

ANÁLISE DE SÉRIES TEMPORAIS PRODUTIVAS NO SETOR DO PETRÓLEO E GÁS NATURAL

Felipe Bezerra Pimentel Araújo¹
José Artur Lima Cabral Marques²
Inocêncio Sanches dos Santos Neto³
Allan Kardec Barros Duailibe Filho⁴

RESUMO

O setor de petróleo e gás natural compreende uma quantidade significativa de séries históricas relevantes (séries de preço, produção, exportação etc.), modelos probabilísticos ou estocásticos podem ser produzidos com objetivo de observar comportamentos e previsões para tomadas de decisão mais eficientes. Nessa perspectiva, duas séries temporais produtivas são analisadas com técnicas de decomposição STL e previsão com modelos ARIMA para caracterizar passado, presente e futuro das séries. Uma delas representa a produção de petróleo e a outra a produção de gás natural no país, sendo assim estratégicas para tomadas de decisão. É notável as características de crescimento e tendência ascendente na produção tanto de barris de petróleo como em metros cúbicos de gás natural. Além disso, pela análise das séries produtivas em questão identifica-se através dos modelos preditivos que esta tendência nos dados deve continuar para o horizonte de 10 anos, com mais fôlego para a produção de gás natural do que para a produção de petróleo.

Palavras-chave: Decomposição STL, Modelos ARIMA, Previsão, Petróleo, Gás Natural.

INTRODUÇÃO

O elevado volume de dados caracteriza uma transformação social e cultural, refletindo-se na nova fase da revolução industrial. Para o meio acadêmico, a análise de dados de séries temporais no setor de petróleo e gás natural pode ser um instrumento que aplica algum tipo de transformação nos dados em busca de conhecimento. Previsões de séries temporais podem ser importantes em várias situações de tomada de decisão, como por

¹ Mestrando do Curso de Engenharia Elétrica da Universidade Federal do Maranhão- UFMA, bolsista do Programa Multidisciplinar de Formação de Recursos Humanos em áreas de interesse do setor do petróleo (PRH-ANP), felipe.pimentel@discente.ufma.br;

² Doutorando pelo Curso de Engenharia Elétrica da Universidade Federal do Maranhão - UFMA, jarturcabral@gmail.com;

³ Doutor pelo Curso de Engenharia Elétrica da Universidade Federal do Maranhão - UFMA, netto.ufma@gmail.com;

⁴ Professor orientador: Dr., Programa de Pós-graduação em Engenharia Elétrica - UFMA, allan.kardec@ufma.br.

exemplo, decidir a construção de uma nova refinaria nos próximos cinco anos a partir de previsões de demanda futura; ou prever a quantidade de suprimentos de um estoque.

O diagnóstico de séries temporais pode auxiliar na investigação do procedimento formador da série temporal, a análise pode ainda ser útil para estabelecer previsões de valores futuros, podendo ser de curto prazo ou a longo prazo. O estudo pode ser utilizado para detalhar o desempenho da série, atestando a presença de tendências, ciclos e variações sazonais (MORETTIN; TOLOI, 2006).

Levando-se em consideração que o setor de petróleo e gás natural compreende uma quantidade significativa de séries históricas relevantes (séries de preço, produção, exportação etc.), modelos probabilísticos ou estocásticos podem ser produzidos com objetivo de observar comportamentos e previsões para tomadas de decisão mais eficientes. Analisar, identificar e prever esses comportamentos são estratégias interessantes para setores de petróleo e gás natural e também são os objetivos deste trabalho. Tomadas de decisão baseadas em análise de séries temporais podem se tornar vantagem competitiva representando um menor risco financeiro e, conseqüentemente menores perdas de investimento e custo operacional.

METODOLOGIA

A metodologia baseia-se em análise de dados abertos presentes em anuários estatísticos da *British Petroleum* no seu relatório chamado de *BP Statistical Review of World Energy 2020*, utilizado também pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis no seu repositório de dados. Além disso, há pesquisa bibliográfica sobre análise, tendência, sazonalidade, estimação, diagnóstico e previsão com modelos de séries temporais. As simulações computacionais são executadas na linguagem de programação R, que apresenta bons recursos de manipulação de séries de tempo e bibliotecas para múltiplas funcionalidades, inclusive, previsão de dados. Dessa forma, muitos profissionais utilizam esta linguagem e suas bibliotecas de código aberto para trabalhos acadêmicos em ciência e educação envolvendo análise de dados.

REFERENCIAL TEÓRICO

O trabalho resume-se a análises de atuação no passado, presente e futuro de dados de séries produtivas do setor de petróleo e gás natural. Para o comportamento do passado e

presente utiliza-se a técnica de decomposição de séries temporais chamada STL (*Seasonal-Trend Decomposition Procedure Based on Loess*) e, para o futuro, técnica de previsão baseada em modelos ARIMA (*Autoregressive Integrated Moving Average Models*). Contudo, para melhor compreensão dessas análises é necessário algumas definições formais acerca da análise univariada de séries temporais.

Ao descrever séries temporais é comum a utilização de termos como tendência, sazonalidade e ciclo. Quando existe um acréscimo ou decréscimo de longo prazo nos dados que constituem a série de tempo, então existe tendência. Já o comportamento sazonal muitas vezes pode ser confundido com o comportamento cíclico, mas esses comportamentos são distintos. Quando fatores sazonais, de frequência fixa, ocorrerem nos dados, então há presença de sazonalidade. Ao passo que, se existirem flutuações com durações médias maiores e magnitudes variáveis, então existe um comportamento cíclico. Diante disso, ao escolher um método de previsão, é importante identificar os padrões em séries temporais (tendência, sazonalidade e ciclo) e, posteriormente, um mecanismo que seja eficiente em se ajustar apropriadamente a esses comportamentos.

Uma forma de verificação de padrões em séries temporais é pela sua decomposição. Usualmente é útil dividir a série de tempo nos três comportamentos, mas também é mais comum combinar a tendência e o ciclo em um único componente chamado ciclo-tendência (às vezes denominado somente tendência para simplificar). Dessa forma, é frequente pensar em uma sequência aditiva ou multiplicativa de séries temporais compreendendo três componentes: um componente de ciclo-tendência, outro sazonal e um remanescente ou de ruído (*remainder*) que contém qualquer outra coisa que apareça nos dados, que não foram contemplados nem pela tendência nem na sazonalidade.

Um procedimento para filtragem de séries temporais nesses componentes é o método STL. Esse método consiste em uma sequência de operações de suavização, cada uma delas, com uma exceção, emprega a mesma suavização: regressão localmente ponderada ou *loess*. Essa estratégia de decomposição é robusta e serve para uma ampla gama de dados com vantagens específicas. Em comparação com métodos clássicos de decomposição, a abordagem STL funciona para qualquer tipo de sazonalidade, pode ser robusta para dados aberrantes e personalizável na escolha das janelas de ciclo-tendência e sazonalidade (CLEVELAND et al., 1990).

Entendendo o comportamento passado e presente da série temporal pela análise de ciclo-tendência, sazonalidade e ruído, é possível então prever com maior clareza os

comportamentos futuros dos dados. Dessa forma surgem algumas estratégias e abordagens de previsão, como por exemplo a abordagem Box e Jenkins e os modelos ARIMA, que são relevantes para séries temporais não estacionárias (BOX et al., 2015).

Modelos ARIMA são utilizados para previsões em séries temporais. Estes modelos utilizam uma abordagem baseada na descrição das autocorrelações para ajustar modelos autoregressivos integrados de médias móveis, ARIMA(p,d,q), a um conjunto de dados. Onde **p** corresponde a ordem da parte autoregressiva, **d** é o grau de primeira diferenciação envolvida para transformar uma série temporal não estacionária em uma série estacionária e, **q** é a ordem da parte de média móvel do modelo. O modelo completo pode ser escrito como: $y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$, onde ϕ_p é a ordem dos pesos da parte autoregressiva, θ_q a ordem dos pesos da parte de média móvel e ε_t representa o ruído branco do processo estocástico. A constante é dada por $c = \mu(1 - \phi_1 - \dots - \phi_p)$, onde μ é a média de y'_t .

Existem alguns casos especiais de modelos ARIMA. Dentre esses casos, existe o ARIMA (0,0,0) que representa o ruído branco, o ARIMA(0,1,0) sem constante e com constante que representam, respectivamente, o passeio aleatório e o passeio aleatório com *drift*. Existem ainda os modelos formadores como o ARIMA(p,0,0) e ARIMA(0,0,q) que representam, respectivamente, o modelo autoregressivo AR(p) e o modelo de médias móveis MA(q).

Para modelagem ARIMA em linguagem R, a abordagem Box e Jenkins segue alguns passos. Primeiramente os dados devem passar por um teste de raiz unitária chamado KPSS (Kwiatkowski-Phillips-Schmidt-Shin) para retornar o número de diferenças **d** e tornar a série estacionária. Nem sempre diferenciações são suficientes para estabilizar a variância dos dados, sendo assim, recorre-se quando necessário, a uma transformação logarítmica chamada Box-Cox. Depois disso, as funções de autocorrelação (ACF) e autocorrelação parcial (PACF) dos dados são exibidas após a série ser diferenciada e esses correlogramas podem ajudar a determinar o valor de **p** ou **q** (HYNDMAN; ATHANASOPOULOS, 2018).

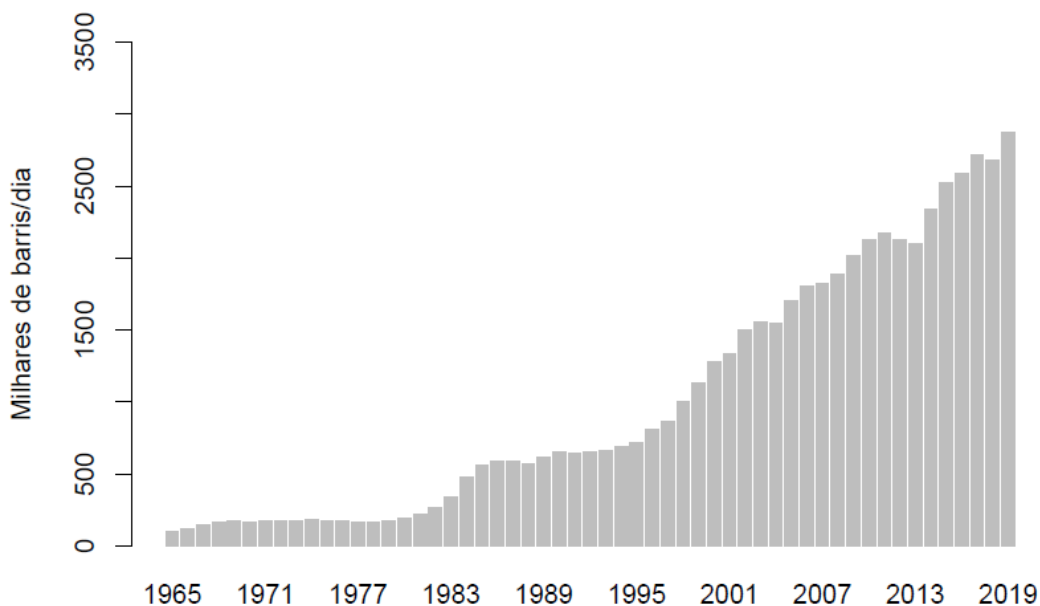
Os dados podem seguir modelos ARIMA(p,d,0) ou ARIMA(0,d,q) se os gráficos ACF e PACF seguirem um padrão em conjunto. Para o ARIMA(p,d,0), a ACF deve apresentar uma decadência senoidal ou exponencial e a PACF um pico significativo na defasagem **p**, mas nenhum pico fora do intervalo de confiança além desta defasagem. Logo, o comportamento inverso nos gráficos ACF e PACF servem para o modelo ARIMA(0,d,q). Se ambos os valores de **p** e **q** forem positivos, gráficos de ACF e PACF não ajudam a encontrar valores adequados de **p** e **q** (HYNDMAN; ATHANASOPOULOS, 2018).

Quando isso acontece, de ambos os valores de p e q serem positivos, critérios de informação serão utilizados para escolha do melhor modelo ARIMA. O critério de informação Akaike corrigido (AICc) que tiver menor valor simulado será reconhecido como o modelo escolhido para checar os resíduos. Verificar os resíduos do modelo escolhido é a última etapa de validação do modelo antes de utilizá-lo para fazer previsões, ou seja, um gráfico ACF dos resíduos deve ser exibido com todas as autocorrelações dentro dos intervalos de confiança, sugerindo um comportamento de ruído branco. Caso esses resíduos se pareçam com um ruído branco, pode-se fazer previsões com o modelo, pois ele se encontrará em equilíbrio estatístico e não carregará vieses para o futuro (HYNDMAN; ATHANASOPOULOS, 2018).

RESULTADOS E DISCUSSÃO

As séries temporais utilizadas para análise foram retiradas do relatório estatístico anual da *British Petroleum*. A primeira série temporal representada na Figura 01, demonstra a evolução produtiva em milhares de barris por dia da produção de petróleo no Brasil de 1965 a 2019.

Figura 01 - Produção de petróleo em milhares de barris/dia entre 1965 a 2019



Fonte: BP Statistical Review of World Energy 2020

Figura 02 – Decomposição STL (tendência e ruído) da produção de petróleo entre 1965 e 2019.

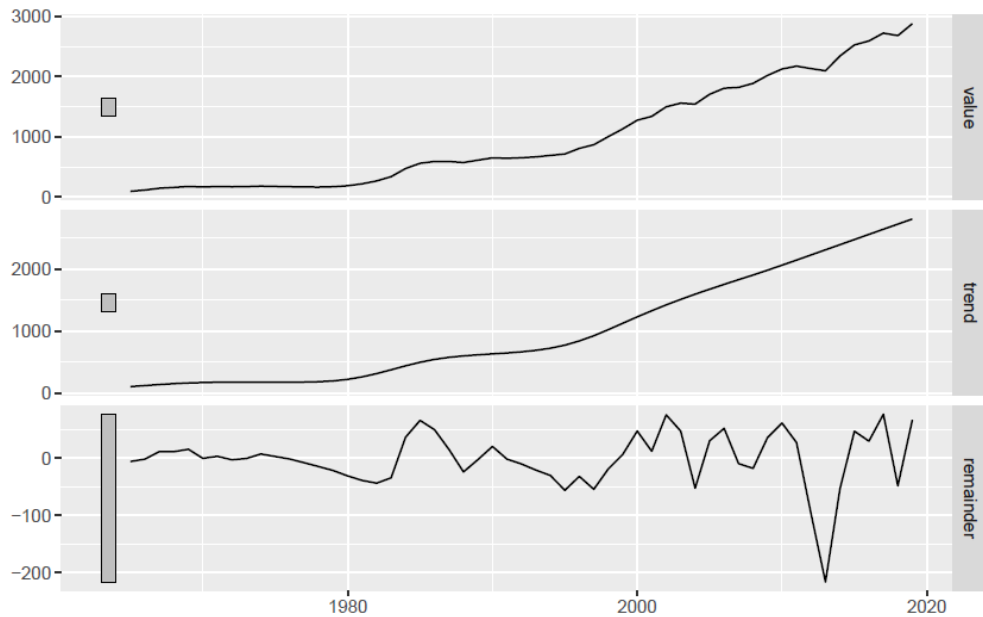
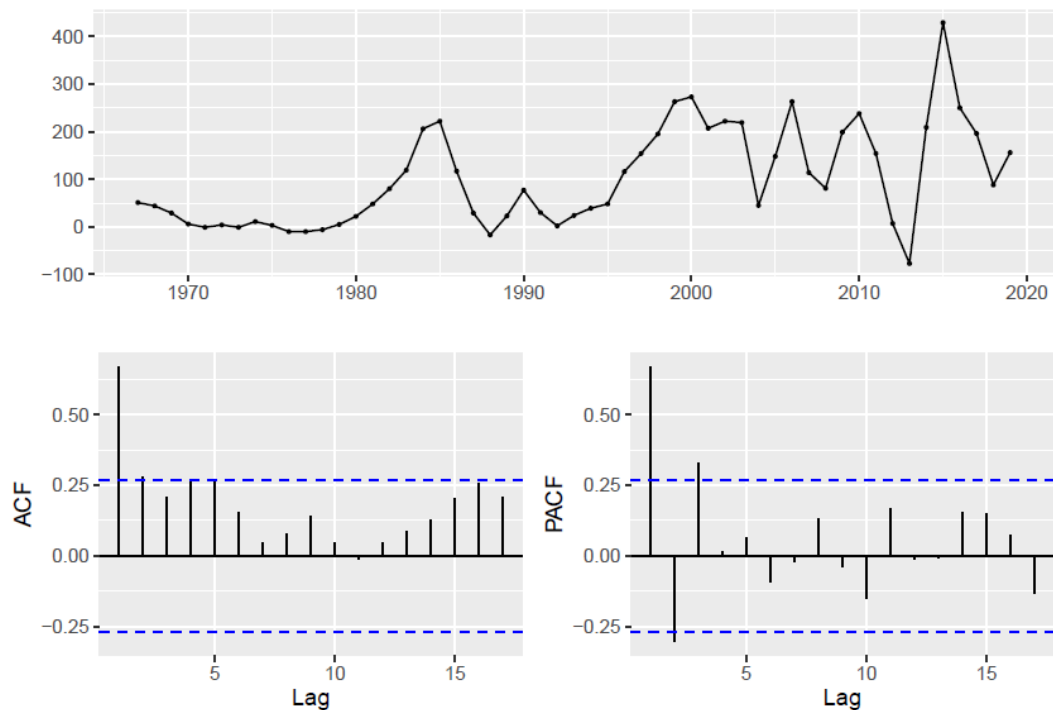


Figura 03 – Representação da série temporal diferenciada duas vezes com suas respectivas ACF e PACF.



Como as funções de autocorrelação e autocorrelação parcial indicam comportamento característico do modelo $ARIMA(p,d,0)$, então ajustou-se um $ARIMA(3,2,0)$ no entanto, o modelo que apresentou menor critério de informação $AICc = 593.88$ foi o $ARIMA(0,2,1)$.

Figura 04 – Representação dos resíduos do modelo e sua respectiva função de autocorrelação representando um ruído branco.

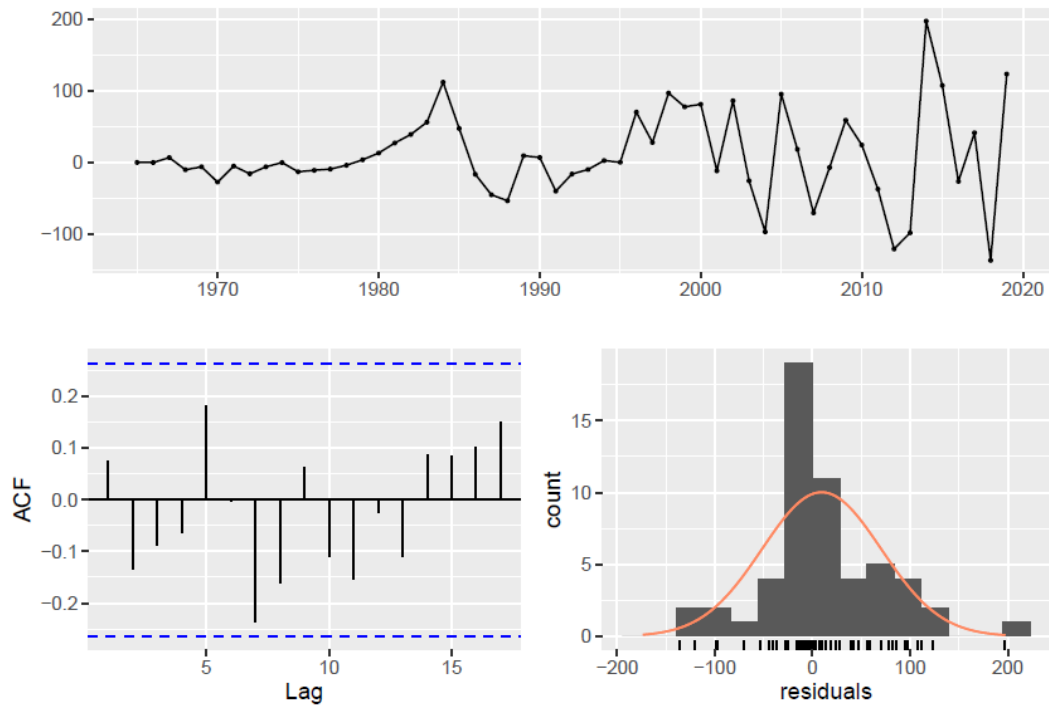


Figura 05 – Previsões em horizonte de 10 anos do ARIMA (0,2,1) e respectivos intervalos de confiança.

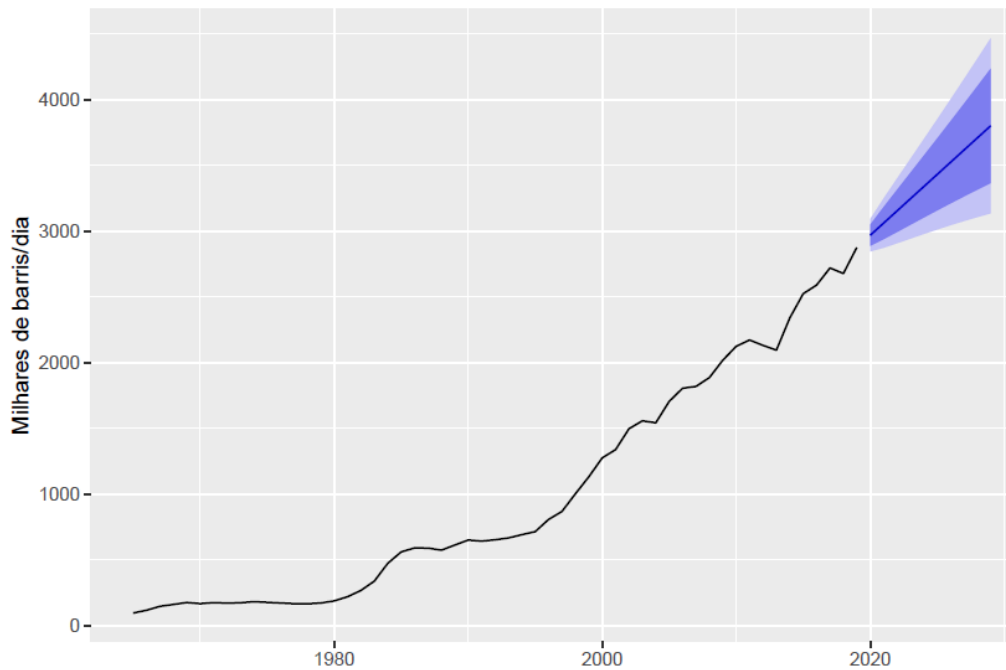
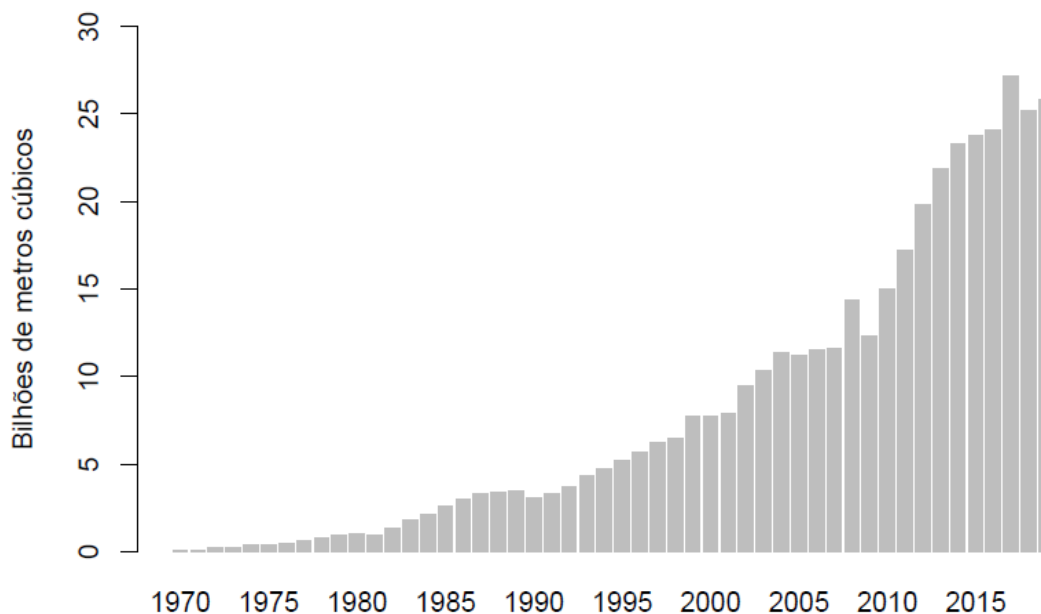


Tabela 01 – Resultados numéricos das previsões com ARIMA (0,2,1) e respectivos intervalos de confiança.

Ano	Pontos Previstos	Limite inferior (80 %)	Limite superior (80 %)	Limite inferior (95 %)	Limite Superior (95 %)
2020	2969.636	2889.078	3050.194	2846.433	3092.839
2021	3062.272	2939.715	3184.829	2874.838	3249.707
2022	3154.908	2994.066	3315.751	2908.921	3400.896
2023	3247.544	3049.255	3445.834	2944.287	3550.802
2024	3340.181	3104.291	3576.07	2979.419	3700.942
2025	3432.817	3158.729	3706.905	3013.635	3851.998
2026	3525.453	3212.341	3838.564	3046.59	4004.315
2027	3618.089	3265.008	3971.169	3078.099	4158.079
2028	3710.725	3316.664	4104.786	3108.061	4313.389
2029	3803.361	3367.274	4239.449	3136.423	4470.299

Já a segunda série, vista na Figura 06, demonstra a evolução de produção de gás natural em bilhões de metros cúbicos no Brasil no período de 1970 a 2019.

Figura 06 – Produção nacional de gás natural em bilhão de metros cúbicos entre 1970 e 2019



Fonte: BP Statistical Review of World Energy 2020

Figura 07 – Decomposição STL (tendência e ruído) da produção de gás natural entre 1970 a 2019.

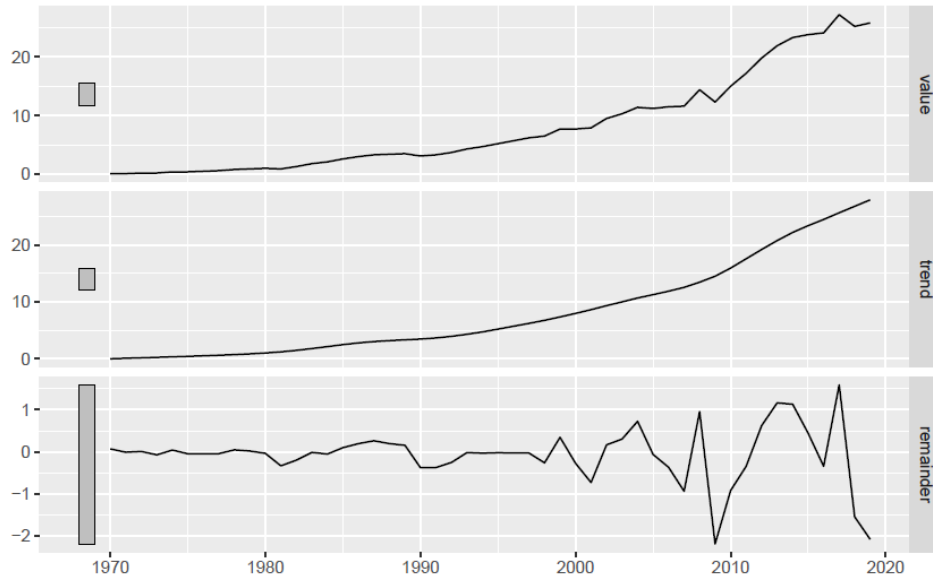
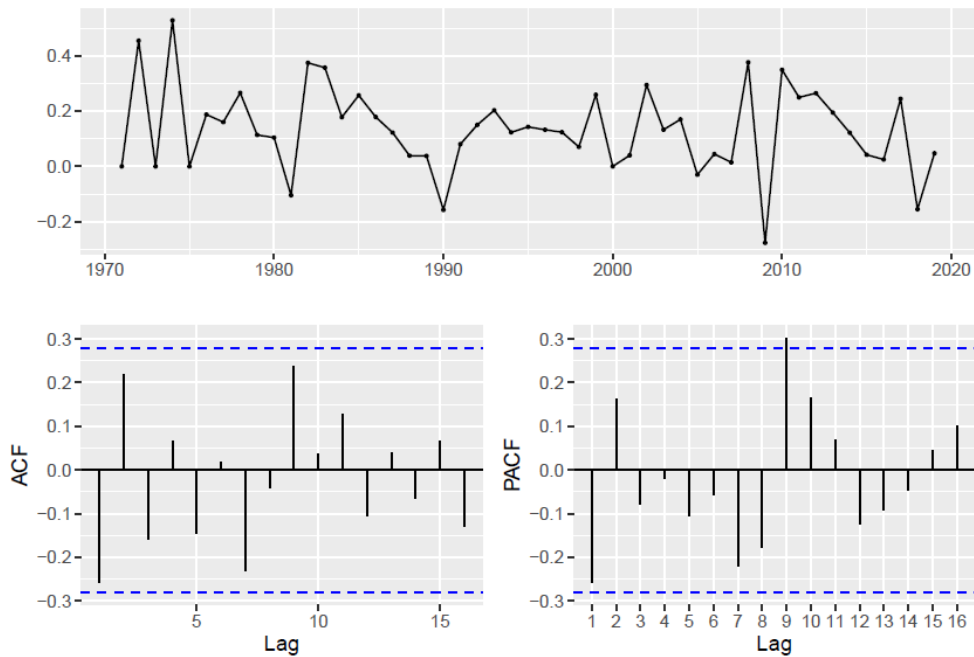


Figura 08 – Representação da série temporal diferenciada uma vez com suas respectivas ACF e PACF.



Como as funções de autocorrelação e autocorrelação parcial não indicam comportamento característico de modelos $ARIMA(p,d,0)$ ou $ARIMA(0,d,q)$, então ajustou-se um $ARIMA(1,1,0)$ com deslocamento (*drift*), e transformação Box-Cox com $\lambda = 0.216385$ para estabilizar a variância dos dados. Este modelo apresentou o menor critério de informação $AIC_c = -40.29$.

Figura 09 – Representação dos resíduos do modelo e sua respectiva função de autocorrelação comum ruído branco.

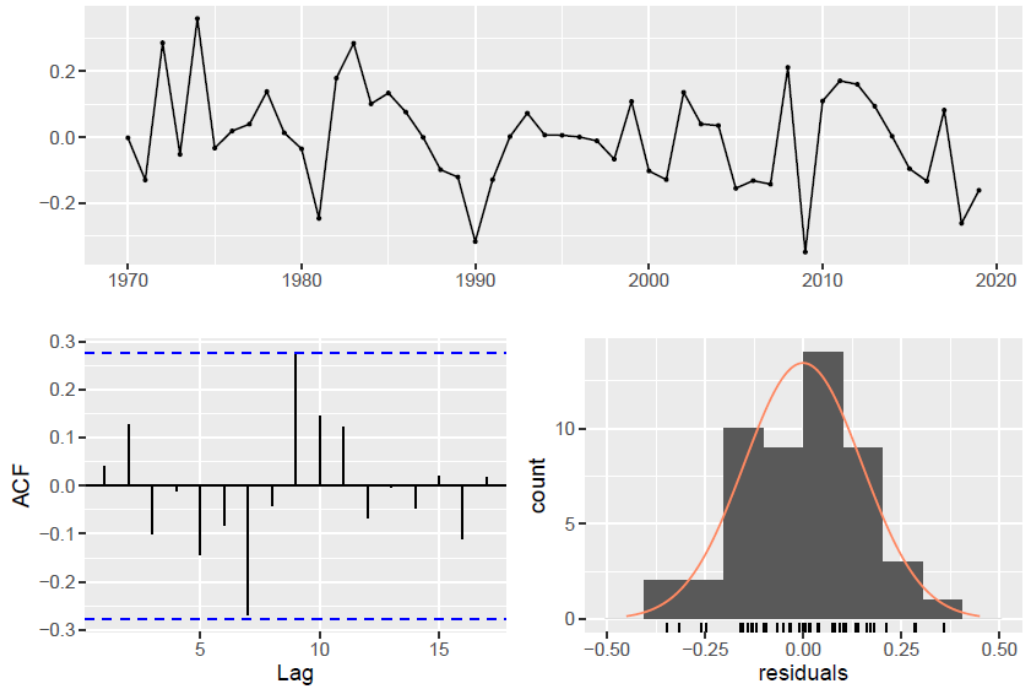


Figura 10 - Previsões para um horizonte de 10 anos do ARIMA (1,1,0) com deslocamento (*drift*) e respectivos intervalos de confiança.

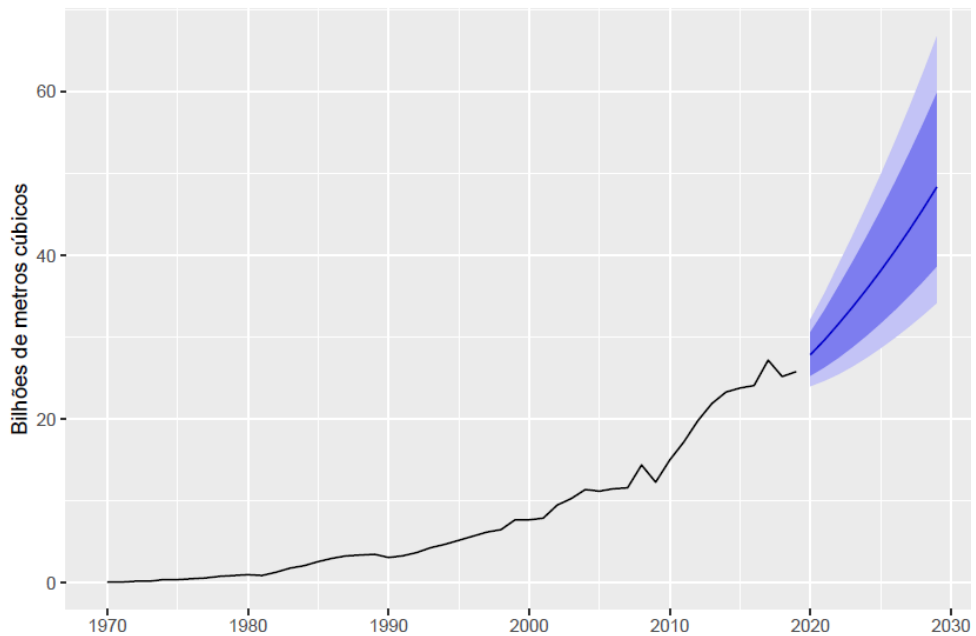


Tabela 02 – Resultados numéricos das previsões com ARIMA (1,1,0) com deslocamento (*drift*) e respectivos intervalos de confiança

Anos	Pontos Previstos	Limite inferior (80 %)	Limite Superior (80 %)	Limite inferior (95 %)	Limite superior (95 %)
2020	27.85956	25.29645	30.62186	24.01661	32.16849
2021	29.64463	26.32091	33.28859	24.68392	35.35461
2022	31.62585	27.4903	36.23377	25.4799	38.87712
2023	33.68022	28.78792	39.20129	26.43423	42.39766
2024	35.84559	30.19398	42.29443	27.49953	46.05754
2025	38.11644	31.69907	45.50993	28.66374	49.85392
2026	40.49907	33.29961	48.86484	29.91842	53.8099
2027	42.99672	34.99385	52.36756	31.25926	57.9367
2028	45.61353	36.78167	56.02689	32.68402	62.24575
2029	48.35351	38.6638	59.85048	34.19184	66.74677

CONSIDERAÇÕES FINAIS

De acordo com a decomposição das séries temporais, a produção de petróleo e gás natural apresentam zonas de crescimento e patamares de estagnação. A produção de petróleo tem seu primeiro patamar entre 1965 e 1982, onde a produção estabilizou-se em média em 170 mil barris/dia. O gás também segue estabilizado em 1 bilhão de m³ anuais até 1981, quando de 1982 a 1984 acontece a duplicação da capacidade produtiva do gás natural, estabilizando-se novamente em 1991 em 3.3 bilhões de metros cúbicos. Já o petróleo apresenta um pequeno crescimento entre 1983 e 1986 atingindo o nível de 339 mil barris/dia no ano, estabilizando novamente em 1996 com 667 mil barris/dia.

O gás apresenta boa taxa de crescimento na produção entre 1992 e 2008, atingindo 12.3 bilhões de metros cúbicos naquele ano. Ainda em 2010 acelera para uma taxa mais ascendente até que em 2019 chega ao nível de 25.8 bilhões. Algo parecido acontece com a produção de petróleo, mas em uma taxa de variação menor entre 1997 e 2019 e com um crescimento menos ascendente que o gás natural, chegando ao nível de produção de 2.87 milhão de barris/ dia.

É notável as características de crescimento e tendência ascendente na produção tanto de barris de petróleo como em metros cúbicos de gás natural. Além disso, pela análise das séries produtivas em questão, identifica-se através dos modelos preditivos que esta tendência nos dados deve continuar para o horizonte de 10 anos, com mais fôlego para a produção de gás natural do que para a produção de petróleo. Cabe ressaltar também que na decomposição das séries temporais o componente de ruído identificou variações pontuais na produção de

ambos os produtos, resultado de choques externos que impactaram isoladamente a produção de alguns anos.

Uma outra abordagem para a pesquisa é um comparativo mais aprofundado sobre outros métodos preditivos relacionados com as componentes de tendência, sazonalidade e aleatoriedade da série, como por exemplo, previsão com modelos de suavização exponencial. Ou ainda utilização de modelos que trabalhem melhor com a volatilidade dos dados para obter previsões e leitura de comportamentos mais precisos. Contudo, a decomposição STL é robusta e os modelos ARIMA respondem a uma grande diversidade de conjuntos de dados e mostram-se boas técnicas para as séries analisadas.

AGRADECIMENTOS

Agradecimentos ao orientador Professor Dr. Allan Kardec Barros Duailibe Filho, aos coautores e ao Programa Multidisciplinar de Formação de Recursos Humanos em áreas de interesse do setor do petróleo (PRH 54.1-ANP) pela oportunidade de pesquisa no setor de petróleo e gás natural.

REFERÊNCIAS

- CLEVELAND, R. B., CLEVELAND, W. S., MCRAE, J. E., TERPENNING, I. J.. STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1), 3–73, 1990.
- BOX, G. E. P., JENKINS, G. M., REINSEL, G. C., & LJUNG, G. M. **Time series analysis: Forecasting and control** (5th ed). Hoboken, New Jersey: John Wiley & Sons, 2015.
- COWPERTWAIT, P. S. P., METCALFE, A. V. **Introductory Time Series with R**. Springer Science + Business Media. New York: LLC, 2009.
- CRYER, J. D.; CHAN, K.-S. **Time Series Analysis With Applications in R**. Second Edition. Springer Science + Business Media. [S.l.]: LLC, 2008.
- HYNDMAN, R. J., ATHANASOPOULOS, G. **Forecasting: principles and practice**. OTexts, 2018.
- HYNDMAN, R. J., KHANDAKAR, Y. **Automatic time series forecasting : The forecast package for R**. *Journal of Statistical Software*, 27(1), 2008.
- LOONEY, B. et al. **Statistical Review of World Energy 2020**. BP Statistical Review, 69th edition, 2020.
- MORETTIN, P.; TOLOI, C. **Análise de séries temporais**. 2ª ed. São Paulo: Edgar Blucher, 2006.
- NEUSSER, K. **Time Series Econometrics**. Springer Texts in Business and Economics. Bern : Springer International Publishing Switzerland, 2016.