

## Modelagem estatística aplicada ao estudo de demanda e consumo energético.

## Statistical modelling applied to the study of energy demand and consumption.

### RESUMO

**Luís Felipe de Paula Miranda**  
[lfeliphemiranda@gmail.com](mailto:lfeliphemiranda@gmail.com)  
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

**Elisângela Aparecida da Silva Lizzi**  
[elisangelalizzi@gmail.com](mailto:elisangelalizzi@gmail.com)  
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

O papel da energia elétrica é notório no desenvolvimento técnico científico das sociedades contemporâneas. Entender as suas formas de obtenção e consumo é de suma importância para o gerenciamento energético inteligente. Neste estudo se empregada análise e modelagem estatística de séries temporais pois se trata de ferramenta analítica que permite estudar e entender o comportamento das séries de consumo energético brasileiro. O objetivo deste estudo é modelar séries de tempo relativas ao consumo mensal de demanda energética no território nacional, descentralizadas por unidades federativas. Os procedimentos utilizados foram baseados em métodos Box e Jenkins, passando-se pela etapa de suposição, obtenção dos parâmetros via estimação de máxima verossimilhança, escolha do modelo mais adequado aos dados com base no critério de Akaike, consolidação e diagnóstico do modelo obtido analisando ruído branco e, por fim fazendo projeções com o intuito de esboçar consumos futuros. Com os resultados obtidos pode-se fazer inferências sob o consumo energético nas unidades federativas do Brasil, servindo como ferramenta analítica na tomada de decisões no âmbito energético e implementação de políticas públicas para a prevenção de eventuais crises de abastecimento.

**PALAVRAS-CHAVE:** Demanda energética. Estatística aplicada. Série Temporal.

### ABSTRACT

**Recebido:** 19 ago. 2019.

**Aprovado:** 01 out. 2019.

**Direito autoral:** Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



The role of electric energy is notorious in the scientific and technical development of contemporary societies. Understanding their forms of procurement and consumption is of paramount importance to intelligent energy management. This study uses analysis and statistical modeling of time series because it is an analytical tool that allows studying and understanding the behavior of the series of Brazilian energy consumption. The objective of this study is to model time series related to the monthly consumption of energy demand in the national territory, decentralized by federal units. The procedures used were based on Box and Jenkins methods, going through the assumption step, obtaining the parameters through maximum likelihood estimation, choosing the most appropriate model to the data based on the Akaike criterion, consolidation and diagnosis of the model obtained by analyzing noise, and finally making projections in order to sketch future consumptions. With the results obtained, it is possible to make inferences about the energy consumption in the federative units of Brazil, serving as an analytical tool in the decision making in the energy field and the implementation of public policies to prevent eventual supply crises.

**KEYWORDS:** Energy demand. Applied statistical analysis. Time series.

## INTRODUÇÃO

O Brasil é um país em que a energia hidrelétrica é predominante (ANEEL, 2019) e que pleiteia grandes colocações mundiais no tocante a capacidade de geração. Entender as peculiaridades locais, consumo e geração permite projetar estratégias de prevenção e planos de ação para eventuais crises energéticas. A metodologia de modelos **Box e Jenkins (1976)** possibilita, por meio de modelagem e análise de séries temporais, estimar modelos matemáticos e estatísticos cujas projeções são factíveis. Então, neste trabalho utilizou-se modelos de séries temporais, especificamente métodos Box e Jenkins, para entender e modelar o consumo de energia elétrica brasileiro nas unidades federais, passando-se pela etapa de suposição, obtenção dos parâmetros via estimação de máxima verossimilhança, escolha do modelo mais adequado aos dados com base no critério de Akaike (AIC), consolidação e diagnóstico do modelo obtido via análise do ruído branco e, por fim fazendo projeções com o intuito de esboçar consumos futuros.

## MATERIAIS E MÉTODOS

O modelo ARIMA (Autorregressivo Integrado de Médias Móveis) é considerado um caso geral dos modelos **Box e Jenkins (1976)** que é apropriado para descrever séries não estacionárias. O modelo ARIMA permite o ajuste de modelos que não possuam componentes sazonais, mas possuam componente de tendência e quando essas componentes possuam comportamento não-estacionário. Geralmente, as séries encontradas têm tendência, sazonalidade e conjuntamente os dois. Segundo **Moretin e Toloi (2006)**, uma série temporal é composta pelos seguintes fatores: séries temporais ( $Z_t$ ) podem ser representadas por uma soma de componentes, sendo elas: tendência ( $T_t$ ), sazonalidade ( $S_t$ ) e o erro aleatório ( $\varepsilon_t$ ). O teste de Cox-Stuart (**Moretin e Toloi, 2006**) tem como finalidade verificar se existe alguma tendência na série e para analisar se há uma componente sazonal utilizou-se uma função de autocorrelação.

O modelo ARIMA (p,d,q) pode ser representado por:

$$\varphi(B)(1 - B)^d Z_t = \theta(B)a_t \quad (1)$$

$$\varphi(B) = 1 - \varphi_1 B^1 - \varphi_2 B^2 - \varphi_3 B^3 - \dots - \varphi_p B^p \quad (2)$$

Na equação (2) tem-se a ordem do polinômio autorregressivo que é representado por  $p$ ,  $B$  é o operador de retardo de modo que  $B_j Z_t = Z_{t-j}$  e  $d$  o número de diferenças que são necessárias para tornar a série estacionária, ou seja, retirar sua tendência;

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \dots - \theta_q B^q \quad (3)$$

Na equação (3) relaciona-se a ordem  $q$  do polinômio de médias móveis, como a série em estudo tem frequência mensal, logo este componente pode ser entendido como à influência do consumo energético dos meses anteriores nos meses atuais.

Composta pelo ciclo de três etapas sequenciais, a metodologia aplicada de Box-Jenkins consiste em: (1) identificação dos modelos; (2) estimativas de parâmetros; (3) diagnóstico do modelo obtido.

Na etapa de identificação dos modelos, os dados devem ser pré processados para a série, obrigatoriamente ser estacionária, deste modo se utiliza diferenciação da série. Por conseguinte, mediante as observações das funções de autocorrelação (FAC) e auto correlação parcial (FACP), seleciona-se o número de parâmetros autorregressivos e médias móveis que podem ser incorporados aos modelos, e na sequência os parâmetros são estimados via métodos de maximaverossimilhança. Para verificar se o modelo é apropriado e demonstra confiabilidade, na fase de diagnóstico verifica-se a sua adequação e se realiza a análise dos resíduos os quais permitem verificar seu comportamento como o de um ruído branco ( $\varepsilon_t \sim N(\mu_\varepsilon; \sigma_\varepsilon)$ , onde  $\lim_{n \rightarrow \infty} \mu_\varepsilon = 0$  e  $\sigma_\varepsilon$  deve ser constante)

Neste trabalho para discriminar o melhor modelo obtido optou-se por utilizar o critério de informação de Akaike (AIC), definido por:

$$AIC = -2\log(L) + 2m \quad (4)$$

sendo  $L$  função de verossimilhança do modelo ARIMA e  $m$  o número de parâmetros do modelo (EHLERS, 2007). Dado que cada modelo tem seu valor de AIC, aquele que o tiver com menor valor deve ser escolhido. Quanto menores estes valores, os modelos mais próximos à realidade são, ou que têm menos perda de informação em relação à realidade (GUJARATI, 2006). Como medidas de comparação, utilizou-se o erro médio percentual absoluto (MAPE) o qual permite verificar a aderência dos modelos estimados. Um valor de  $MAPE$  menor igual a 15% significa uma boa aderência, ajuste do modelo e também indica boas projeções *out of sample*. O mesmo é calculado pela seguinte relação:

$$MAPE = \left\{ \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \right\} \cdot 100 \quad (5)$$

Onde  $\hat{y}$  é a série estimada,  $y$  a série original e  $n$  o número de observações. Logo, o MAPE é uma validação da projeção dos resultados estimados pelo modelo.

Os dados avaliados foram retirados por meio de bancos de dados disponibilizados pela **Agência Nacional de Energia Elétrica (ANEEL)** e **Empresa de Pesquisas Energéticas (epe)**. Tais dados referem-se ao consumo energético mensal na rede das Unidades Federativas (UF) do Brasil. No período de janeiro de 2004 a junho de 2018.

Vale ressaltar neste texto que somente serão ilustrados e mostrados os resultados dos estados de São Paulo, Paraná, Minas Gerais e Rio Janeiro, pois não há espaço suficiente para discorrer e explorar os resultados das vinte sete unidades federativas nesse resumo.

## RESULTADOS E DISCUSSÕES

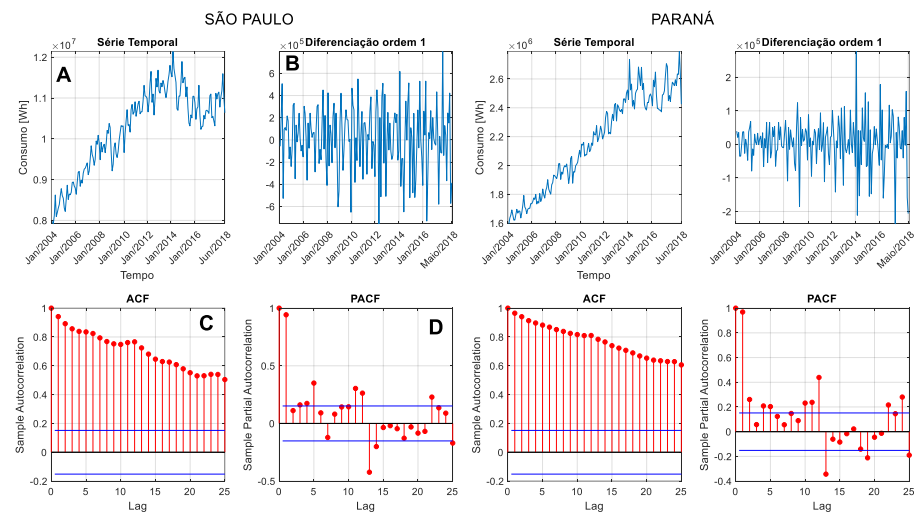
Analisados os dados, testou-se o modelo  $ARIMA(p, d, q)$  para as vinte e sete Unidades Federativas (UF) do Brasil, ou seja, implementou-se um modelo para cada UF. Na etapa de diferenciação, as séries temporais mostraram-se estabilizadas com componente de diferenciação  $d = 1$ , para os parâmetros de autorregressivos ( $p$ ) e de média móvel ( $q$ ), pode-se estabelecer os seguintes intervalos  $\{p, q \in N \mid 1 \leq p \leq 5\}$  e  $\{1 \leq q \leq 10\}$ , respectivamente.

Nas Figuras 1 e 2 se faz possível a visualização no painel (A) dos gráficos de comportamento das séries em um período de janeiro de 2004 a junho de 2018

para o consumo energético, onde evidencia-se que todos os gráficos se mostram com comportamento não estacionário. No painel (B) tem-se o gráfico da série após sua primeira diferenciação,  $d = 1$ , com isto é possível perceber que a série se torna estacionária e possibilita a implementação dos métodos de Box e Jenkins. No painel (C) tem-se o gráfico de autocorrelação e no painel (D) o gráfico de autocorrelação parcial.

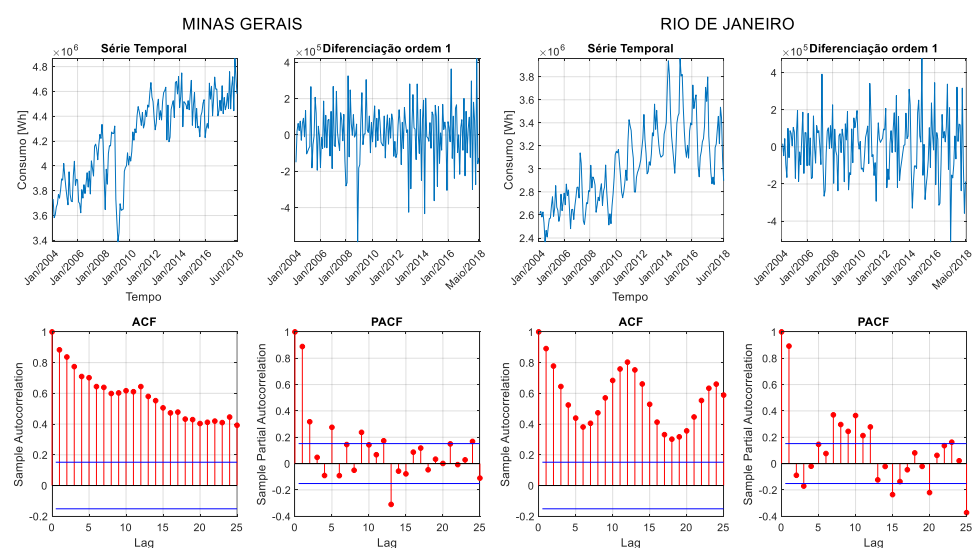
Estes gráficos são essenciais e norteadores para que seja possível escolher os melhores valores elegíveis de componentes autorregressivos e de média móvel. Para São Paulo, Figura 1, é possível ver no gráfico de ACF um padrão de decaimento lento, isto é, indica-se que é necessário altos valores no componente autorregressivo para que seja capaz de interpolar uma boa função nos dados. Já em relação ao gráfico PACF é possível verificar valores de médias móveis, com isso inicia-se, pelo processo, para os outros estados a interpretação é análoga.

Figura 1 – Análise da série temporal em São Paulo e Paraná.



Fonte: Autoria própria (2019).

Figura 2 – Análise da série temporal em Minas Gerais e Rio de Janeiro.



Fonte: Autoria própria (2019).

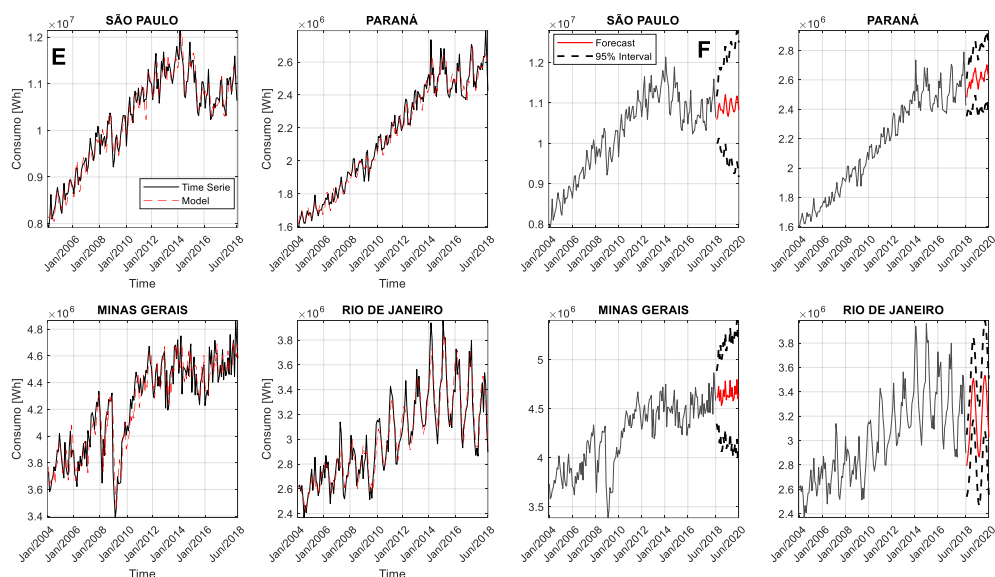
Na tabela 1, segue-se os modelos obtidos com os menores valores de AIC para os estados de São Paulo, Paraná, Minas Gerais e Rio de Janeiro. Vale salientar que para cada estado foi feito todas as etapas de modelagem, e os resultados exibidos na tabela 1, já são o resultado do modelo com melhor ajuste e adequação segundo os critérios descritos nos métodos.

Tabela 1 – Modelos ARIMA(p,d,q) melhores ajustados e critério AIC para SP, PR, MG e RJ.

UF	Modelo Ajustado	AIC
São Paulo	ARIMA(10,1,4)	4763,53
Paraná	ARIMA(9,1,5)	4309,74
Minas Gerais	ARIMA(9,1,5)	4556,56
Rio de Janeiro	ARIMA(6,1,2)	4546,72

Fonte: Autoria própria (2019).

Figura 3 – Ajuste dos modelo para SP, PR, MG e RJ e previsão obtida com as bandas de confiança de 95%.



Fonte: Autoria própria (2019).

Na Figura 3 é possível observar no painel (E) os valores reais sobre o perfilamento do modelo ajustado (em vermelho), deste modo é possível visualizar que todos os modelos gerados tem ótimo ajuste e seguem a tendência da série original, evidenciando a ótima qualidade de ajuste destes métodos ao problema proposto. Observa-se no painel (F) os valores da previsão e seus respectivos intervalos de confiança com 95% de confiança, geralmente esses modelos são utilizados por no máximo 24 meses, depois disso precisam de uma atualização para gerar boas previsões. Para validar o resultado das previsões, foram comparados os valores previstos com os valores originais dos últimos seis meses que estão disponíveis pela ANEEL utilizando o MAPE, conforme Tabela 2.

Tabela 2 – Comparação das projeções e dos valores reais para o período de julho a dezembro de 2018 utilizando o critério MAPE.

UF	Critério MAPE (6 meses)
São Paulo	2,04 %
Paraná	1,97 %
Minas Gerais	2,31 %
Rio de Janeiro	3,51 %

Fonte: Autoria própria (2019).

A Tabela 2 exibe os resultados obtidos pela análise de predição dos modelos obtidos, ou seja, compara-se com os dados originais para o entendimento da aderência do modelo e se o mesmo pode ser utilizado para futuras tomadas de decisões relacionado a administração de tais informações. Para as demais UF utilizou-se igual método.

## CONCLUSÃO

Pode-se concluir que os métodos de modelagem estatística no âmbito de séries temporais são adequados porquanto permitem subsidiar planejamento de consumo e demanda energética nas UF do Brasil. É importante salientar que os resultados obtidos foram compilados pelas UF, em razão de saber-se que a política pública energética é descentralizada entre união, estado e município. Consequentemente, apresentando os resultados por estado é possível indicar os melhores modelos para cada estado e deste modo possibilitar estratégias de vigilância e de políticas públicas de tais ações para cada estado onde comiserar-se-á as particularidades dos mesmos.

## REFERÊNCIAS

### ANEEL. MATRIZ DE ENERGIA ELÉTRICA.

C. A. J. Schmidt e M.A.M. Lima: **A demanda por energia elétrica no Brasil**. Revista Brasileira de Economia, 58:67–98, 2004.

GUJARATI, D. **Econometria básica**. Rio de Janeiro: Elsevier, 2006. 812 p.

G. P. BOX e G.M. JENKINS. **Time series analysis, forecasting and control**. San Francisco: Holden-Day, 1976.

P.A. MORRETIN E C.M.C TOLOI. **Análise de séries temporais**. São Paulo: Edgard Blucher, 2006.

PIRES, J. C. L.; GOSTKORZEWICZ, J.; GIAMBIAGI, F. **O cenário macroeconômico e as condições de oferta de energia elétrica no Brasil**. Rio de Janeiro: BNDES, mar. 2001.

R.S. EHLERS. **Análise de séries temporais**. Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná.

## AGRADECIMENTOS

A UTFPR pelo aporte fornecido durante o período de pesquisa e bolsa pesquisa concedida, a Profa. Dra. Elisângela Aparecida da Silva Lizzi pela oportunidade e disponibilidade de passar seu conhecimento e aos meus pais pela confiança e empenho no que tange a educação.