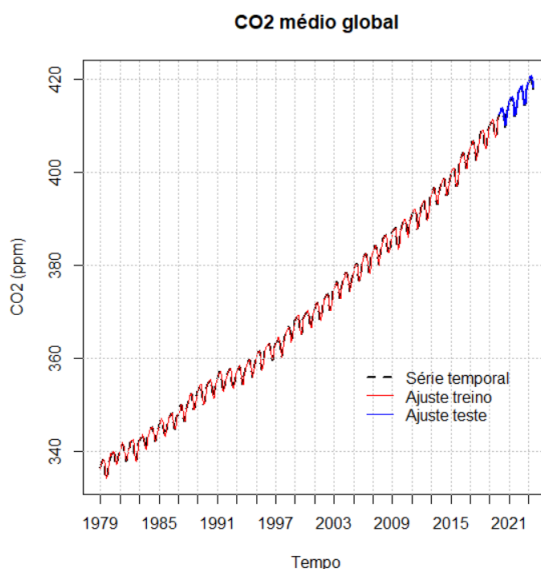


Figura 1: Ajuste do modelo  $ARIMA(0,1,3)(0,1,1)$  à série temporal do da concentração global de  $CO_2$ . Fonte: Autores.

Conclui-se que o modelo  $ARIMA(0,1,3)(0,1,1)$  demonstrou ser a escolha mais adequada, capaz de ajustar tanto os padrões não sazonais quanto os sazonais presentes na série temporal. A avaliação de seu desempenho confirmou uma previsão precisa e confiável, evidenciando a utilidade do modelo para análises futuras e aplicações práticas na compreensão e previsão dos padrões na série temporal da concentração atmosférica de  $CO_2$ .

## Referências

- [1] K.S. Chan e J.D. Cryer. **Time series analysis with applications in R**. New York: Springer, 2008. ISBN: 9780387759593.
- [2] R.J. Hyndman. “Measuring forecast accuracy.” Em: **Business forecasting: Practical problems and solutions**. Ed. por M. Gilliland, L. Tashman e U. Sglavo. John Wiley & Sons, 2015. Cap. 3, pp. 177–183.
- [3] R.J. Hyndman e Y. Khandakar. “Automatic time series forecasting: the forecast package for R”. Em: **Journal of statistical software** 27 (2008), pp. 1–22. DOI: 10.18637/jss.v027.i03.
- [4] P. A. Morettin e C. M.C. Toloi. **Análise de séries temporais: modelos lineares univariados**. Editora Blucher, 2018. ISBN: 9786555060041.
- [5] N. Global Monitoring Laboratory US DEPARTMENT OF COMMERCE. **Carbon Cycle Greenhouse Gases**. Acessado em 07/03/2024, <https://gml.noaa.gov/ccgg/trends/mlo.html>.