

# Previsão da Emissão de CO<sub>2</sub> e Gases do Efeito Estufa em Lavouras de Arroz no Estado do Rio Grande do Sul

Dalvana L. Ribeiro<sup>1</sup>, Viviane L. D. de Mattos<sup>2</sup>, André A. Longaray<sup>3</sup>

Programa de Pós-Graduação em Modelagem Computacional (PPGMC)/FURG, Rio Grande, RS

**Resumo.** Nos últimos anos, os grandes volumes de emissões de gases do efeito estufa e CO<sub>2</sub> tem preocupado e gerado discussões em todo mundo. A elevação da temperatura e os desastres ambientais, cada vez mais constantes, evidenciam a necessidade da abordagem desse tema. Nesse sentido, o objetivo do estudo foi realizar a previsão da emissão de CO<sub>2</sub> e gases do efeito estufa em lavouras de arroz no estado do Rio Grande do Sul. Os dados utilizados foram provenientes do Sistema de Estimativas de Emissões e Remoções de Gases de Efeito Estufa (SEEG). Para as análises foi utilizada uma série temporal anual no período de 1970 a 2021, e calculadas as previsões por meio de métodos básicos (média, naive e drift) e o modelo mais complexo ARIMA. A acurácia dos modelos foi avaliada por meio de valores RMSE, MAPE E MAE. Entre os modelos analisados, o modelo ARIMA obteve resultados satisfatórios ao comparar com os valores reais da série estudada.

**Palavras-chave.** Séries temporais, Gases do Efeito Estufa, ARIMA, Previsão.

## 1 Introdução

A intensidade das alterações e variações climáticas causadas pelo efeito estufa é apontada como uma ameaça às condições de vida da população mundial [7]. Entre os principais gases responsáveis pelo efeito estufa (GEE) estão: dióxido de carbono (CO<sub>2</sub>), clorofluorcarbonos (CFCs), Dióxido nitroso (NO<sub>2</sub>) e o metano (CH<sub>4</sub>), produzido em solos naturalmente inundados e em lavouras de arroz. De acordo com [10], economias emergentes, especialmente o Brasil, China e Índia, estão se tornando os principais emissores de GEE à medida que aumentam sua produção de alimentos. Nesse contexto, o Rio Grande do Sul é um dos maiores produtores de arroz do Brasil, com uma produção anual média de cerca de 8.209.443 toneladas do grão[4]. Contudo, a expansão da agricultura tem influenciado também no aumento das emissões de gases de efeito estufa, contaminação e degradação dos recursos hídricos e do solo. Nessa perspectiva, o estudo de séries temporais de emissão desses gases pode contribuir para determinação de medidas controladoras e eficazes para a atenuação desse fenômeno. A análise de séries temporais é um importante instrumento na compreensão de fenômenos. Sendo assim, este trabalho tem como objetivo realizar previsões de emissões de GEE em lavouras de arroz no estado do Rio Grande do Sul. Para tanto, foram abordados métodos básicos de previsão como média, naive e drift e o modelo ARIMA. A organização deste trabalho está estruturada nos seguintes tópicos: embasamento teórico onde são abordados alguns conceitos teóricos a cerca dos modelos utilizados, seguido pela metodologia utilizada na análise da série escolhida, na seção 4 os resultados obtidos são mostrados e discutidos e por fim, na seção 5 uma conclusão do trabalho é apresentada.

---

<sup>1</sup>lopesribeirodalvana20@gmail.com

<sup>2</sup>vivianemattos@furg.br

<sup>3</sup>andrelongaray@furg.br

## 2 Embasamento Teórico

Uma série temporal consiste em um conjunto de dados observados ao longo de um período de tempo [6] e pode ser constituída de diversas formas como dados mensais, diários, anuais etc. Para seu estudo, inúmeras técnicas de modelagem matemática baseadas em estatísticas podem ser usadas.

A Figura 1 mostra algumas categorias de modelos.

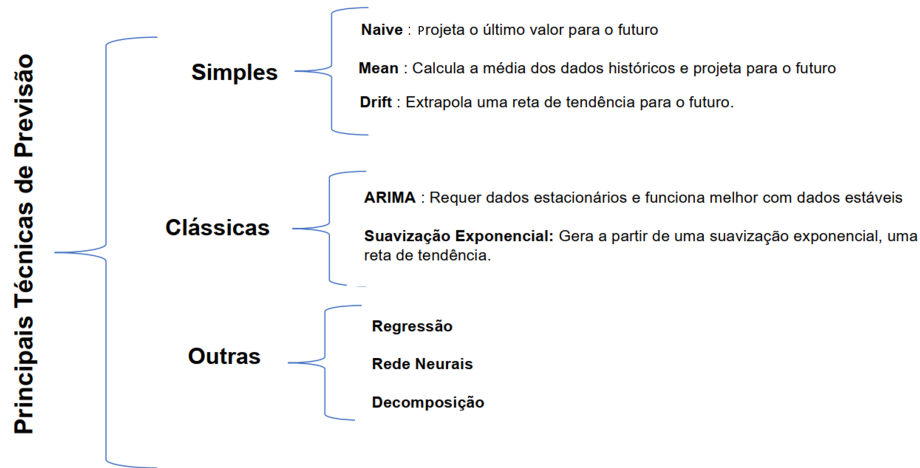


Figura 1: Principais técnicas de previsão. Fonte: Adaptado de [3]

### 2.1 Modelagem ARIMA

O modelo Auto Regressivo Integrado de Média Móvel (ARIMA) é uma abordagem de previsão de série temporal usada para prever o valor futuro de uma variável a partir de seus próprios valores passados. Consiste na combinação de três componentes: Auto-regressivo (AR), integração (I) e o componente de médias móveis (MA) e sua notação é representada como ARIMA (p,d,q) onde p e q representam as suas ordens. O modelo é expresso conforme a Equação 1:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \phi_1 \epsilon'_{t-1} + \dots + \phi_q \epsilon_{t-q} + \epsilon_t \quad (1)$$

onde  $y'_t$  representa a série diferenciada e é calculado levando em consideração os valores defasados de  $y_t$  bem como os erros defasados,  $\phi_p$  é o coeficiente da parcela auto-regressiva na p-ésima defasagem,  $\phi_q$  é o coeficiente de médias móveis na q-ésima defasagem e  $\epsilon_t$  o erro no t-ésimo tempo.

A determinação dos valores p,d,q requerem testes e otimização. [5], sugerem a seguinte sequência para a construção do modelo ARIMA: primeiro verifica-se se a série é estacionaria, caso não seja realiza-se diferenciações torná-la estacionária. Já para escolher os valores p e q, deve-se inspecionar visualmente os gráficos da função de autocorrelação (ACF) e função de autocorrelação parcial (PACF), que provaram ser úteis na determinação dos valores dos parâmetros p e q [2] Estimados os parâmetros os critérios de informação de Akaike (AIC), de Akaike Corrigido (AICc) e o Critério Bayesano de Informação de Schwarz (BIC) podem ser usados para elencar o modelo apropriado. Para medir a precisão de previsão, o Erro Quadrado Médio Médio (RMSE), o Erro Médio Absoluto (MAE) e o Erro Percentual Absoluto Médio (MAPE), são os mais utilizados. Estes são definidos nas Equações 2, 3 e 4.

$$RMSE = \sqrt{\frac{1}{n-m} \sum_{t=m+1}^n (y_t - \hat{y}_t)^2} \quad (2)$$

$$MAPE = \frac{1}{n-m} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \quad (3)$$

$$MAE = \frac{1}{n-m} \sum_{t=m+1}^n |y_t - \hat{y}_t| \quad (4)$$

Assumindo por exemplo, que o conjunto de dados observado é de tamanho  $n$  e o conjunto de teste é de tamanho  $m$ , onde  $m < n$ . Idealmente,  $m$  deve ser tão grande quanto o período de tempo previsto. Um erro de previsão é a diferença entre o valor observado ( $y_t$ ) e a sua previsão  $\hat{y}_t$  [3].

### 3 Metodologia

Os dados históricos da emissão anual de CO<sub>2</sub> de lavouras de arroz, em toneladas ( $t$ ), do estado do Rio Grande do Sul foram obtidos através do Sistema de Estimativas de Emissões e Remoções de Gases de Efeito Estufa (SEEG | <http://seeg.eco.br/>). O SEEG é uma iniciativa do Observatório do Clima que compreende a produção de estimativas anuais das emissões de gases de efeito estufa (GEE) no Brasil por meio de documentos analíticos. Inicialmente foi realizado um estudo exploratório da série para determinar algumas características para a construção dos modelos. Em seguida a modelagem foi realizada com o uso dos métodos de média, naive e drift. Em todos foram realizadas análise de resíduos e teste de Ljung-Box para a verificação das propriedades de autocorrelação. Partindo para os métodos clássicos, deu-se início na modelagem ARIMA. A ordem de diferenciação ( $d$ ) foi obtida por meio de análise de estacionariedade do comportamento da série. Para tanto, foi aplicado o teste Dickey Fuller Aumentado – ADF [1], considerando o critério de [8] para encontrar o número máximo de defasagens a ser considerado. Feito isso, a análise gráfica da função de autocorrelação (FAC) e da função de Autocorrelação Parcial (FACP) foram conduzidas a fim de determinar as componentes  $p$  e  $q$  do modelo. Com o teste de estacionariedade e as análises das funções de autocorrelação foram considerados seis modelos candidatos. Estes modelos foram então comparados entre si, utilizando os critérios de informação de Akaike (AIC), Akaike Corrigido (AICc) e o Critério Bayesano de Informação de Schwarz (BIC) para elencar o modelo mais propício a ser usado. Os testes para verificar homoscedasticidade (ARCH), normalidade (Jarque-Bera) e autocorrelação (Ljung-Box) nos resíduos foram aplicados, sendo estes indispensáveis na modelagem ARIMA. A capacidade preditiva foi medida por RMSE, MAE e MAPE. E que foram construídos intervalos com 95% de confiança.

Também é usual dizer que as análises foram desenvolvidas no RStudio [9].

### 4 Resultados e Discussão

Para obter a previsão da emissão de CO<sub>2</sub> em lavouras de arroz no estado do Rio Grande do Sul, foi utilizada uma série temporal anual no intervalo de 1970-2021 (Figura 2) .

Para a modelagem foram considerados os dados relativos ao período 1970 a 2010, utilizando o intervalo de 2011 a 2021 para a validação. Inicialmente, com a execução de uma análise exploratória pode-se observar que a série não possui sazonalidade, com média de 1332962 ( $t$ ), desvio padrão de 317057.6 ( $t$ ) e variabilidade de 0.2378594. Além disso, a série possui uma distribuição platycúrtica com caudas mais leves do que uma distribuição normal e assimetria com o valor -0.6768629.



Figura 2: Emissão anual de CO<sub>2</sub> e GEE em lavouras de arroz no RS

Após feita a análise exploratória passou-se a modelagem da série com os métodos básicos. Na Tabela 1 é exposto a avaliação dos resíduos destes modelos e o modelo Mean é o que apresenta o pior desempenho em relação aos demais por terem sido encontradas evidências de autocorrelação.

Tabela 1: Avaliação dos resíduos dos modelos obtidos.

Modelo	Ljung Box	
	Estatística	Valor p
Mean	28.668	0,0001
Naive	9.1862	0.3268
Drift	9.1862	0.3268

O passo seguinte foi a construção de um modelo ARIMA. Conforme o teste de estacionariedade pode ser considerado  $d = 1$ , pois a série foi diferenciada uma vez para se tornar estacionaria. A partir disso, passa-se a investigar os possíveis valores de  $p$  e  $q$ , que podem ser determinados pelo gráfico da função de autocorrelação, mostrado na Figura 3.

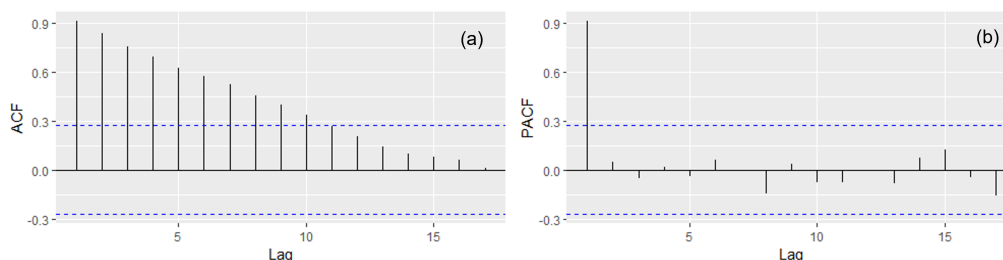


Figura 3: Gráfico da função de autocorrelação (a) e da função de autocorrelação parcial (b)

Observando os gráficos da função de autocorrelação (ACF) e da autocorrelação parcial (PACF) 5 da série, percebe-se as características de um modelo AR(1), pois tem-se uma queda lenta no gráfico da ACF e um queda brusca na defasagem 1 no PACF. De acordo com as observações realizadas, pode-se identificar os seguintes modelos candidatos: ARIMA(0,1,0), ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,1), ARIMA(2,1,1), ARIMA(2,1,0). A determinação de um modelo apropriado que explique satisfatoriamente o comportamento de um fenômeno é uma etapa importante na análise de dados. Nesse sentido, como ferramentas auxiliares neste processo estão os critérios de informação de Akaike (AIC) e o critério Bayesiano de Schwarz (BIC) que podem ser utilizados com a finalidade de ordenação de modelos ARIMA. Na Tabela 2 são mostrados os valores dos respectivos critérios de informação para os modelos candidatos sugeridos.

Tabela 2: Modelos candidatos e critérios de informações.

Modelo	AIC	AICc	BIC
ARIMA (0,1,0)	1033.26	1033.58	1036.64
ARIMA (1,1,0)	1033.70	1034.37	1038.77
ARIMA (0,1,1)	1032.93	1033.59	1037.99
ARIMA (1,1,1)	1033.50	1033.65	1039.26
ARIMA (2,1,1)	1034.34	1036.11	1042.79
ARIMA (2,1,0)	1035.38	1036.52	1042.13

De acordo com a Tabela 2, o modelo ARIMA(0,1,0) foi o modelo mais parcimonioso apresentando os menores valores nos critérios de informações. A normalidade dos modelos foi testada nos resíduos conforme a Tabela 3 em que todos cumprem as suposições de independência, normalidade e homocedasticidade dos resíduos

Tabela 3: Avaliação dos resíduos dos modelos candidatos.

Modelo	Ljung-Box		Jarque-Bera		ARCH-LM	
	Estatística	valor p	Estatística	valor p	Estatística	valor p
ARIMA (0,1,0)	10.357	0.4097	1.6659	0.4348	7.7387	0.6543
ARIMA (1,1,0)	9.6642	0.4704	1.8405	0.3984	8.1498	0.6142
ARIMA (0,1,1)	10.063	0.4350	2.4115	0.2995	8.6679	0.5639
ARIMA (1,1,1)	6.6203	0.7607	2.2591	0.3232	8.6800	0.5627
ARIMA (2,1,1)	6.7080	0.7527	2.6981	0.2595	8.2565	0.6038
ARIMA (2,1,0)	9.3881	0.4957	2.4416	0.2950	8.9553	0.5364

Para estabelecer a capacidade preditiva dos modelos foram adotados os critérios de erros como Erro Quadrático Médio Raiz (RMSE), Erro Médio Absoluto (MAE) e Erro Médio de Percentual Absoluto (MAPE). A Tabela 4 mostra a acurácia dos modelos baseados nos critérios descritos.

Tabela 4: Medidas de acurácia dos modelos candidatos.

Modelo	RMSE	MAE	MAPE
ARIMA (0,1,0)	92227.40	69259.32	3.981124
ARIMA (1,1,0)	81757.35	66785.54	3.898234
ARIMA (0,1,1)	123660.77	97231.72	5.557946
ARIMA (1,1,1)	87434.15	75809.96	4.486392
ARIMA (2,1,1)	89732.90	77082.89	4.553019
ARIMA (2,1,0)	84037.96	70508.47	4.135090

Considerando a análise realizada, o modelo mais adequado por apresentar melhor capacidade preditiva é o ARIMA(1,1,0), pois apresenta os menores valores de RMSE, MAE e MAPE.

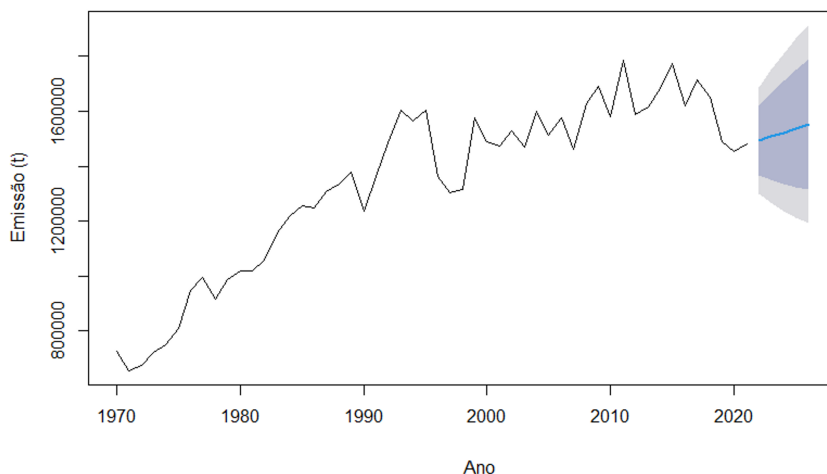


Figura 4: Representação do modelo ARIMA(1,1,0) para a emissão de CO<sub>2</sub> e gases do Efeito Estufa em lavouras de arroz do estado do Rio Grande do Sul no período entre 1970-2021

Na Tabela 5 é possível averiguar a proximidade com os dados reais observados e os previstos com o modelo ARIMA. Percebe-se uma maior diferença entre os valores dos últimos anos, porém todos se encontram entre o intervalo de confiança estipulado para ambos os modelos. Tais valores apontam a boa capacidade preditiva do modelo utilizado.

Tabela 5: Dados observados e dados previstos pelo modelo ARIMA (1,1,0).

Ano	Observado (t)	Previsto (t)	Intervalo de Confiança (0,95)	
			Limite Inferior	Limite Superior
2011	1783878	1630054	1445563	1814545
2012	1588056	1646975	1410824	1883126
2013	1616843	1670405	1388004	1952806
2014	1682515	1692527	1371167	2013886
2015	1772523	1714911	1358701	2071121
2016	1621279	1737243	1349324	2125162
2017	1714489	1759585	1342357	2176814
2018	1652685	1781926	1337316	2226536
2019	1490116	1804266	1333866	2274667
2020	1452317	1826607	1331759	2321455
2021	1479389	1848948	1330804	2367091

## 5 Conclusões

O presente estudo teve como objetivo realizar uma previsão da emissão de CO<sub>2</sub> de lavouras de arroz do estado do Rio Grande do Sul. Para tanto, foi realizada uma análise exploratória que mostrou parâmetros interessantes de caracterização da série estudada e uma visão geral do seu comportamento. Feito isso, foram aplicados os métodos básicos (Média, Naive e Drift) e ARIMA. Os modelos selecionados foram examinados em termos de medidas de parcimonia (AIC, AICc e BIC) e acurácia (RMSE, MAE e MAPE). Além disso, os valores previstos se aproximaram aos valores reais com uma diferença maior entre os últimos anos considerados. Uma observação interessante foi considerar os limites dos intervalos de confiança dos modelos que em sua totalidade continham os valores reais. Vale ressaltar que a previsão de emissão de CO<sub>2</sub> por meio de série de dados não podem evidenciar a situação completa deste fenômeno. Entretanto, a discussão sobre este assunto torna-se relevante diante dos eventos enfrentados atualmente, mostrando que os modelos de previsão podem contribuir para elaboração de políticas públicas que visem minimizar seus efeitos.

## Referências

- [1] D. A. Dickey e W. A. Fuller. “Likelihood ratio statistics for autoregressive time series with a unit root”. Em: **Econometrica: journal of the Econometric Society** (1981), pp. 1057–1072.
- [2] A. Haneen et al. “On the accuracy of ARIMA based prediction of COVID-19 spread”. Em: **Results in Physics** 27 (2021), p. 104509.
- [3] R. J. Hyndman e G. Athanasopoulos. **Forecasting: principles and practice**. OTexts, 2018.
- [4] IBGE. **Site oficial do Instituto Brasileiro de Geografia e Estatística: Projeção da População do Brasil por sexo e idade**. Online. Acessado em 07/11/2022, <http://www.ibge.gov.br>.
- [5] P. A. Morettin e C. M. C. Toloi. **Análise de séries temporais**. 2a. ed. São Paulo: Blucher, 2006. ISBN: 8521203896.
- [6] P. A. Morettin e C. M. C. Toloi. **Análise de séries temporais: modelos lineares univariados**. 3a. ed. São Paulo: Blucher, 2018. ISBN: 9788521213512.
- [7] D. S. Paiva et al. “Mercado voluntário de carbono: análises de cobenefícios de projetos brasileiros”. Em: **Revista de Administração Contemporânea** 19 (2015), pp. 45–64.
- [8] G. W. Schwert. “Why does stock market volatility change over time?” Em: **The journal of finance** 44.5 (1989), pp. 1115–1153.
- [9] R. C. Team et al. “R: A language and environment for statistical computing. R Foundation for Statistical Computing.” Em: (<https://www.R-project.org/>).
- [10] H. Tian et al. “A comprehensive quantification of global nitrous oxide sources and sinks”. Em: **Nature** 586.7828 (2020), pp. 248–256.