

Previsão de produção de cana-de-açúcar e seus derivados, através de ARIMA, no período de 2020 a 2025, na macrorregião Centro-Sul do Brasil

Forecast of the production of sugarcane and its derivatives, through ARIMA, in the period from 2020 to 2025 in the Centro-South macroregion of Brazil

Pronóstico de producción de caña de azúcar y sus derivados, a través de ARIMA, en el período 2020 a 2025, en la macroregión Centro-Sur de Brasil

Recebido: 16/02/2022 | Revisado: 22/02/2022 | Aceito: 01/03/2022 | Publicado: 11/03/2022

Celso Correia de Souza

ORCID: <https://orcid.org/0000-0002-2689-8264>
Pesquisador Produtividade do CNPq, Brasil
E-mail: csouza939@gmail.com

Joao Bosco Arbués Carneiro Junior

ORCID: <https://orcid.org/0000-0001-7094-2245>
Universidade Federal de Rondonópolis, Brasil
E-mail: jbacj@hotmail.com

Márcia Ferreira Cristaldo

ORCID: <https://orcid.org/0000-0001-8790-7163>
Instituto Federal de Mato Grosso do Sul, Brasil
E-mail: marcia.cristaldo@ifms.edu.br

Raul Asseff Castelão

ORCID: <https://orcid.org/0000-0002-4677-6088>
Universidade Anhanguera Uniderp, Brasil
E-mail: raulassefcastelao@gmail.com

Daniel Massen Frainer

ORCID: <https://orcid.org/0000-0003-0813-214X>
Universidade Estadual do Mato Grosso do Sul, Brasil
E-mail: danielfrainer@gmail.com

Hevelyne Henn da Gama Viganó

ORCID: <https://orcid.org/0000-0001-6951-7903>
Instituto Federal de Mato Grosso do Sul, Brasil
E-mail: hevelyne.vigano@ifms.edu.br

Resumo

Os principais subprodutos da cana-de-açúcar são o açúcar, o etanol anidro e o etanol hidratado, muito importantes para a economia brasileira. Fazer as previsões anuais de suas produções são interessantes para as tomadas de decisões do governo brasileiro. Desse modo, o objetivo deste trabalho foi usar recursos de autoregressivo integrated moving average (ARIMA) em séries temporais, para fazer a previsão da produção de cana-de-açúcar e dos seus subprodutos na macrorregião Centro-Sul brasileira, principal produtora de cana, no período de 2020 a 2025. Utilizou-se para isso séries históricas da produção anual de cana-de-açúcar e de seus subprodutos no período de 1980 a 2020, num total de quatro séries históricas, cada uma delas com quarenta observações. Nessas quatro séries foram aplicados testes estatísticos de correlação de Pearson, de autocorrelação e autocorrelação parcial e diferenciação de primeira ordem visando o bom desempenho de previsões utilizando ARIMA. Foi realizado o teste de previsão no horizonte de seis anos posteriores, de 2021 a 2025, para essas quatro séries. Os resultados podem ser considerados satisfatórios, apesar dos erros médios absolutos em percentual (MAPE) serem relativamente altos para as quatro previsões feitas, em média de 13,6%. Como ainda não existem dados de produções de cana, açúcar, etanol anidro e etanol hidratado já consolidados para o ano de 2021, não foi possível realizar comparações entre as produções reais e as previsões realizadas. A importância deste trabalho está no fornecimento de informações necessárias no planejamento agrícola e nas políticas governamentais para o setor do agronegócio.

Palavras-chave: Desenvolvimento regional; Previsão de produção de cana; Previsão de açúcar; Previsão de etanol anidro; Produção de etanol hidratado; Macrorregião Centro-Sul do Brasil.

Abstract

The main by-products of sugarcane are sugar, anhydrous ethanol and hydrated ethanol, which are very important for the Brazilian economy. Making annual forecasts of its productions is interesting for the decision-making of the Brazilian government. Thus, the objective of this work was to use autoregressive integrated moving average (ARIMA) resources in time series, to forecast the production of sugarcane and its by-products in the Brazilian Center-South macro-region, the

main producer of sugarcane, in the period from 2020 to 2025. For this purpose, historical series of the annual production of sugarcane and its by-products from 1980 to 2020 were used, in a total of four historical series, each with forty observations. In these four series, statistical tests of Pearson's correlation, of autocorrelation and partial autocorrelation and of first order differentiating were applied aiming at the good performance of forecasts using ARIMA. The forecast test was carried out in the horizon of six years later, from 2021 to 2025, for these four series. The results can be considered satisfactory, despite the mean absolute errors in percentage (MAPE) being relatively high for the four forecasts made, on average of 13.6%. As there is still no data on production of cane, sugar, anhydrous ethanol and hydrated ethanol already consolidated for the year 2021, it was not possible to make comparisons between the actual productions and the forecasts made. The importance of this work lies in providing the necessary information in agricultural planning and government policies for the agribusiness sector.

Keywords: Regional development; Sugarcane production forecast; Sugar forecast; Anhydrous ethanol forecast; Hydrated ethanol production; Center-South macro-region of Brazil.

Resumen

Los principales subproductos de la caña de azúcar son el azúcar, el etanol anhidro y el etanol hidratado, que son muy importantes para la economía brasileña. Hacer pronósticos anuales de sus producciones es interesante para la toma de decisiones del gobierno brasileño. Así, el objetivo de este trabajo fue utilizar características de media móvil integrada autorregresiva (ARIMA) en series de tiempo, para pronosticar la producción de caña de azúcar y sus derivados en la macrorregión Centro-Sur de Brasil, principal productora de caña de azúcar, en el periodo de 2020 a 2025. Para ello se utilizaron series históricas de la producción anual de caña de azúcar y sus derivados de 1980 a 2020, en un total de cuatro series históricas, cada una con cuarenta observaciones. En estas cuatro series se aplicaron pruebas estadísticas de correlación de Pearson, de autocorrelación y autocorrelación parcial y de diferenciación de primer orden, buscando el buen desempeño de los pronósticos utilizando ARIMA. La prueba de pronóstico se realizó en el horizonte de seis años después, de 2021 a 2025, para estas cuatro series. Los resultados pueden considerarse satisfactorios, a pesar de que los errores absolutos medios en porcentaje (MAPE) son relativamente elevados para las cuatro previsiones realizadas, en promedio del 13,6%. Como aún no hay datos de producción de caña, azúcar, etanol anhidro y etanol hidratado ya consolidados para el año 2021, no fue posible realizar comparaciones entre las producciones reales y las previsiones realizadas. La importancia de este trabajo radica en brindar la información necesaria en la planificación agropecuaria y las políticas de gobierno para el sector de agronegocio.

Palabras clave: Desarrollo regional; Pronóstico de producción de caña de azúcar; Pronóstico de azúcar; Pronóstico de etanol anhidro; Producción de etanol hidratado; Macrorregión Centro-Sur de Brasil.

1. Introdução

A cultura da cana-de-açúcar teve origem no Brasil Martim Affonso de Souza, que iniciou o seu cultivo na Capitania Hereditária de São Vicente, em 1532. Nessa mesma Capitania foi construído o primeiro engenho de açúcar do Brasil (UDOP, 2003). A cana-de-açúcar é uma planta perene, podendo produzir por um período médio de quatro a seis anos, com manejo relativamente fácil. Atualmente, a cana-de-açúcar é plantada em quase todo o Brasil, com destaque para as regiões Centro-Sul e Norte-Nordeste do País. O estado de São Paulo é o maior produtor de cana-de-açúcar e seus derivados, com a maior área plantada da cultivar.

Segundo Leite & Cortez (2020), o etanol anidro e o etanol hidratado são subprodutos do etanol etílico e usados como biocombustíveis veiculares, com o etanol anidro, que é misturado à gasolina, sendo caracterizado pelo teor alcoólico mínimo de 99,3°, determinado pelo Instituto Nacional de Pesos e Medidas (INPM). O etanol hidratado, utilizado como combustível de veículos flex, é uma mistura de etanol etílico e água, com teor alcoólico mínimo de 92,6° (INPM).

Em decorrência da crise do petróleo na década de 1970, houve a iniciativa brasileira da produção de etanol anidro e hidratado para ser usado como combustíveis para veículos, com a finalidade de reduzir a dependência do país do petróleo importado (Carvalho et al., 2013). O governo brasileiro lança o Programa Nacional do Álcool (PROÁLCOOL), em 1975, com o objetivo de aumentar a produção de etanol anidro e hidratado e incentivar a produção de veículos *flex-fuel*, veículos movidos a etanol hidratado ou gasolina, ou ainda, uma mistura dos dois. Em virtude da grande volatilidade do preço do petróleo, que depende da conjuntura mundial de produção, consumo e conflitos entre nações, os preços do petróleo recuaram logo após a criação do PROÁLCOOL, de modo que a importância do programa reduziu significativamente, abortando toda iniciativa de investimentos no PROÁLCOOL. Somente em 1979 esse programa voltou a ter uma razoável importância, com uma nova crise do petróleo, surgindo daí segunda

fase do programa (PROÁLCOOL, 2008). Esse processo de altas e baixas do petróleo tem sido comum nos anos posteriores, até os dias atuais.

Apesar dos altos e baixos do programa PROÁLCOOL, o cultivo da cana-de-açúcar vem contribuindo para a melhoria do cenário social e industrial brasileiro, propiciando o desenvolvimento do agronegócio do país e creditando ao etanol um importante espaço no setor automobilístico nacional pelo fato de ser um combustível ambientalmente correto e plenamente sustentável (Rodrigues, 2010).

Por outro lado, a participação do Brasil no mercado internacional de açúcar é elevada, aproximadamente, 40% das exportações mundiais são brasileiras, além do que, o mercado interno brasileiro, consumidor de açúcar, também é muito relevante. A exportação de etanol produzido no país é quase que irrelevante, pois, tem dificuldades de atender o consumo interno, fazendo com que o país importe etanol produzido a partir do milho dos EUA. Enquanto uma tonelada de cana produz apenas 89,5 litros de etanol, uma tonelada de milho pode produzir até 407 litros de etanol. Por outro lado, no Brasil, a produtividade média de milho por hectare é de aproximadamente 5 toneladas, enquanto a produtividade média de cana é de 70 toneladas. Daí, em um hectare pode-se produzir em média, aproximadamente, 2035 litros de etanol de milho ou 6265 litros de etanol de cana. Além do mais, em termos financeiros, quando se compara o preço de produção do etanol brasileiro, que é de R\$1,16/L, o etanol americano, que é produzido a partir do milho, tem o seu custo a R\$1,85/L, portanto, altamente competitivo o custo do etanol brasileiro (Viegas, 2010).

Conhecer o comportamento comercial dos subprodutos da cana-de-açúcar no mercado interno e externo é vantajoso para o desenvolvimento de projetos de comercialização e tomadas de decisão. Como o Brasil tem uma participação importante no mercado internacional de açúcar, um forte mercado interno desse produto, além de produzir um elevado volume de etanol (anidro e hidratado), é esperada a previsão das produções dessas *commodities* tanto para a exportação quanto para suprir o mercado interno, visto que, esses produtos têm origem na cana-de-açúcar.

Um modelo matemático que possui a capacidade de estabelecer relações entre diversas variáveis é o indicado para resolver o seguinte problema de pesquisa deste estudo: a previsão de produção de um produto ao longo do tempo, entre outros que utilizam a mesma cadeia produtiva, pode facilitar nas tomadas de decisão na condução dessa produção? Existem vários métodos para a previsão econômica por meio de séries temporais, mas o presente trabalho se concentrará apenas no processo autoregressivo integrado de média móvel (ARIMA).

Sendo assim, este artigo teve por objetivo aplicar conceitos de processos autoregressivos integrados de médias móveis (ARIMA) nas previsões das produções de cana-de-açúcar e de seus subprodutos em um horizonte de até cinco anos, a partir de dados de produções anteriores.

2. Metodologia

Em relação aos procedimentos metodológicos, este trabalho classifica-se como pesquisa documental, com a utilização de dados secundários, extraídos de relatórios gerenciais históricos de órgãos públicos e autarquias brasileiras. Quanto aos objetivos, segundo Raupp e Beuren (2004), esta pesquisa se enquadra como descritiva, pois se pretende estudar e relatar o comportamento de uma série temporal econômica utilizando uma técnica estatística.

Este trabalho também pode ser caracterizado como uma pesquisa exploratória-explicativa, que trata de explorar e aprofundar conhecimentos sobre um tema muito relevante ao Brasil, e explicativa, pois ambas possibilitam comprovar, aprofundar, explicar e identificar fatores que explicam a ocorrência de fatores que expliquem a relevância e importância do tema proposto. Inicialmente, foi realizada uma pesquisa bibliográfica sobre conceitos de séries temporais, dando ênfase às propriedades das séries temporais estacionárias e não-estacionárias. Abordou-se os aspectos conceituais do Modelo Auto-Regressivo Integrado de Médias Móveis (ARIMA) e os métodos de avaliação de desempenho de previsões.

Segundo Morettin e Toloi (2006) os operadores utilizados no modelo ARIMA são dados pelas equações 1 a 4.

$$BZ_t = Z_{t-1}, B^m Z_t = Z_{t-m} \quad (1)$$

$$FZ_t = Z_{t+1}, F^m Z_t = Z_{t+m} \quad (2)$$

$$\Delta Z_t = Z_t - Z_{t-1} = (1 - B)Z_t = W_t \quad (3)$$

$$SZ_t = \sum_{j=0}^{\infty} Z_{t-j} = Z_t + Z_{t-1} + \dots = (1 + B + B^2 + \dots)Z_t \quad (4)$$

Onde: B é o operador translação para o passado; F é o operador translação para o futuro; Δ é o operador diferença ($\Delta = 1 - B$); S é o operador soma ($S = \Delta^{-1}$; $SZ_t = (1 - B)^{-1}Z_t = \Delta^{-1}Z_t$); Z_t é um vetor do conjunto de variáveis da série temporal que varia em função do tempo (t); W_t é a diferença de Z_t ; j é o contador de tempo das translações para o passado e m é o contador de tempo das translações para o futuro.

Considerando que W_t é uma diferença de Z_t , então Z_t é uma integral de W_t , portanto, Z_t segue um modelo auto-regressivo, integrado de médias móveis equação 5:

$$\phi(B)W_t = \theta(B)a_t \quad (5)$$

Onde: $\phi(B)$ é o operador auto-regressivo; $\theta(B)$ é o operador de médias móveis; a_t é o ruído.

O modelo ARIMA, pode ser representado de três formas de acordo com Morettin & Toloi (2006), sendo elas: *i*) em termos de valores prévios de Z_t e do valor atual e prévios de a_t ; *ii*) em termos do valor atual e prévios de a_t ; *iii*) em termos dos valores prévios de Z_t e do valor atual de a_t .

A forma usual do modelo, para o cálculo das previsões a partir das diferenças é dada pela equação 6.

$$Z_t = \varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + \dots + \varphi_{p+d} Z_{t-p-d} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (6)$$

Onde: $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_{p+d} B^{p+d}$ e $\varphi(B)$ é o operador auto-regressivo não-estacionário, de ordem $p + d$, com d raízes iguais a um (sobre o círculo unitário) e as restantes p fora do círculo unitário.

O cálculo da variância dos erros de previsão é dado pela equação 7.

$$Z_t = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \psi(B)a_t \quad (7)$$

Onde: $\varphi(B)Z_t = \varphi(B)\psi(B)a_t$ e $\varphi(B)\psi(B) = \theta(B)$; ψ_j são os pesos que podem ser identificados a partir dos coeficientes de B , B^2 etc.

A identificação do modelo ARIMA é a fase mais crítica do processo, pois implica no ajuste dos dados na técnica aplicada, critérios como. O procedimento mais utilizado consiste em determinar os valores de p, d, q do modelo ARIMA e suas estimativas preliminares dos parâmetros a serem usados no estágio de estimação. São dados em três partes, conforme a literatura de Box & Jenkins. (1976):

(a) verificar a existência da necessidade de transformação na série original, objetivando estabilizar sua variância;

(b) obter a série estacionária a partir das diferenciações, de modo que d seja igual a zero e $\Delta^d Z_t$ seja reduzido a um processo ARMA(p, q);

(c) identificar o processo ARMA(p, q) resultante, utilizando a análise de autocorrelações e autocorrelações parciais estimadas.

Para a estimação do modelo ARIMA é necessário usar um procedimento iterativo de estimação não-linear de mínimos quadrados e as estimativas preliminares encontradas na fase de identificação são utilizadas como valores iniciais para as estimativas preliminares.

Dados os parâmetros $p + q + 1$ do modelo ARIMA(p, d, q), inseridos no vetor $\xi = (\phi, \theta, \sigma_a^2)$, onde $\phi =$

$(\phi_1, \phi_2, \dots, \phi_q)$, $\theta = (\theta_1, \theta_2, \dots, \theta_q)$ e σ_a^2 é uma estimativa da variância residual, a estimativa de ξ é feita pelo método de máxima verossimilhança (EMV) para as N observações Z_1, Z_2, \dots, Z_N .

Segundo Koprinska et al. (2015), o diagnóstico consiste em verificar se o modelo estimado representa ou não, adequadamente, os dados. A técnica mais utilizada para essa etapa é o super ajustamento, onde é realizada uma estimativa do modelo com parâmetros extras, é examinado se estes são significativos e se sua inclusão diminui significativamente a variância residual. Outras técnicas podem ser aplicadas nessa etapa, baseadas nas autocorrelações estimadas dos resíduos, tais como: o teste de autocorrelação residual; o teste de Box-Pierce; teste da autocorrelação cruzada; teste do periodograma acumulado, dentre outros. Uma vez encontrado o modelo satisfatório, a última etapa é a obtenção do modelo de previsão.

A previsão, denotada por $\hat{Z}_t(h)$, tem como objetivo encontrar um valor Z_{t+h} , com $h \geq 1$, supondo as observações ..., Z_{t-2}, Z_{t-1}, Z_t , até o instante t , denominado origem das previsões e h o horizonte alcançado. A forma da equação das diferenças pode ser reescrita de acordo com a equação 8.

$$Z_{t+h} = \phi_1 Z_{t+h-1} + \dots + \phi_{p+d} Z_{t+h-p-d} - \theta_1 a_{t+h-1} - \dots - \theta_q a_{t+h-q} + a_{t+h} \quad (8)$$

E a previsão através do modelo ARIMA(p, d, q) será dada pela equação 9:

$$\hat{Z}_t(h) = \phi_1 [Z_{t+h-1}] + \dots + \phi_{p+d} [Z_{t+h-p-d}] - \theta_1 [a_{t+h-1}] - \dots - \theta_q [a_{t+h-q}] + [a_{t+h}] \quad (9)$$

Pode-se observar que os termos das médias móveis desaparecem para $h > q$; para calcular $\hat{Z}_t(h)$ é necessário obter os valores de $\hat{Z}_t(h-1)$ e $\hat{Z}_t(h-2)$, calculados recursivamente; os valores da sequência de previsões são calculados a partir da introdução de valores iniciais finitos; as inovações de a_t são obtidas também recursivamente. A atualização da previsão Z_{t+h-1} , feita no instante t , é obtida quando um novo dado Z_{t+1} é observado, portanto, a previsão Z_{t+h+1} na origem $t+1$, é feita adicionando-se à $\hat{Z}_t(h+1)$ um múltiplo do erro de previsão $a_{t+1} = Z_{t+1} - \hat{Z}_t(1)$.

Como o modelo ARIMA está ligado aos conceitos de autocorrelação e autocorrelação Parcial, julgou-se interessante, no próximo parágrafo colocar as suas conceituações. A autocorrelação é caracterizada como uma observação em um determinado instante de uma série de valores está relacionada às observações passadas de diversas ordens, denominadas de lags (número de períodos de tempo que separa os dados de séries temporais. Uma autocorrelação de primeira ordem caracteriza séries onde uma observação está correlacionada com a observação imediatamente anterior. A autocorrelação de segunda ordem está correlacionada com as observações a 2 unidades de tempo no passado e, assim, sucessivamente. A identificação da autocorrelação é feita através da Função de Autocorrelação (ACF – Autocorrelation function), discutida mais adiante. Além disso, testes como o de Durbin Watson auxiliam na identificação da autocorrelação de primeira ordem. A autocorrelação parcial é a correlação entre as observações em uma série temporal que não é contabilizada por todos os intervalos mais curtos entre essas observações (Koprinska et al., 2015).

O modelo ARIMA foi desenvolvida na década de 1970, conhecido como modelo de Box e Jenkins, e consiste em ajustar modelos auto-regressivos integrados de médias móveis, ARIMA (p, d, q), a um conjunto de dados, (onde p é o número de parâmetros auto-regressivos, d é o número de diferenciações da série e q é o número de parâmetros de médias móveis) (Morettin & Toloi, 2006). As séries estacionárias podem ser identificadas por três modelos básicos, sendo denominados de auto-regressivos (AR), média móvel (MA), e média móvel auto-regressivo (ARMA), que ao se combinarem em adição à diferenciação regular (RD), tornam-se o modelo ARIMA, onde ocorre então a integração e o procedimento de diferenciação (Box & Jenkins., 1976).

O modelo AR é semelhante ao modelo de regressão linear, com a diferença de que as suas variáveis independentes os valores da variável dependente defasados no tempo, daí, o nome auto-regressivo. O modelo MA trata de uma média móvel ponderada de valores da série produzidos no passado, cujos pesos não precisam ser iguais e nem somar 1. Observa-se na técnica MA que o valor mais recente da série possui um peso maior que os valores atrasados mais distantes, isto é, cada previsão produz

um erro no passado dos valores gerados. Para o modelo ARMA usa-se a função autocorrelação (ACF) e a função autocorrelação parcial (PACF) para identificar quantas defasagens no tempo, a significância das correlações entre a observação presente e as observações passadas se encontrarão, e determinar os coeficientes da regressão linear da série temporal utilizando os valores defasados como variáveis independentes (Morettin & Toloi, 2006).

Segundo Box e Jakins (1976), Campos et al. (2006) e Marin et al. (2016), o modelo ARIMA é capaz de estabelecer relações entre diversas variáveis para a realização de previsões, desde que as séries em estudo sejam estacionárias. Segundo Costa (2012), a aplicação do modelo ARIMA deve ser precedido de três passos:

I – Identificação: se a série não for estacionária, ela deve ser diferenciada até atingir a condição de estacionariedade. Segundo Fracaro (2018), as séries temporais na sua maioria são não estacionárias, sendo necessária a aplicação de uma transformação nas séries temporais.

Uma transformação consiste em tomar as diferenças $\Delta Z_t = Z_t - Z_{t-1}$, denominada operação de diferenciação, podendo repetir n vezes $\Delta^n Z_t$, até se atingir a estacionariedade da série. Dependendo da distribuição dos dados, uma diferenciação pode não ser suficiente, necessitando uma nova diferenciação, obtendo valor de $d = n$, sendo n a quantidade de diferenciações.

II – Estimação e teste: como o valor de d já foi obtido no passo I, obtém-se intuitivamente os valores apropriados para os parâmetros p e q do modelo ARIMA. Para isso, emprega-se as representações gráficas da Função de Autocorrelação (FAC) e Função de Autocorrelação Parcial (FACP) contra o tamanho da defasagem. Alternativamente, a estimação dos parâmetros do modelo pode ser feita por meio de mínimos quadrados ou pelo método de estimativa não-linear. O modelo estimado deve ajustar bem aos dados se os resíduos do modelo são ruídos brancos. Caso sejam, aceita-se o modelo. Caso contrário, o processo deve ser recommçado a fim de se identificar outro modelo e;

III – Previsão: o modelo obtido dos passos anteriores é usado para fazer previsões. Observe que a previsão pode não ser tão precisa, pois, a modelagem matemática é complexa e idealizada para permitir a inclusão de incertezas no modelo.

Para Montgomery et al. (2008), o indicador básico de erro de previsão para um período de tempo t (θ_t) é a diferença entre o valor real Z_t e o valor previsto no tempo correspondente \hat{Z}_t , isto é $\theta_t = Z_t - \hat{Z}_t$. Segundo Montgomery et al. (2008), existem diversas ferramentas para avaliar a qualidade do ajuste e o desempenho de um modelo de previsão, entre elas os mais importantes são o coeficiente de ajustamento R^2 , a Raiz do Erro Médio Quadrático (REMQU), o Erro Médio Absoluto (MAPE) e o BIC normalizado. As equações (10) e (11) representam matematicamente esses erros.

$$REMQU = \sqrt{\frac{1}{N} \sum_{t=1}^N (\theta_t)^2} \quad (10)$$

$$MAPE = \frac{1}{N} \sum_{t=1}^N |\theta_t| \quad (\%) \quad (11)$$

Onde N é o número de períodos da série temporal e $|\theta_t|$ é o valor absoluto da variável θ_t . Quanto menor forem estas medidas, melhor será o desempenho do modelo de previsão tendo em vista que essas medidas avaliam o quanto o modelo está errando. O BIC Normalizado (Normalized Bayesian Information Criterion), critério de informação bayesiano, também chamado de Critério de Schwarz é um critério de seleção de modelo para ser aplicado considerando-se a densidade a *posteriori* dos parâmetros do modelo ajustado. Esse critério pondera entre a função do logaritmo verossimilhança maximizado e o número de parâmetros do modelo. O melhor modelo é aquele que apresenta o menor valor do BIC.

Inicialmente, abordar-se-á artigos que tratam da produção, consumo e efeitos socioambientais causados pelos derivados da cana-de-açúcar, com ênfase especial às produções de açúcar, etanol anidro e etanol hidratado, subprodutos da cana que tem grande influência na economia brasileira.

No artigo “Produção de álcool anidro mais limpa”, Santos et al. (2020) abordam os efeitos causados ao meio ambiente pela emissão de alguns gases poluentes oriundos dos combustíveis fósseis, principais causadores do aquecimento global devido ao

efeito estufa. Esse fato causou a grande procura de substitutos desses combustíveis por fontes de energia mais limpa que possuam sustentabilidade. Os autores afirmam que o Brasil se destaca nessa procura, com a produção do álcool combustível, especificamente o álcool anidro, que é conhecido como um combustível renovável e tem sua produção realizada para substituir o chumbo que é um metal venenoso adicionado à gasolina. A forma de produção do etanol anidro é aparentemente mais cara, porém, rentável a longo prazo garantindo um *payback* bastante rápido. Daí conclui-se que esse tipo de produção além de ser mais limpa enquadra-se em melhoria econômica e ambiental.

Segundo Pacheco (2011), a produção de etanol não impacta muito pouco o meio ambiente, e o Brasil se beneficia disso, em vista da escassez e da alta dos preços dos combustíveis fósseis que penalizam as economias do País. A produção de etanol no Brasil é bastante vantajosa, pelo uso de tecnologias avançadas no cultivo da cana-de-açúcar e na sua produção. Por outro lado, o aumento do cultivo da área plantada de cana-de-açúcar não exige nenhum desmatamento e não concorre com áreas destinadas à produção de alimentos. O Brasil tem uma matriz energética plenamente sustentável, pois, possui 46,8% das fontes de energias renováveis, enquanto a média mundial é de apenas 14%. A população brasileira se beneficia da produção de álcool, pelo fato de ter uma importante indústria automobilística, com destaque de produção de carros flex, que utilizam tanto combustíveis fósseis quanto etanol, colaborando sobremaneira para uma menor poluição atmosférica, reduzindo a emissão de CO₂.

Segundo Viegas (2010), em termos financeiros, quando se compara o preço de produção do etanol brasileiro, que é de US\$0,22/L, o etanol americano, que é produzido a partir do milho, tem o seu custo a US\$0,35/L, portanto, altamente competitivo o custo do etanol brasileiro.

Segundo Lorenzoni (2020), é preciso ficar atento à evolução da comercialização do etanol para um mercado consumidor cuja tendência é crescer, tanto nacional quanto internacionalmente devido às legislações ambientais que tende a priorizar os biocombustíveis em detrimento dos combustíveis fósseis podendo, com isso, estabelecer uma concorrência entre o uso da terra para a produção de cana-de-açúcar e o seu uso para a produção de alimentos, principalmente, naquelas regiões que já não dispõem de terras férteis para essas produções. Daí, é necessário investir no etanol de segunda geração, também conhecido como etanol celulósico, que é uma nova alternativa para o uso energético da biomassa, que traz vantagens ambientais e econômicas, mas cujas tecnologias ainda estão em fase de aperfeiçoamento.

Segundo Lorenzi & Andrade (2020), o Brasil apresenta vantagens competitivas na produção de etanol de segunda geração devido à grande disponibilidade das duas matérias primas que entram na produção desse tipo de etanol devido à melhor eficiência das caldeiras que resulta em excedentes de bagaço de cana, e a proibição da queima da cana para a colheita, resultando em excedente de palha. Esses dois subprodutos da cana-de-açúcar são matérias primas para a produção de etanol de segunda geração. Essa nova tecnologia de produção de etanol utiliza a mesma planta de produção do etanol comum.

Carvalho et al. (2013) no artigo “Cana-de-açúcar e álcool combustível: histórico, sustentabilidade e segurança energética”, afirmam que a cana-de-açúcar tem importância desde os primórdios na história do Brasil, desde o seu descobrimento, e que essa cultura tem papel importante na economia brasileira, sendo um dos principais produtos comercializados. Na década de 1970 com o advento da crise econômica, geopolítica e da possibilidade de esgotamento de petróleo, os países dependentes desse combustível buscaram novas alternativas energéticas. Em 1975 criou-se o Programa Nacional do Álcool (PROALCOOL), que teve alguns anos de ascensão, pelo aumento de destilarias e comercialização de carros movidos com mistura de etanol. Em decorrência da diminuição do preço do barril de petróleo a importância do programa reduziu significativamente, voltando ao auge apenas em 1979, ou seja, a segunda fase do programa. Concebido como um dos vetores da resposta nacional à crise dos preços do petróleo dos anos 1970, o programa persistiu em momentos em ascensão em outros não, alcançando por mais de três décadas. O Brasil é o segundo maior produtor de etanol, perdendo apenas para o EUA, em que usa como matéria prima o milho, o que torna um gargalo a produção de biodiesel, pois compete com a produção de alimentos.

Com o objetivo de fazer previsões sobre produções de cana-de-açúcar e seus subprodutos, analisou-se alguns artigos que

tratam do tema sobre séries temporais, com destaque para o método ARIMA.

Em Brito (2020) há um exemplo de aplicação do modelo ARIMA fora do âmbito financeiro, para previsão de atendimentos em pontos de atenção em um hospital no estado de Minas Gerais. Mostrando a versatilidade e o grande leque de aplicação do modelo ARIMA. Foram criados dois modelos, um ARIMA (3,1,1) para previsão de atendimentos cardiológicos e um ARIMA (2,1,1) para previsão de atendimentos no serviço de Emergência e Urgência feito pelo enfermeiro. Oferecendo uma ferramenta de auxílio na tomada de decisão dos gestores do hospital.

É importante estabelecer previsões da produção da cana e de seus derivados, para que as autoridades afins possam tomar acertadas decisões. Campos et al. (2006), com o objetivo de elaborar um modelo de previsões para o preço do frango inteiro resfriado no grande atacado do estado de São Paulo, utilizou a metodologia ARIMA ou de Box-Jenkins, de previsões de séries temporais. Para isso, foi utilizada uma série histórica mensal de preços de frangos entre os anos de 1996 a 2005, de modo que os testes de previsão ex post foram realizados para os anos de 2004 e 2005. A série de preços foi deflacionada atualizando os índices IGP-DI da Fundação Getúlio Vargas (FGV) para o mês de dezembro de 2005, como forma de eliminar o efeito da inflação. Após os testes foram determinados 4 modelos que demonstraram consistência estatística e bom desempenho de previsões. Foram realizados testes de previsões trimestrais, anuais e para 2 anos, em que todos os modelos apresentaram bons desempenhos, o que não era esperado para os testes de 2 anos, já que a metodologia ARIMA univariada é reconhecidamente eficiente apenas para previsões de curto prazo. Os bons resultados alcançados pelo modelo fornecem boas expectativas para seu uso como apoio ferramental a profissionais de diversas áreas como planejamento, orçamento, investimentos, entre outras.

Segundo Martin et al. (2016), em um ambiente cada vez mais competitivo, produzir a quantidade estritamente necessária é um fator fundamental para manter a sustentabilidade da empresa. Ao se prever o comportamento de crescimento do número de automóveis no país, são vários os benefícios que vão além da previsibilidade para o setor automobilístico, pois, auxilia também nas tomadas de decisões pelos setores públicos. Dentre os vários métodos para prever o comportamento de uma empresa, optou-se em usar as séries temporais que consideram os dados de uma variável ordenados cronologicamente e, que podem fornecer informações futuras através vários métodos, como os autorregressivos integrados de média móvel (ARIMA) ou suavização exponencial. O trabalho teve como objetivo analisar o crescimento do número de automóveis no município de Joinville/SC por meio do método ARIMA e suavização exponencial, utilizando dados mensais do número de automóveis, caminhonetes, camionetes e utilitários, entre janeiro de 2003 e março de 2014. Constatou-se que os modelos SARIMA(0,2,1)x(1,0,0)₁₂ apresentaram os menores erros dentro e fora da amostra, portanto, foi o mais adequado para realizar as previsões.

Em Azevedo et al. (2015) os modelos ARIMA criados para os estados de SP, MG, RJ, RS, PR, BA aumentaram a acurácia das previsões dos cálculos de arrecadação do ICMS, dando perspectivas diferentes para as tomadas de decisões em cada Estado. Entretanto, esse trabalho alerta sobre dois aspectos importantes, a necessidade de atualização do modelo com a entrada de novos dados e a questão da fragilidade do modelo em caso de mudanças bruscas no aspecto macroeconômico.

Carneiro Junior et al. (2020), em seu artigo Aplicação de redes neurais artificiais na previsão do produto interno bruto do Mato Grosso do Sul (MS) em função da produção de cana-de-açúcar, açúcar e etanol buscou responder à questão qual a capacidade preditiva das Redes Neurais Artificiais (RNAs) para explicar o Produto Interno Bruto de MS em função da das variáveis independentes, produção de cana-de-açúcar, açúcar e etanol? Assim, o objetivo geral do trabalho é estimar o Produto Interno Bruto de MS relativo à produção de cana-de-açúcar, de açúcar e de etanol utilizando as RNAs. A relevância da pesquisa foi a utilização desta importante ferramenta de predição para pode prever o fluxo monetário que será demandado com o incremento da produção da cana-de-açúcar, do açúcar e do etanol, podendo planejar investimentos no sentido de propiciar melhores condições de vida à população, não só nos entornos das usinas, mas no MS que será beneficiado como um todo. A série temporal em análise compreende o intervalo de 2002 a 2015. O software utilizado para construir a RNA foi o SPSS - Statistical Package for the Social Sciences, pacote estatístico para as Ciências Sociais. Concluiu-se que o modelo predito de RNA foi capaz de explicar o comportamento do

PIB do MS em função das variáveis independentes descritas, além disso, o modelo apresentou valores e observações satisfatórias para sua validação. Por fim, utilizando dados hipotéticos para as variáveis independentes, o modelo predito e validado foi utilizado para realizar uma previsão com significativa segurança.

3. Resultados e Discussão

Neste item, serão usados recursos de ARIMA em séries temporais para fazer previsões de produções de cana-de-açúcar, de açúcar, de etanol anidro e de etanol hidratado, na macrorregião Centro-Sul brasileira, no período de 2020 à 2025. Inicialmente, foram coletados os dados desses produtos via internet, com periodicidade anual, no período de 1980 a 2020, totalizando 40 observações. No Quadro 1 estão apresentadas as coletas realizadas das produções de cana-de-açúcar, de açúcar, de etanol anidro e de etanol hidratado na macrorregião Centro-Sul, no período de 1980 a 2020.

Quadro 1. Produções de cana, açúcar, etanol anidro e etanol hidratado na macrorregião Centro-Sul do Brasil, no período de 1980 a 2020.

Ano	Cana-de-Açúcar (mil t)	Açúcar (mil t)	Etanol Anidro (mil m ³)	Etanol Hidratado (mil m ³)
1980	83432	5253	1823	1233
1981	58387	5146	1014	2400
1982	65320	5613	3003	1632
1983	66664	5512	2124	4610
1984	145780	5282	1827	5761
1985	163374	4834	2813	6986
1986	156608	4809	1759	6542
1987	169795	5027	1788	7880
1988	166487	5253	1513	8451
1989	162436	4140	1303	8598
1990	170195	4509	1088	8620
1991	179031	5761	1802	9125
1992	176295	6118	1938	8120
1993	173264	7304	2379	8003
1994	196037	9407	2577	8555
1995	204383	10185	2588	8256
1996	231558	11618	3821	8257
1997	248730	11354	4762	8489
1998	269749	15160	4804	7412
1999	263894	16900	5370	6234
2000	207099	12643	4802	4262
2001	244218	15972	5746	4430
2002	270407	18778	6270	4882
2003	298772	20426	8029	4984
2004	327806	22149	7298	6265
2005	336783	22015	7275	7036
2006	372754	25796	7066	9000
2007	431114	26201	7143	13191
2008	505116	26750	8247	16868

2009	541962	28645	6206	17479
2010	556945	33501	7413	17971
2011	493159	31304	7454	13089
2012	532758	34097	8730	12632
2013	597061	34295	11008	14568
2014	573145	32011	10740	15492
2015	617709	31221	10643	17581
2016	607137	35628	10655	14996
2017	596330	36060	10419	15670
2018	573169	26510	9141	21812
2020	590361	26761	9946	23313
2020	605462	38465	9688	20675

Fonte: Adaptado de UNICA (2021).

Nas análises das séries temporais do quadro 1 foi utilizado o software Statistical Package for the Social Sciences (SPSS), que é um software que tem aplicações em diversas áreas do conhecimento, especialmente, a Estatística (Corres et al., 2008). Para a utilização do modelador de séries temporais do SPSS foi necessário conhecer os valores dos parâmetros do modelo ARIMA (p, d, q), que possuem os seguintes significados: (p) auto-regressivo – o valor de (p) indica o número de períodos de tempos anteriores (lags) da série que serão usados para prever os valores atuais; (d) diferença – o valor de (d) indica o número de diferenciações aplicadas à série para torná-la estacionária. Quando a série temporal apresenta tendência, certamente, não é estacionária, daí, a diferenciação é necessária para remover o seu efeito e; (q) - média móvel – o valor de (q) indica o número de observações que serão consideradas para o cálculo da média no modelo para a previsão.

Desse modo, tomando as séries históricas de produções de cana-de-açúcar, de açúcar, de etanol anidro e de etanol hidratado, de 1980 a 2020, na macrorregião Centro-Sul do Brasil, (Quadro 1), um primeiro teste foi realizado para verificar os níveis de associações entre as variáveis, e facilitar a determinação das variáveis dependentes e independentes do modelo, foi a utilização do software SPSS através do comando “Analisar – Correlacional – Bivariável”, para determinar a matriz de correlação entre as variáveis consideradas, conforme Tabela 1.

Tabela 1. Matriz das correlações entre as variáveis Produções de Cana, de Açúcar, de Etanol Anidro e Etanol Hidratado da macrorregião Centro-Sul do Brasil, no período de 1980 a 2020.

Produção	Cana	Açúcar	Etanol_Anidro	Etanol_Hidratado
Cana	1			
Açúcar	0.960854	1		
Etanol_Anidro	0.931671	0.952736	1	
Etanol_Hidratado	0.887555	0.754326	0.707221	1

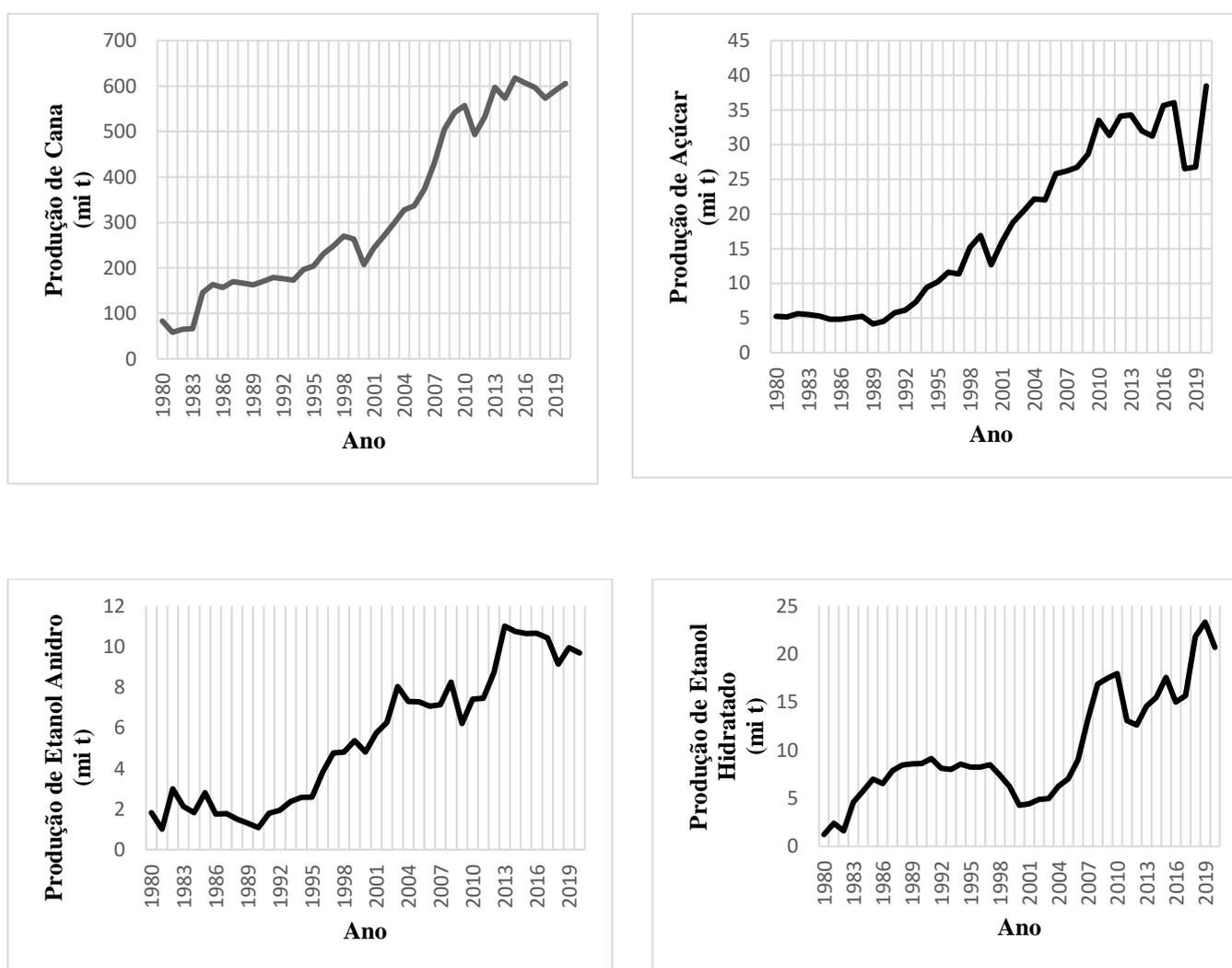
Fonte: Autores (2022).

Da Tabela 1 observa-se fortes dependências entre todas as variáveis, retratadas pelas altas correlações entre as mesmas, particularmente, com fortes dependências entre as variáveis Produções de Açúcar, de Etanol Anidro e Etanol Hidratado com a variável Produção de Cana, que surgiria naturalmente como variável dependente das demais, mas como já afirmado, os níveis de produções dependem mais de políticas governamentais. As altas correlações são devido a cana de açúcar ser a matéria prima de todos os outros produtos analisados.

Para a aplicação da metodologia ARIMA em uma série temporal alguns pré-requisitos devem ser satisfeitos. Primeiramente, a série deve ser estacionária, caso não seja, deve-se diferenciá-la sequencialmente, sazonalmente ou de ambos os modos, para torná-la estacionária. Para a verificação da estacionariedade das séries optou-se por construir as suas representações gráficas ao longo do tempo, neste caso, utilizando o software Microsoft Excel, considerando como variável independente a variável “Ano” e como variáveis dependentes, as variáveis “Produção de Cana”, “Produção de Açúcar”, “Produção de Etanol Anidro” e “Produção de Etanol Hidratado”.

A Figura 1 apresenta os gráficos das séries históricas das produções de cana, açúcar, etanol anidro e etanol hidratado na macrorregião Centro Sul do Brasil, no período de 1980 a 2020.

Figura 1. Gráficos das séries históricas das produções de cana-de-açúcar, de açúcar, etanol anidro e etanol hidratado na macrorregião Centro Sul do Brasil, no período de 1980 a 2020.

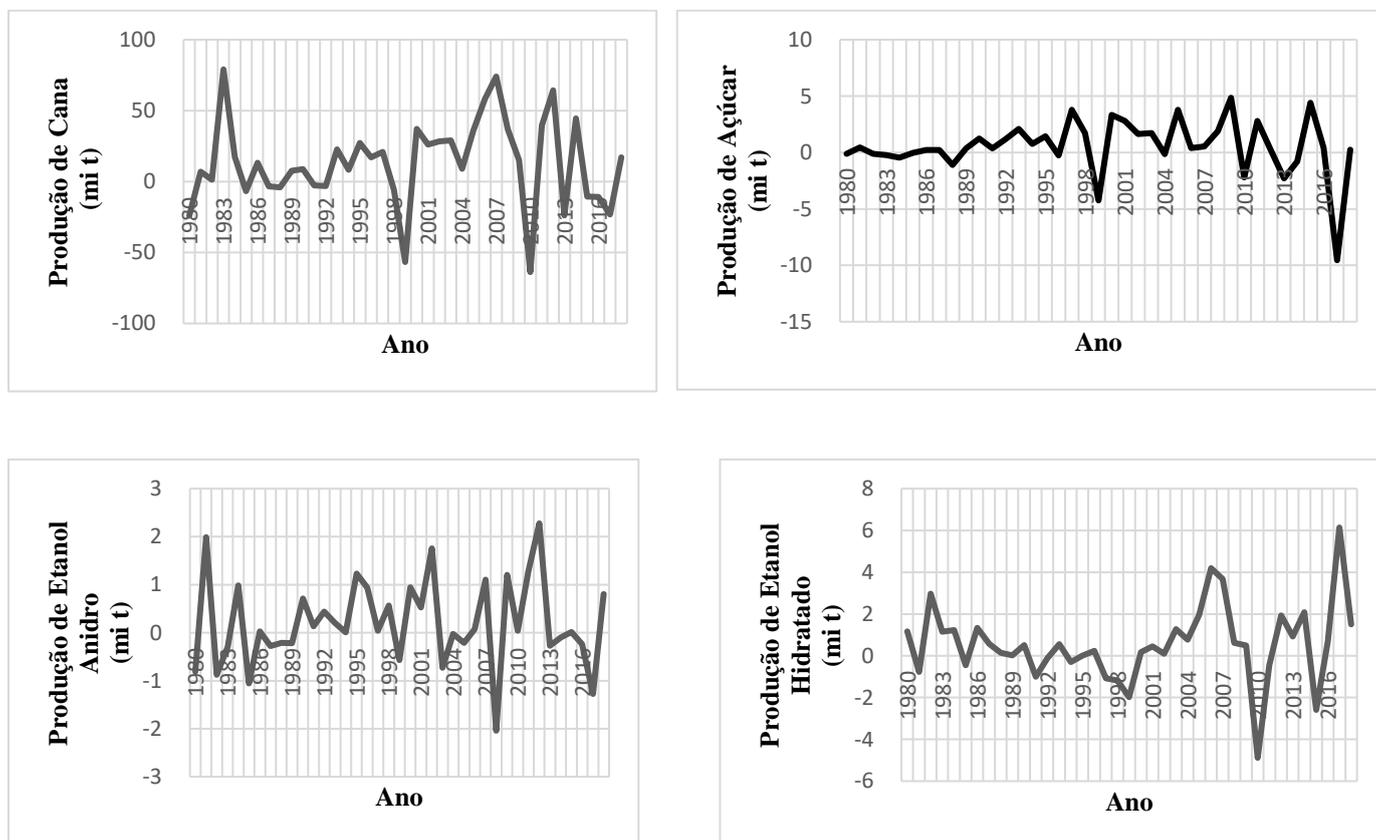


Fonte: Autores (2022).

Observa-se, da Figura 1, que nenhuma das séries são estacionárias, todas elas apresentando tendências de crescimento. Desse modo, para se fazer previsões utilizando o método ARIMA, essas séries precisam ser diferenciadas, sequencialmente, até atingirem as estacionariedades. No SPSS essa diferenciação é feita através do comando “Analisar – Previsão – Gráficos de Sequências - Diferença”, com a indicação do número de diferenciações necessárias para tornar a série estacionária, já para o traçado

dos gráficos. Neste caso, diferenciando as quatro séries, utilizando um intervalo de tempo (no caso, $d = 1$ ano), já foi suficiente para transformá-las em séries estacionárias, cujos gráficos diferenciados estão na Figura 2.

Figura 2. Gráficos das quatro séries históricas, diferenciadas uma vez ($d = 1$ ano), das produções de cana-de-açúcar, de açúcar, etanol anidro e etanol hidratado na macrorregião Centro Sul do Brasil, no período de 1980 a 2020.



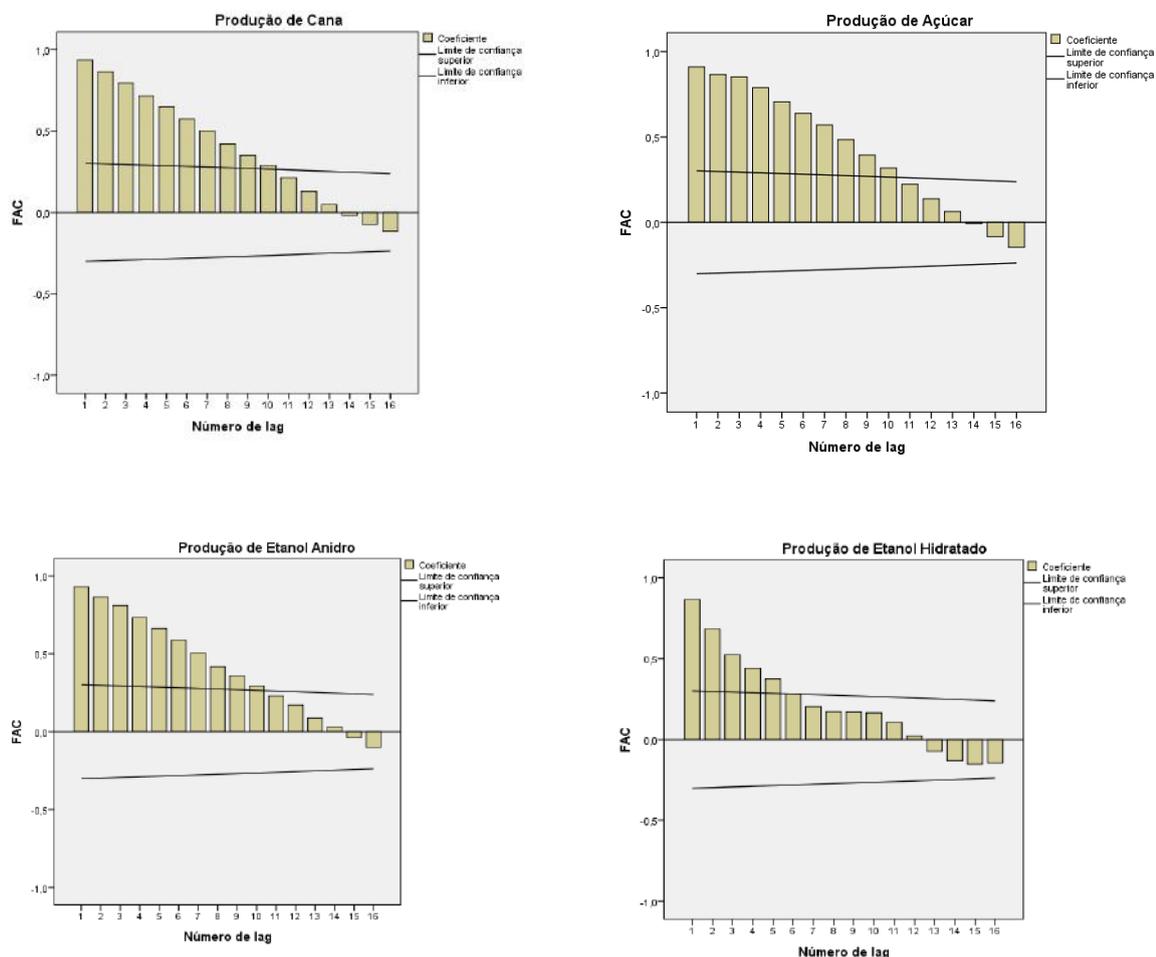
Fonte: Autores (2022).

Da Figura 2, pode-se observar que todas as quatro séries passaram a ser estacionárias. Percebe-se pelos gráficos que existem volatilidades em todas as séries, mas essas variações ocorrem em torno de uma média. Por outro lado, percebe-se que já não existem mais tendências. Portanto, as quatro séries diferenciadas em relação a um intervalo de tempo (1ª. ordem) já permitem as aplicações da metodologia ARIMA. A segunda condição para a aplicação do ARIMA é verificar se os valores atuais das séries estão correlacionados com valores passados das mesmas, fatos que acontecem frequentemente para séries temporais de dados econômicos.

O número de lags a ser considerado como o mais relevante é dado pelo maior valor da autocorrelação e autocorrelação parcial correspondentes, levando-se em conta a sua significância. Esse fenômeno caracteriza a sigla AR (autoregressive) do ARIMA, isto é, dados atuais estão fortemente correlacionados a dados anteriores e estatísticas complementares podem ser obtidas no SPSS através do comando “Analisar – Previsão – Criar Modelos Tradicionais”, em que se insere como variáveis dependentes “Produção de Cana”, “Produção de Açúcar”, “Produção de Etanol Anidro” e “Produção de Etanol Hidratado”, considerando que os seus níveis de produção estão fortemente correlacionados com os seus níveis de produções anteriores, portanto, não havendo variáveis independentes em cada um dos modelos. A figura 3 apresenta os gráficos das autocorrelações das séries temporais Produção de Cana, Produção de Açúcar, Produção de Etanol Anidro e Produção de Etanol Hidratado na macrorregião Centro-Sul

do Brasil, no período de 1980 a 2020.

Figura 3. Gráficos das autocorrelações das séries temporais Produções de Cana, de Açúcar, de Etanol Anidro e de Etanol Hidratado na macrorregião Centro-Sul do Brasil, no período de 1980 a 2020.



Fonte: Autores (2022).

Os gráficos da Figura 3 apresentam os lags e suas significâncias. O lag 1 é o mais significativo para todos os gráficos, e vai decaindo exponencialmente com o passar dos tempos. As duas linhas uma acima e outra abaixo do eixo horizontal apresenta os limites de confiança inferior e superior, respectivamente. As colunas representativas dos lags que ultrapassam os limites de confiança são significativas, quanto maior mais significativa. Desse modo, o primeiro lag, ou lag = 1, é o mais significativo em todos os gráficos, confirmando o que já foi exposto, de que o valor atual depende somente do que aconteceu há um ano atrás. Assim, a Tabela 2 apresenta os valores das autocorrelações com as suas respectivas significâncias das quatro séries em análise.

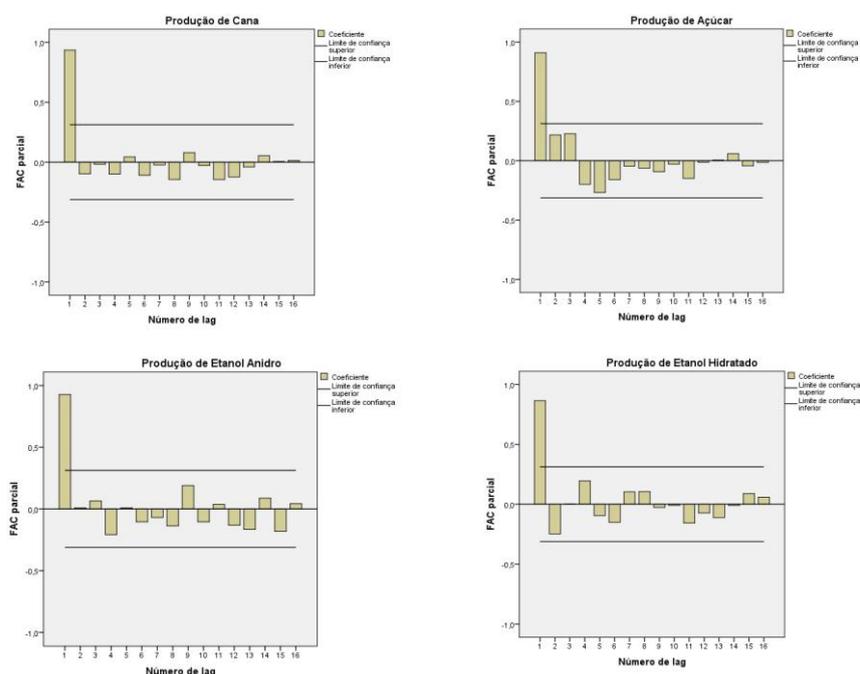
Tabela 2. Valores dos lags, das autocorrelações e das estatísticas de Ljung-Box das séries Produções de Cana, de Açúcar, de Etanol Anidro e de Etanol Hidratado na macrorregião Centro-Sul do Brasil, no período de 1980 a 2020.

Série	Lag	Autocorrelação	Estatística de Ljung - Box		
			Valor	gl	Sig.
Produção de Cana	1	0,935	35,645	1	0,000
Produção de Açúcar	1	0,952	39,018	1	0,000
Produção de Etanol Anidro	1	0,922	36,603	1	0,000
Produção de Etanol Hidratado	1	0,824	29,283	1	0,000

Fonte: Autores (2022).

A Tabela 2 apresenta para todas as séries os lags iguais a um como os mais importantes, com os maiores valores das autocorrelações, que são muito significantes pelo teste de Ljung-Box. Desse modo, conclui-se que os acontecimentos atuais para todas as séries temporais dependem mais fortemente do que aconteceu há um ano atrás com cada uma delas. Na Figura 4 apresenta-se os gráficos das autocorrelações parciais, com os respectivos limites de confiança inferior e superior.

Figura 4. Gráficos das autocorrelações parciais, com os respectivos limites de confianças inferior e superior das séries temporais Produções de Cana, de Açúcar, de Etanol Anidro e de Etanol Hidratado na macrorregião Centro-Sul do Brasil, no período de 1980 a 2020.



Fonte: Autores (2022).

Pelos gráficos da Figura 4 observa-se que só apresenta significância o primeiro lag de cada série temporal. Após o lag 1, existe uma queda abrupta dos valores das autocorrelações parciais, de modo a torna-las não significativas. Desse modo, conclui-se que não existe parâmetros de média móvel, pois, caso houvesse aconteceria o contrário, com os valores das autocorrelações caindo abruptamente após o primeiro lag e os valores das autocorrelações parciais decrescendo exponencialmente a partir do primeiro lag. Coincidentemente, o valor da autocorrelação parcial do primeiro lag de cada série é o mesmo valor da autocorrelação das respectivas

séries (Tabela 1). Desse modo, já se tem todos os pré-requisitos para se fazer a previsão da produção de cana, de açúcar, de etanol anidro e etanol hidratado.

Para realizar a previsão, utilizando o SPSS, invoca-se o seguinte comando “Analisar – Previsão – Criar Modelos Tradicionais – Modelo – Estatísticas”. Em “Modelo - Critérios” optou-se por “ARIMA (1, 1, 0), baseado nas análises anteriores em que se obteve $p = 1$ e $d = 1$. Em “Estatísticas” solicitou-se “exibir medidas de ajustes com as suas significâncias (Sig.), R^2 - raiz quadrada do erro quadrático médio, (REMQ) - erro absoluto médio em percentual (MAPE %), BIC normalizado, Estatísticas para modelos individuais, Exibir previsões”. A tabela 3 apresenta parte dos resultados desses comandos.

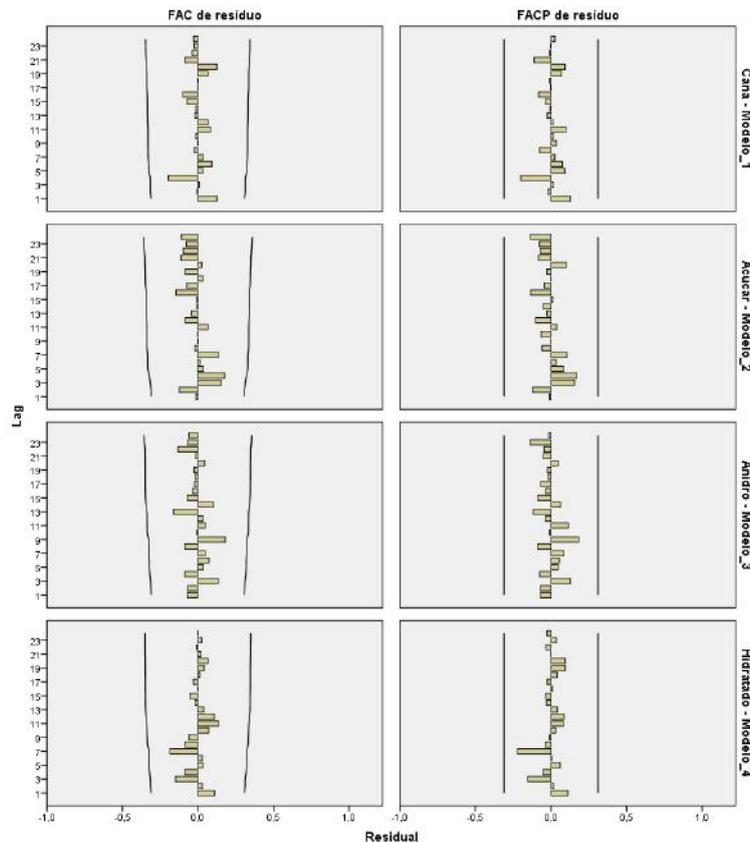
Tabela 3. Estatísticas de ajuste R^2 , REMQ, MAPE (%), BIC Normalizado e Sig (Ljung-Box) dos modelos determinados pelas quatro séries temporais produções de cana, açúcar, etanol anidro e etanol hidratado na macrorregião Centro-Sul do Brasil, no período de 1980 a 2020.

Modelo	Estatísticas de ajuste do modelo				
	R^2	REMQ	MAPE (%)	BIC normalizado	Sig. (Ljung_Box)
Produção de Cana – Modelo 1	0,974	30424,525	9,102	20,830	0,403
Produção de Açúcar – Modelo 2	0,933	3040,714	12,285	16,224	0,089
Produção de Etanol Anidro - Modelo_3	0,933	866,273	18,777	13,713	0,766
Produção de Etanol Hidratado - Modelo_4	0,886	1878,596	14,436	15,261	0,699

Fonte: Autores (2022).

Da Tabela 3 observa-se que, pelos valores dos parâmetros, que as qualidades dos ajustes para os modelos estão boas, com R^2 altos para todos os modelos, o BIC Normalizado para a comparação de modelos indica que o modelo mais bem ajustado é o Modelo 3, pois, tem o menor valor, os valores das significâncias de Ljung-Box mostram que os modelos realmente estão bem ajustados (Sig > 0,05), ou seja não-significativas. Na Figura 5 apresenta-se os gráficos das autocorrelações (FAC) e autocorrelações parciais (FACP) de resíduos, com seus respectivos limites de confiança inferior e superior das séries temporais produções de cana, de açúcar, de etanol anidro e de etanol hidratado na macrorregião Centro-Sul do Brasil, no período de 1980 a 2020.

Figura 5. Gráficos das autocorrelações (FAC) e autocorrelações parciais (FACP) de resíduos, com seus respectivos limites de confiança inferior e superior das séries temporais produções de cana, açúcar, etanol anidro e etanol hidratado na macrorregião Centro-Sul do Brasil, no período de 1980 a 2020.



Fonte: Autores (2022).

Analisando a Figura 5, que apresenta os gráficos das autocorrelações (FAC) e autocorrelações parciais (FACP) de resíduos das quatro séries temporais, após a retirada das sazonalidades (que neste caso não houve, visto que o valor de $q = 0$ em cada um dos modelos), das tendências, parâmetros autorregressivos e outliers, conclui-se que os quatro modelos estão bem ajustados, pois, todos os gráficos da figura são ruídos brancos, mostrando aleatoriedades desses resíduos, principalmente, nos gráficos de autocorrelações parciais de resíduos das séries. A Tabela 4 apresenta as previsões de produções de cana, açúcar, etanol anidro e etanol hidratado na macrorregião Centro-Sul do Brasil, no período de 2021 a 2025, coroando o objetivo principal deste trabalho. Estão apresentados também os limites superiores de confiança (LSC), bem como os limites inferiores de confiança (LIC).

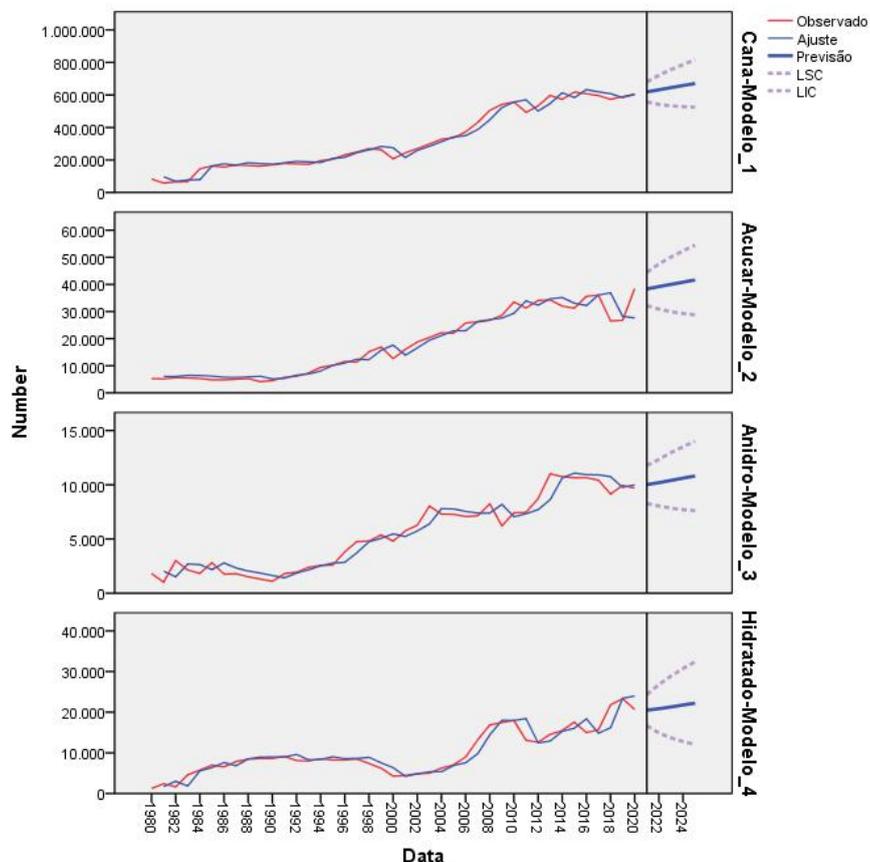
Tabela 4. Previsões de produções de cana, açúcar, etanol anidro e etanol hidratado na macrorregião Centro-Sul do Brasil, no período de 2021 a 2025.

Modelo		2021	2022	2023	2024	2025
Produção de Cana_Modelo 1	Previsão	618591	631588	644577	657565	670553
	LSC ¹	680176	721630	756193	787223	816033
	LIC ²	557005	541546	532960	527906	525072
Produção de Açúcar_Modelo 2	Previsão	38326	39219	40022	40833	41643
	LSC	44481	47553	50100	52392	54515
	LIC	32170	30884	29944	29274	28772
Produção de Etanol Anidro_Modelo 3	Previsão	10025	10192	10408	10609	10815
	LSC	11776	12344	12972	13509	14020
	LIC	8274	8040	7843	7710	7610
Produção de Etanol Hidratado_Modelo 4	Previsão	20509	20849	21293	21758	22227
	LSC	24312	26803	28904	30742	32404
	LIC	16706	14895	13681	12773	12050

¹LSC: Limite Superior de Confiança; ²LIC: Limite Inferior de Confiança. Fonte: Autores (2022).

As previsões de produções de cana, de açúcar, de etanol anidro e de etanol hidratado na macrorregião Centro-Sul do Brasil, no período de 2021 a 2025 se mostram adequadas para o problema, que dependem muito de decisões governamentais e, por isso, observa-se da Tabela 3, que os quatro modelos apresentaram erros percentuais em torno de 13,6%, em média. Observa-se, pela Tabela 4, que todas as produções são crescentes, indicando que o setor da cana-de-açúcar na macrorregião Centro-Sul do Brasil tem um grande potencial para crescimento, impulsionando a economia da região. Na Figura 6 estão expostos os gráficos dos valores observados, o ajuste dos modelos, a previsão e os limites de confiança inferior e superior das séries históricas de produções de cana, açúcar, etanol anidro e etanol hidratado na macrorregião Centro-Sul do Brasil, no período de 1980 a 2020 e de 2021 a 2025 correspondentes as previsões, bem como os limites de confiança.

Figura 6. Gráficos dos valores observados, o ajuste dos modelos, a previsão e os limites de confiança inferior e superior das séries históricas de produções de cana, açúcar, etanol anidro e etanol hidratado na macrorregião Centro-Sul do Brasil, no período de 1980 a 2020 e as previsões correspondentes, bem como os limites de confiança, de 2021 a 2025.



Fonte: Autores (2022).

Observa-se da figura 6 que, apesar do erro médio, em percentual, ter ficado em torno de 13,6%, em média, foi possível ter uma boa visão da evolução desse importante setor da economia brasileira, mostrando que a mesma está em plena evolução, e que depende de mais pesquisas no setor para uma melhor adequação das políticas públicas do Governo Federal.

4. Conclusão

Neste estudo ficou evidenciado que o Brasil apresenta uma participação elevada no mercado internacional de açúcar, aproximadamente 40% das exportações mundiais são brasileiras, o que, cada vez mais, justifica estudos dessa natureza. A escolha das variáveis produção de açúcar e produção de etanol anidro para a realização das previsões se deve ao fato de que o açúcar tem uma pauta elevada nas exportações brasileiras e o etanol anidro é utilizado como combustível no mercado interno brasileiro e adição à gasolina. Também, existe um alto potencial para a exportação do etanol brasileiro.

Os resultados obtidos podem ser usados pelos órgãos públicos e também responsáveis pelo setor canavieiro para as tomadas de decisão de produção, tributária, comerciais, entre outros. Sugere-se que novos estudos sejam realizados, com a inserção de novas variáveis no modelo como área plantada, preços de vendas, entre outras, com o intuito de se obter previsões mais ligadas à realidade.

Utilizando uma série de observações, de 1980 a 2020, relativas as produções de cana-de-açúcar, produção de açúcar, de etanol anidro e hidratado, foi possível fazer previsões das produções de açúcar e etanol anidro com seis anos de antecedência, isto é, de 2020 a 2025, o que mostra a eficácia do software utilizado, o SPSS. Os erros médios absolutos em percentuais (MAPE) foram

relativamente altos para as quatro previsões realizadas, em torno de 13,6%, em média. Para os padrões de trabalhos científicos na área é de alta dimensão devido a não aleatoriedade dos resultados, pois, dependem de decisões governamentais. Mesmo assim, o trabalho é de grande valia, pois, propicia condições para novas tomadas de decisão.

Futuras pesquisas sobre a aplicação dessa metodologia em outras séries econômicas, principalmente, daquelas commodities cujos volumes de produções não dependam de decisões governamentais, para avaliar o seu grau de adaptação. Estudos também podem ser realizados com outras metodologias de previsão de séries temporais utilizando, por exemplo, redes neurais artificiais (RNA), para comparações dos resultados. Um complemento a este trabalho seria a inclusão de outras variáveis ao modelo aqui estudado tais como áreas plantadas, preços, etc., para verificar se há ganhos significativos de predição.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da Bolsa de Produtividade em Pesquisa.

Referências

- Box, G. E. P. & Jenkins, G. M. (1976). *Time series analysis: forecasting and control*. Holden-Day.
- Campos, P. A. C., Clemente, A. & Cordeiro, A. A. L. de. (2006). Aplicação do modelo ARIMA para previsão do preço do frango inteiro resfriado no grande atacado do estado de São Paulo. *XIII Congresso Brasileiro de Custos – Belo Horizonte*, MG, Brasil, 30 de outubro a 01 de novembro de 2006.
- Carneiro Junior, J. B. A. & Souza, C. C. (2020). Aplicação de redes neurais artificiais na previsão do produto interno bruto do Mato Grosso do Sul em função da produção de cana-de-açúcar, açúcar e etanol. *Revista Ibero-Americana de Ciências Ambientais (RICA)*. 10(5).
- Carvalho, L. C., Bueno, R. C. O. F., Carvalho, M. M., Favoretto, A. L. & Godoy, A. F. (2013). Cana-de-açúcar e álcool combustível: histórico, sustentabilidade e segurança energética. *Enciclopédia Biosfera*. Centro Científico Conhecer. Goiânia. 9(16), 530.
- Corres, G.; Esteban, A.; García, J. & Zárata, C. (2008). Metodología Box – Jenkins (ARIMA) con SPSS. Parte I. *Revista de la EPIO*. 29, 161-176.
- Fracaro, N. (2018). *Estacionaridade das séries temporais do modelo matemático arimax de propulsores eletromecânicos*. Universidade Regional do Noroeste do Estado do Rio Grande do Sul - UNIJUÍ.
- Koprinska, I.; Rana, M. & Agelidis, V. G. (2015). Correlation and instance based feature selection for electricity load forecasting. *Knowledge-Based Systems*, 82, 29-40.
- Leite, R. S. & Cortez, L. A. B. (2020). *O Etanol Combustível no Brasil*. <http://alunosonline.uol.com.br/geografia/producao-etanol-no-brasil.html>.
- Lorenzi, B. R. & Andrade, T. N. (2020). O etanol de segunda geração no Brasil: políticas e redes sociotécnicas. *Revista Brasileira de Ciências Sociais*. 34(100), 1-32.
- Martin, A. C., Henning, E., Carvalho, W. O. M. F. & Konrath, A. C. (2016). Análise de séries temporais para previsão da evolução do número de automóveis no Município de Joinville. *Revista Espacios*. 37(06), 29-44.
- Montgomery, D. C., Jennings, C. L. & Kulahci, M. (2008). *Introduction to time series analysis and forecasting*. John Wiley & Sons. Inc. 287p.
- Costa, W. S. L. (2012). Tendências do preço da arroba do boi gordo, uso de séries temporais e modelo arima no SPSS. *Cadernos do leite*. 12(12), 72-81.
- Morettin, P. A. & Toloi, C. M. C. (2006). *Análise de Séries Temporais*. Blucher, 541p.
- Pacheco, T. F. (2011). *Produção de Etanol: primeira ou segunda geração?* <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/32985/1/CITE-04.pdf>.
- PROÁLCOOL – Programa Brasileiro de Álcool. (2008). *Proálcool: Lição de Vida*. <http://www.biodieselbr.com/proalcool/proalcool.htm>.
- Raupp, F. M. & Beuren, I. M. (2004). Metodologia da pesquisa aplicada às ciências sociais. In: Beuren, I.M. (Org.). *Como elaborar trabalhos monográficos em contabilidade: teoria e prática*. (2a ed.), Atlas. 76-97.
- Rodrigues, L. D. (2010). *A cana-de-açúcar como matéria-prima para a produção de biocombustíveis: impactos ambientais e o zoneamento agroecológico como ferramenta para mitigação*. <http://atividadarural.com.br/artigos/5601927a79cad.pdf>.
- Santos, L. M. A., Marques, H. L., Sousa, V. C. de; Rabesco, G. C., Ferreira, R. M., Marques, H. P. G., Oliveira, D. A. de & Vieira, P. T. (2020). Produção de álcool anidro mais limpo. *Científica Multidisciplinar Núcleo do Conhecimento*. Ano 04, 11ed. 03, 67- 80.
- UDOP – União Nacional de Bioenergia. (2003). *A História da Cana-de-açúcar: Da Antiguidade aos Dias Atuais*. <https://www.udop.com.br/noticia/2003/01/01/a-historia-da-cana-de-acucar-da-antiguidade-aos-dias-atuais.html>.
- UNICA. Observatório da Cana. (2021). *Históricos de moagem de cana-de-açúcar, área plantada e colhida, produção de açúcar e etanol no Brasil*. Disponível em <https://observatoriocana.com.br/>.
- Viegas, T. (2010). *Competitividade internacional do etanol brasileiro: oportunidades e ameaças*. Rio de Janeiro: Blog Infopetro, Instituto de Economia: Universidade Federal do Rio de Janeiro.