

Análise de séries temporais aplicadas a furtos e fraudes de energia elétrica

Marcelle Bezerra (UFPR) marcellezacarias@gmail.com
Angelo Cabral (UFPR) angeloscabral@gmail.com
Anelize Clair Flor (UFPR) anelizeflor@yahoo.com.br
Sonia Isoldi Marty Gama Müller (UFPR) soniaisoldi@ufpr.br
Débora Morales (UFPR) debora_mo2@hotmail.com

Resumo:

A energia elétrica é preponderante na qualidade de vida da sociedade bem como, para o desenvolvimento e crescimento industrial. Desta forma, a concessionária de distribuição de energia possui a responsabilidade, definida pela ANEEL, de combater as perdas não-técnicas, para que o valor referente à diminuição dessas perdas seja aplicado em melhorias e expansão das redes de distribuição e subestações, assim como na modicidade tarifária para os seus clientes. Este estudo buscou encontrar um modelo da análise de série temporal, para que as concessionárias de distribuição de energia tenham uma previsão do valor das perdas não-técnicas. A partir do banco de dados da ANEEL, procurou-se avaliar os métodos automáticos Brown e Holt-Winters e não-automático SARIMA de série temporal para a previsão. Com os resultados obtidos no presente estudo, verificou-se que nenhum dos métodos teve resultado satisfatório e nenhum se destacou frente às opções, podendo se afirmar que entre os métodos testados não foi encontrado nenhum modelo adequado para fazer previsões significativas das perdas não-técnicas de energia. Portanto, recomenda-se a análise de outros métodos, como as redes neurais artificiais, que tem apresentado bons resultados na academia.

Palavras chave: Perda não-técnica, Série temporal, SARIMA.

Time series analysis applied to electricity theft and fraud

Abstract

Electrical energy is predominant in Brazilian society's quality of life as well as for the development and industrial growth. Thus, the energy distribution company has a responsibility to combat non-technical losses, defined by ANEEL, so that the income recovered to decrease these losses are applied in improvements and expansion of distribution networks and substations, as well as low tariffs for their customers. This study seeks a methodology to find a model of the time series analysis for the energy distribution companies have an estimate of the value of the non-technical losses. From the database of ANEEL, we tried to evaluate the automated and non-automated methods of time series for the forecast. With the results obtained in this study, it was found that none of the methods had a satisfactory result and none of that stood out from options, we can say that among the tested methods is not found a suitable model to make meaningful forecasts of non-technical losses of

energy. Therefore, we recommend analyzing other methods, like artificial neural networks which has shown good result in academy.

Key-words: Non-technical lost, Time series, SARIMA.

1. Introdução

1.1 Energia elétrica

Na era moderna, dois fatores foram determinantes para o avanço da humanidade, a Revolução Industrial e a descoberta da maior fonte de energia, o petróleo. Esses dois acontecimentos selaram a dependência energética das sociedades, cujo progresso material pode ser medido pelo seu consumo de energia (GOLDEMBERG, 1979).

A Revolução Industrial caracterizou-se pela substituição da força laboral pelas máquinas, esta evolução é associada a uma sucessão de combustíveis fósseis (carvão, petróleo, gás natural) e a conversão de combustíveis e energia hidráulica em uma nova fonte de energia, a eletricidade.

A energia elétrica, além de ser facilmente transportável, podendo ser produzida no local mais conveniente e transmitida para consumidores distantes por uma rede elétrica, é eficientemente transformável em outras formas de energia, como, por exemplo, calor, luz, movimento. Ela é fundamental para a ocorrência de muitos fenômenos físicos e químicos necessários para operar diversos equipamentos.

1.1.2 Perdas globais de energia elétrica

As perdas globais são definidas por Penin (2008) e a Agência Nacional de Energia Elétrica - ANEEL (2008) como a diferença entre a energia injetada na rede elétrica e a energia faturada/medida nessa mesma rede, estas podem ser divididas em dois tipos, as perdas técnicas e as perdas não-técnicas. As perdas técnicas ocorrem naturalmente nos sistemas elétricos, causadas por ações internas nos materiais e consistem, principalmente, na dissipação de energia nos diversos equipamentos que compõem a rede elétrica de energia. As perdas técnicas podem ter seus valores estimados por modelos matemáticos, considerando o fluxo de potência, a carga e as características do sistema elétrico.

Já as perdas não-técnicas, ou perdas comerciais, são causadas pela falta de faturamento da energia consumida pelos usuários. Elas ocorrem pela deficiência em equipamentos de medição, falhas no cadastro ou no faturamento, furto de energia ou fraudes nos equipamentos de medição. É possível fazer o cálculo das perdas não-técnicas, identificando as perdas globais e subtraindo dela o valor referente às perdas técnicas. A diferença é referente às perdas comerciais ou não-técnicas (PENIN, 2008; ANEEL, 2008).

Segundo a ANEEL em 2010 constatou-se que os consumidores que pagam regularmente pela energia que consomem, arcaram no ano passado com uma conta de quase R\$8,1 bilhões das companhias brasileiras de energia elétrica que deixaram de receber este valor em 2009 por causa de ligações clandestinas na rede e fraudes na medição do consumo de energia elétrica.

Assim sendo, o objetivo deste artigo é, a partir do banco de dados disponibilizado pela ANEEL, analisar os dados de distribuidoras das cinco regiões brasileiras, buscando o melhor método de previsão entre os automáticos e não-automáticos em relação aos furtos e fraudes de energia elétrica destas distribuidoras.

Com esta análise é possível mostrar a tendência dos valores das fraudes e furtos de energia elétrica nas distribuidoras das regiões brasileiras, possibilitando utilizar o resultado na busca de estratégias para a redução dos valores, para que o ganho financeiro venha a ser aplicado na

melhoria das redes elétricas e, consequentemente, aumento na qualidade da energia elétrica fornecida para as indústrias, comércios, residências, etc.

2. Séries Temporais

Para Ehlers (2009), uma série temporal é uma coleção de observações feitas sequencialmente ao longo do tempo. A característica mais importante deste tipo de dados é que as observações vizinhas são dependentes e o interesse maior é analisar e modelar esta dependência.

De acordo com Granger (1988), o objetivo inicial da análise de séries temporais é a realização de inferências sobre as propriedades ou características básicas do mecanismo gerador do processo estocástico das observações da série.

Assim, através da análise das regularidades contidas nos fenômenos observáveis de uma série temporal, existe a possibilidade de se construir um modelo estatístico como uma representação simplificada da realidade.

Aquelas metodologias na qual a série de interesse é explicada não só pelo seu passado histórico, como também por outras séries temporais não-correlatas entre si, são conhecidas como funções de transferência. Esta classe de métodos de previsão envolve, portanto, mais de uma série temporal, com a ressalva de que a relação de causalidade entre estas séries é perfeitamente conhecida.

Os métodos multivariados abrangem os procedimentos de previsão que associam mais de uma série temporal na efetivação de prognósticos sem, no entanto, qualquer imposição com relação à causalidade entre essas séries.

2.1 Métodos

2.1.1 Método de Brown

Segundo Chaves Neto (2014), quando o processo de amortecimento no tempo apresenta uma tendência linear, então, aplicando-se inicialmente o AES (alisamento exponencial simples) aos dados, o amortecimento exponencial à saída do primeiro amortecimento já aplicado, assim, o valor futuro $E[Z_t]$ será:

$$E[Z_t] = 2E[\hat{\mu}_T] - E[\hat{\mu}_T^{[2]}]$$

$$\hat{Z}_T = 2\hat{\mu}_T - \hat{\mu}_T^{[2]}$$

A previsão h períodos a frente é feito usando-se a equação:

$$\hat{Z}_T(h) = (2 + h \frac{\alpha}{1-\alpha})\hat{\mu}_T - (1 + h \frac{\alpha}{1-\alpha})\hat{\mu}_T^{[2]}$$

$$\hat{\beta}_0(T) = 2\hat{\mu}_T - \hat{\mu}_T^{[2]} - T[\frac{\alpha}{1-\alpha}(\hat{\mu}_T - \hat{\mu}_T^{[2]})]$$

No amortecimento duplamente exponencial de Brown, há necessidade de valores iniciais para $\hat{\mu}_0$ e $\hat{\mu}_0^{[2]}$. O usual é fazer estimativas para β_0 e β_1 através de um ajuste linear simples e entrando com esses valores em:

$$\hat{\beta}_1(T) = \frac{\alpha}{1-\alpha}(\hat{\mu}_T - \hat{\mu}_T^{[2]})$$

$$\hat{\beta}_0(T) = 2\hat{\mu}_T - \hat{\mu}_T - T[\frac{\alpha}{1-\alpha}(\hat{\mu}_T - \hat{\mu}_T^{[2]})]$$

$$\hat{\mu}_0 = \hat{\beta}_0(0) - \frac{1-\alpha}{\alpha}\hat{\beta}_1(0), \hat{\mu}_0^{[2]} = \hat{\beta}_0(0) - 2\frac{1-\alpha}{\alpha}\hat{\beta}_1(0)$$

2.1.2 Método de Holt-Winters

Os modelos de Holt-Winters são muito utilizados quando da existência de uma série temporal que apresente, além da tendência, um componente de sazonalidade. Uma série com esse componente é caracterizada pela ocorrência de padrões cíclicos de variação, que se repetem

em intervalos relativamente constantes de tempo, são muito observadas em indústrias do ramo alimentícias, de vestuários, cosméticos, entre outras. Os modelos de Holt-Winters são classificados em dois grupos: aditivo e multiplicativo, no modelo aditivo, a amplitude da variação sazonal é constante ao longo do tempo, ou seja, a diferença entre o maior e menor valor de demanda dentro das estações permanece relativamente constante no tempo. No modelo multiplicativo, a amplitude da variação sazonal aumenta ou diminui como função do tempo (MORETTIN; TOLOI, 2006).

2.1.2.1 Modelo sazonal multiplicativo

O modelo multiplicativo de Holt-Winters se ajusta, de maneira mais adequada, a séries com tendência e sazonalidade multiplicativa, ou seja, àquelas em que a amplitude da variação sazonal aumenta com o acréscimo no nível médio da série temporal (KOEHLER et al., 2001, p.269). Seu algoritmo de previsão utiliza as seguintes expressões:

$$L_t = \alpha \frac{x_t}{S_{t-s}} + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

$$S_t = \gamma \frac{x_t}{L_t} + (1 - \gamma)S_{t-s}$$

$$\hat{x}_{t+m} = (L_t + mT_t)S_{t-s+m}$$

Nas expressões acima, $\hat{x}_{t+m}$ é a previsão para o período $t + m$, enquanto α, β e γ são constantes de suavização, cujos valores encontram-se entre 0 e 1, e x_t é a mais recente observação. Nelas, L_t representa uma estimativa do nível da série no tempo t , T_t , uma estimativa da declividade da série no mesmo período t e S_t , o componente de sazonalidade também no período t . A sazonalidade, ou seja, o número de subperíodos por ano é representado por s . Assim como os modelos descritos anteriormente, o método multiplicativo de Holt-Winters funciona através da aplicação recursiva de suas equações aos dados da série.

Dessa forma, tal aplicação deve iniciar em algum período no passado, onde os valores de L_t, T_t e S_t devem ser estimados (MAKRIDAKIS et al., 1998, p.168). Uma maneira simples de se fazer essa estimativa é através da inicialização do nível e da tendência no mesmo período s . O nível é determinado através da média de primeira estação. Para se inicializar a tendência, é recomendado o uso de duas estações completas, ou seja, $2s$ períodos e por último, os índices sazonais iniciais podem ser determinados através da razão entre as primeiras observações com a média do primeiro ano.

2.1.2.2 Modelo sazonal aditivo

O modelo aditivo de Holt-Winters apresenta maior capacidade de explicação em séries que possuem tendência e sazonalidade aditiva. Ou seja, ele é utilizado naquelas onde o efeito sazonal não é função do nível médio corrente da série temporal e pode ser adicionado ou subtraído de uma previsão que dependa apenas de nível e tendência (KOEHLER et al., 2001, p. 269). Seu algoritmo de previsão é baseado nas seguintes expressões:

$$L_t = \alpha(x_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

$$S_t = \gamma(x_t - L_t) + (1 - \gamma)S_{t-s}$$

$$\hat{x}_{t+m} = L_t + mT_t + S_{t-s+m}$$

A equação T_t é idêntica à equação do modelo multiplicativo de Winters, o que difere os dois modelos é o fato das outras equações, agora, apresentarem os índices de sazonalidade somados e subtraídos, ao invés de multiplicados e divididos (MAKRIDAKIS et al., 1998). As inicializações de L_s e T_s são idênticas às do modelo multiplicativo.

2.1.3 Método de Box & Jenkins

Box & Jenkins propuseram que um processo estocástico estacionário, por possuir média, variância invariantes no tempo e autocorrelação que depende apenas do intervalo de defasagens (lags), pode ser representado por um modelo auto regressivo e/ou médias móveis – ARMA (modelo auto regressivo de médias móveis) (p,q). – obtido por intermédio da passagem de um processo estocástico tipo ruído branco $a_t \sim N(0, \sigma_a^2)$ por um filtro linear, o que significa que a série resultante poderá ser vista como uma combinação linear dos termos da série original (ruído branco).

A metodologia de Box & Jenkins consiste na busca de um modelo ARIMA (modelo auto regressivo integrado de médias móveis) que represente o processo estocástico gerador da série temporal, a partir de um modelo ARMA aplicável na descrição de séries temporais estacionárias, estendendo esse conceito para séries temporais não-estacionárias (NELSON, 1975).

A identificação da estrutura do processo ARIMA (p,d,q) é uma das fases mais complicadas da metodologia. O objetivo da identificação é determinar os valores de p, d e q do modelo ARIMA (p, d, q), além de estimativas preliminares dos parâmetros a serem usadas no estágio de estimação. Esta escolha é feita com base nas autocorrelações estimadas e autocorrelações parciais estimadas e em critérios de informações. Após a identificação do modelo que seja uma representação adequada do mecanismo gerador da série, a estimação dos parâmetros desse modelo é efetuada. Os parâmetros do processo AR são estimados através de métodos de regressão; caso o processo MA esteja envolvido, a estimação dos parâmetros deste modelo é obtida pela aplicação de algum algoritmo de otimização não-linear.

2.1.3.1 Modelos auto regressivos (AR)

Introduzido por G. U. Yule (YULE, 1926), um modelo auto regressivo de ordem p, denotado por AR (p) é dado pela expressão:

$$Z_t = \delta + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t$$

$$Z_t - \delta - \phi_1 Z_{t-1} - \phi_2 Z_{t-2} - \dots - \phi_p Z_{t-p} = a_t \Rightarrow Z_t - \phi_1 B Z_t - \dots - \phi_p B^p Z_t = a_t, \delta = 0$$

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Z_t = a_t \Rightarrow \phi(B) Z_t = a_t, \delta = 0$$

Sendo que o polinômio característico é:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

Considerando-se o modelo AR (p) de termo constante δ dado por:

$$Z_t = \delta + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p}$$

$$\delta = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p), t = 1, 2, 3, \dots$$

E a média do processo dada por:

$$\mu = \frac{\delta}{(1 - \phi_1 - \phi_2 - \dots - \phi_p)}$$

2.1.3.2 Modelos de médias móveis (MA)

Os modelos de médias móveis foram introduzidos por Eugen Slutsky (SLUTZKY, 1937). Considere o modelo linear:

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

E admita que $\psi_j = 0$ para $j > q$. Obtém-se com isso um processo denominado de médias móveis (Moving Average) de ordem q, MA(q) que denota-se:

$$Z_t = \delta_t + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

2.1.3.3 Modelos auto regressivos de médias móveis (ARMA)

Herman O. A. Wold (WOLD, 1938) mostrou que qualquer processo estocástico estacionário discreto pode ser representado por modelos auto regressivos e de médias móveis ARMA. Os modelos que utilizam a estrutura ARMA possuem uma parte auto regressiva e outra de médias móveis e costumam ser mais parcimoniosos. No caso da estrutura ARMA (p,q), ou seja, quando se tem a ordem p para auto regressiva e a q para médias móveis, tem-se os modelos da forma:

$$Z_t = \delta_t + \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}$$

Com os polinômios correspondentes aos operadores auto regressivos e de médias móveis, a expressão acima pode ser escrita na forma:

$$(\phi)Z_t = \delta_t + \theta(B)a_t$$

Analisando a estacionariedade e a invertibilidade do processo ARMA (p,q), tem-se que Z_t pode ser escrito na forma de médias móveis de ordem infinita:

$$Z_t = \psi(B)a_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$$

E na forma auto regressiva infinita:

$$\pi(B)Z_t = Z_t - \sum_{j=1}^{\infty} \pi_j Z_{t-j} = a_t$$

Sendo, com os parâmetros ψ_j e π_j absolutamente calculáveis, satisfaz a igualdade:

$$\psi_j = \phi_1 \psi_{j-1} + \phi_2 \psi_{j-2} + \dots + \phi_p \psi_{j-p}, j > 0$$

Com $\psi_0 = 1, \psi_j = 0$ para $j < 0$, e $\theta_j = 0$ para $j > q$, satisfaz a igualdade:

$$\pi_1 = \theta_1 \pi_{j-1} + \theta_2 \pi_{j-2} + \dots + \theta_q \pi_{j-q}, q > 0 \text{ e } j > 0$$

Com $\pi_0 = -1, \pi_j = 0$ para $j < 0$, e $\phi_j = 0$ para $j > p$. De modo que o processo será estacionário se $\sum \psi_j < \infty$, ou seja, se $|\phi| < 1$ e o processo será invertível se $\sum \pi_j < \infty$, ou seja, $|\theta| < 1$. Observa-se que a condição de estacionariedade e de invertibilidade do processo ARMA é a mesma para o processo AR (p) e MA (q). Assim o processo ARMA (p,q) pode ser escrito como uma das equações abaixo:

$$Z_t = (B)a_t = \theta_q(B)\phi_p^{-1}(B)a_t$$

$$\pi(B)Z_t = \phi_p(B)\theta_q^{-1}(B)Z_t = a_t$$

2.1.3.4 Modelos auto regressivos integrados de médias móveis (ARIMA)

Se a série $W_t = \nabla^d Z_t$ onde ∇ é o operador de diferenças, for estacionária, é possível representar W_t por um modelo ARMA (p,q), ou seja:

$$\phi_p(B)W_t = \theta_q(B)a_t$$

$$\nabla^d Z_t = (1-B)^d Z_t$$

Se W_t for uma diferença de Z_t , então Z_t é uma integral de W_t , daí diz-se que Z_t segue um modelo auto regressivo, integrado, de médias móveis, ou modelo ARIMA. De ordem (p,d,q) e escreve-se ARIMA (p,d,q), onde p e q são as ordens dos polinômios $\phi(B)$ e $\theta(B)$, respectivamente:

$$\phi(B)\nabla^d Z_t = \theta(B)a_t$$

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

O modelo acima descrito supõe que a d-ésima diferença da série Z_t pode ser representada por um modelo ARMA, estacionário e invertível obedecendo às condições de estacionabilidade e

invertibilidade já vistos. O modelo ARIMA (p,d,q) é um caso especial de um processo integrado. Em geral, diz-se que Z_t é integrado de ordem d se $\nabla^d Z_t$ for estacionário, e escreve-se $Z_t \sim I(d)$, se d = 0, Z_t é estacionário. O modelo ARIMA descrito pode ser representado de três formas:

- Em termos de valores prévios de Z_t e do valor atual e prévio de a_t , pelo uso direto da equação de diferença.
- Somente em termos do valor atual e dos prévios de a_t ;
- Em termos de valores prévios de Z_t e do valor atual de a_t .

2.1.3.5 Modelos auto regressivos integrados de médias móveis sazonais (SARIMA)

Em grande parte das séries temporais existe a presença da componente periódica sazonal, está se repete a cada s (s > 1) observações. Em dados mensais, por exemplo, com período s = 12, espera-se que Z_t dependa de Z_{t-12} e, talvez, de Z_{t-24} , além de $Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots$. Em situações como esta, tomar uma diferença não é suficiente para eliminar um padrão sazonal acentuado e, conseqüentemente, deixar a série estacionária. É necessário, então, tomar diferenças no período sazonal.

Segundo Box et al (2015), o modelo ARIMA pode ser generalizado para se lidar com a sazonalidade, assim definiram um modelo ARIMA sazonal multiplicativo, denominado SARIMA (p, d, q) X (P, D, Q), representado por:

$$\phi(B^s)\phi(B)\nabla_s^D\nabla^d Z_t = \theta(B^s)\theta(B)a_t$$

Onde: $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ é o operador auto regressivo AR (p);

$\phi(B^s) = 1 - \phi_1 B^s - \dots - \phi_P B^{Ps}$ é o operador auto regressivo SAR (P), estacionário;

$\nabla^d = (1 - B)^d$ é o operador de diferença e d é o número de diferenças necessárias para retirar a tendência da série;

$\nabla_s^D = (1 - B^s)^D$ é o operador de diferença generalizado;

$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ é o operador de médias móveis MA (q);

$\theta(B^s) = 1 - \theta_1 B^s - \dots - \theta_Q B^{Qs}$ é o operador de médias móveis sazonal SMA (Q), invertível.

Para se realizar uma boa análise é necessário seguir os seguintes procedimentos (MÜLLER, 2011):

- 1º Passo: Construir o gráfico da série e observar o comportamento da série e verificar se existe estacionariedade ou sazonalidade;
- 2º Passo: Construir os gráficos da função de autocorrelação e da função de autocorrelação parcial para identificar o provável modelo;
- 3º Passo: Construir o gráfico do periodograma acumulado e identificar se existe parte sistemática na série para ser modelada;
- 4º Passo: Verificar quais modelos da série pode ser ajustada e que possam ser adequados;
- 5º Passo: Determinar a raiz quadrada do erro médio (RMSE) dos modelos propostos e escolher o modelo que apresentou o menor valor;
- 7º Passo: Analisar o teste de Box-Pierce dos modelos para confirmar que não há parte sistemática nos resíduos;
- 8º Passo: Construir o periodograma dos resíduos para os modelos sugeridos do modelo para verificar se restou somente parte aleatória.

2.2 Comparação dos métodos

A RMSE será o teste utilizado para a escolha do modelo neste estudo, quanto menor o valor do teste melhor será o modelo. Depois deste, analisa-se o teste Box-Pierce para ver se o modelo é adequado ou não adequado para as estatísticas apresentadas neste estudo. Segundo Morettin & Toloí (2006), sugerem o uso do teste de Box-Pierce (BOX et al, 2015) para verificar se o resíduo é branco. Este teste é baseado nas k primeiras estimativas das autocorrelações, $\hat{r}_k^2$, dos resíduos.

A hipótese de ruído branco, ou seja, se $Q(k) < X_{k-p-q}^2$, são: $H_0: \rho_1 = \rho_2 = \dots = \rho_t = 0$ e $H_a: \rho_t \neq 0$. Se o Q tabelado para determinado nível de significância, for maior que Q não se aceita a hipótese nula H_0 de que os resíduos não são autocorrelacionados. Porém, isto implica em aceitar a hipótese alternativa H_a de que pelo menos uma autocorrelação é diferente de zero, ou seja, o erro não é ruído branco. No caso da aceitação da hipótese alternativa H_a , rejeita-se o modelo.

2.2.1 Teste do periodograma acumulado dos resíduos

O teste consiste na comparação dos periodogramas acumulados da série dos resíduos estimados $\hat{a}_t$ e de um ruído branco a_t para verificar a existência de componentes periódicas na série dos resíduos estimados $\hat{a}_t$. Neste caso, a análise da série temporal é baseada na suposição de que ela é constituída de ondas senoidais e cosenoidais com diferentes frequências.

Para o ruído branco t o espectro acumulado de potência varia linearmente com a frequência. Assim, para a série dos resíduos $\hat{a}_t$, calcula-se:

- O periodograma da série $I(f_i)$ com $i = 1, 2, \dots, n/2$;
- O periodograma normalizado acumulado: $c(f_i) = \frac{1}{n\hat{\sigma}_a^2} \sum_{j=1}^i I(f_j)$ (estimador de $\frac{P(f)}{\sigma_a^2}$).

Então, quando a série $\hat{a}_t$ se aproxima do ruído branco, a função $c(f_i) \times f_i$ tem comportamento linear entre os pontos (0;0) e (0,5;1) e aceita-se o modelo fixado como válido. Ainda, como o periodograma acumulado está para o espectro teórico assim como a distribuição de frequências acumuladas observadas está para a experimental, com isso pode ser usado o teste de Kolmogorov-Smirnov para traçar linhas limites próximas da linha teórica e isto serve para indicar, com uma probabilidade determinada, a aproximação de $\hat{a}_t$ para o ruído branco.

3. Metodologia

O banco de dados disponibilizado pela ANEEL para a elaboração deste estudo apresenta dados de 64 distribuidoras de energia elétrica no período entre janeiro de 2003 a setembro de 2012. Com isso, o critério utilizado para a escolha das distribuidoras foram as distribuidoras com os maiores números de consumidores, escolhendo assim a maior de cada região do país com exceção da região sudeste devido ao grande número de consumidores e pelo fato de possuir grandes empresas com valores muito altos de perdas não-técnicas. Foram analisadas somente as perdas não-técnicas de 7 distribuidoras de energia elétrica em 117 meses, sendo os 3 últimos meses deixados para teste, sendo que, cada distribuidora tem seu modelo de série temporal individual.

Foi utilizado o *software* “Statgraphics Centurion XVI” para descrever e analisar o melhor modelo de série temporal para cada distribuidora de energia elétrica, e o “Microsoft Excel” para a montagem de tabelas e banco de dados.

4. Resultados

Foram analisadas 7 distribuidoras de energia elétrica em 117 meses sendo os três últimos meses deixados para teste. Cada distribuidora tem seu modelo de série temporal individual.

4.1 Modelo para cada distribuidora

4.1.1 Distribuidora centro-oeste

Com os valores apresentados na previsão (TABELA 1) pode-se dizer que o melhor modelo para a distribuidora centro-oeste é o de Brown, mas como o que define o melhor modelo é o menor valor do RSME, a melhor escolha passa a ser o SARIMA, sendo que o teste de Box-Pierce indica que esse é o único aceitável entre os três. Nesta distribuidora as previsões apresentaram a maior diferença relativa em relação ao valor real observado.

Período	Valor Real	SARIMA (2,0,2)x(3,2,3) ₁₂		Brown Linear		Holt-Winters	
		Valor	%	Valor	%	Valor	%
jul/12	99.052,20	47.487,60	-52,06%	38.553,80	-61,08%	38.926,70	-60,70%
ago/12	74.747,50	14.325,50	-80,83%	36.658,20	-50,96%	36.041,60	-51,78%
set/12	15.886,40	36.908,30	132,33%	27.434,40	72,69%	28.275,40	77,98%
RMSE		4.892,86		6.758,40		6.853,70	
P-valor Box-Pierce		0,681585		7,20808E-05		0,00105861	

Tabela 1 – Distribuidora centro-oeste

4.1.2 Distribuidora nordeste

Os valores da previsão (TABELA 2) apresentam menor diferença no modelo de Holt-Winters, mas o melhor modelo é o SARIMA devido ao menor valor de RSME.

Período	Valor Real	SARIMA (0,0,0)x(3,2,3) ₁₂		Brown Linear		Holt-Winters	
		Valor	%	Valor	%	Valor	%
jul/12	91.047,80	20.637,70	-77,33%	51.031,30	-43,95%	52.963,10	-41,83%
ago/12	45.701,90	33.178,40	-27,40%	64.456,90	41,04%	63.011,40	37,87%
set/12	72.520,70	25.721,70	-64,53%	51.008,90	-29,66%	51.063,20	-29,59%
RMSE		22.622,60		26.836,90		26.918,00	
P-valor Box-Pierce		0,49522		0,321134		0,624068	

Tabela 2 – Distribuidora nordeste

4.1.3 Distribuidora norte

Os modelos se ajustaram bem aos dados da distribuidora norte. O modelo de Holt-Winters apresenta valores com a menor diferença na previsão, mas o modelo SARIMA tem o menor valor de RSME e é o único considerado aceitável pelo teste de Box-Pierce. Esta foi a distribuidora onde os modelos apresentaram as melhores previsões.

Período	Valor Real	SARIMA (0,1,1)x(3,0,3)12		Brown Linear		Holt-Winters	
		Valor	%	Valor	%	Valor	%
jul/12	222.997,00	200.496,00	-10,09%	213.278,00	-4,36%	205.792,00	-7,72%
ago/12	238.700,00	216.337,00	-9,37%	254.804,00	6,75%	237.108,00	-0,67%
set/12	222.599,00	200.863,00	-9,76%	207.160,00	-6,94%	201.629,00	-9,42%
RMSE		7.481,14		11.339,20		10.640,30	
P-valor Box-Pierce		0,752367		5,15879E-05		0,0158986	

Tabela 3 – Distribuidora norte

4.1.4 Distribuidora sudeste 1

Na distribuidora sudeste 1 (TABELA 4), o melhor modelo em relação a diferença é o de Holt-Winters, mas o modelo com menor RMSE e teste Box-Pierce aceitável é o SARIMA.

Período	Valor Real	SARIMA (0,0,2)x(2,1,2)12		Brown Linear		Holt-Winters	
		Valor	%	Valor	%	Valor	%
jul/12	421.902,00	368.709,00	-12,61%	416.296,00	-1,33%	418.774,00	-0,74%
ago/12	438.562,00	515.408,00	17,52%	478.027,00	9,00%	468.426,00	6,81%
set/12	423.767,00	363.365,00	-14,25%	403.187,00	-4,86%	401.715,00	-5,20%
RMSE		53.638,90		72.877,70		75.375,20	
P-valor Box-Pierce		0,162589		1,01474E-13		1,44298E-06	

Tabela 4 – Distribuidora sudeste 1

4.1.5 Distribuidora sudeste 2

As melhores previsões para a distribuidora sudeste 2 foram as do modelo SARIMA (TABELA 5), exceto no mês de setembro. Este modelo apresenta o menor valor RMSE.

Período	Valor Real	SARIMA (3,2,3)x(3,1,3)12		Brown Linear		Holt-Winters	
		Valor	%	Valor	%	Valor	%
jul/12	79.934,90	71.470,90	-10,59%	67.596,20	-15,44%	70.496,30	-11,81%
ago/12	69.637,60	66.770,20	-4,12%	79.723,40	14,48%	78.251,00	12,37%
set/12	57.268,40	29.606,60	-48,30%	55.292,30	-3,45%	63.033,20	10,07%
RMSE		12.001,30		20.139,60		20.095,40	
P-valor Box-Pierce		0,0560469		0,0147878		0,278392	

Tabela 5 – Distribuidora sudeste 2

4.1.6 Distribuidora sudeste 3

Na distribuidora sudeste 3 (TABELA 6), as melhores previsões no período foram do modelo Holt-Winters, exceto no mês de agosto. O modelo SARIMA é o melhor devido ao menor valor RMSE.

Período	Valor Real	SARIMA (1,0,0)x(2,1,3) ₁₂		Brown Linear		Holt-Winters	
		Valor	%	Valor	%	Valor	%
jul/12	282.563,00	187.966,00	-33,48%	257.978,00	-8,70%	302.073,00	6,90%
ago/12	189.351,00	287.778,00	51,98%	275.248,00	45,36%	303.174,00	60,11%
set/12	250.846,00	163.127,00	-34,97%	162.131,00	-35,37%	187.791,00	-25,14%
RMSE		57.620,50		76.904,10		76.912,20	
P-valor Box-Pierce		0,159032		0,00710477		0,166186	

Tabela 6 – Distribuidora sudeste 3

4.1.7 Distribuidora sul

As melhores previsões nos meses de julho e agosto para a distribuidora sul foram as do modelo SARIMA. O valor de RMSE mostra que este modelo é o mais indicado.

Período	Valor Real	SARIMA (3,0,3)x(3,0,3) ₁₂		Brown Linear		Holt-Winters	
		Valor	%	Valor	%	Valor	%
jul/12	80.913,00	70.238,50	-13,19%	65.458,70	-19,10%	60.965,60	-24,65%
ago/12	78.658,00	68.822,40	-12,50%	62.728,50	-20,25%	59.542,60	-24,30%
set/12	35.984,00	48.618,70	35,11%	41.248,00	14,63%	37.673,40	4,69%
RMSE		5.358,93		7.641,11		7.884,65	
P-valor Box-Pierce		0,330518		0,299716		0,467711	

Tabela 7 – Distribuidora sul

Portanto, o melhor método identificado para todas as distribuidoras é o SARIMA (Box & Jenkins).

5. Conclusão

Conclui-se que os modelos não-automáticos da metodologia Box & Jenkins são melhores para fazer previsões de séries temporais entre os métodos escolhidos para comparação. As previsões que tiveram melhor qualidade foram da distribuidora norte, onde o erro de previsão foi menor que 10% em todos os três meses. As piores previsões foram das distribuidoras centro-oeste e nordeste.

Observou-se neste trabalho que os erros de previsão são não bons para as distribuidoras observadas. Um trabalho que pode ser realizado futuramente para analisarem estes dados seria a utilização de outros métodos. Sugere-se para trabalhos futuros a utilização do método de redes neurais artificiais, que vem apresentando na academia melhores resultados que os métodos estatísticos.

Referências

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA – ANEEL. Acesso aos dados no período de janeiro de 2003 a setembro de 2012.

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA – ANEEL. *Por dentro da conta de luz: informação de utilidade pública/Agência Nacional de Energia Elétrica*. 4. ed. Brasília: Aneel, 2008.

BOX, G. E., JENKINS, G. M., & REINSEL, G. C. *Time series analysis: forecasting and control*. 5. ed. New York: John Wiley & Sons, 2015.

CHAVES NETO, A.C. *Notas de aulas de análise séries temporais*. 2014.

- EHLERS, R.S.** *Análise de Séries Temporais*. 2009. Disponível em:
<<http://www.icmc.usp.br/ehlers/stemp/stemp.pdf>>. Acesso em: 03/03/2014.
- GOLDEMBERG, J.** *Energia no Brasil*. Editora: Livros Técnicos e Científicos, 1979.
- GRANGER, C.W.J.; NEWBOLD, P.** *Forecasting economic time series*. v. 83, n. 40, p. 574. Journal of the American Statistical Association, 1988.
- KOEHLER, A.B.; SNYDER, R.D.; ORD, J.K.** *Forecasting models and prediction intervals for the multiplicative Holt-Winters method*. v.17, n.2, p.269-286, apr./jun. International Journal of Forecasting, 2001.
- LJUNG, G.M.; BOX, G.E.P.** *On a measure of lack fit in time series models*. Biometrika, 1978.
- MAKRIDAKIS, S.; WHEELWRIGHT, S.; HYNDMAN, R.J.** *Forecasting Methods and Applications*. 3. ed. New York: John Wiley & Sons, 1998.
- MORETTIN, P.A.; TOLOI, C.M.C.** *Análise de Séries Temporais*. Editora: Blücher, 2006.
- MÜLLER, S.I.M.G.** *Notas de aulas da disciplina Séries Temporais*. DEST/UFPR, 2011.
- NELSON, C.R.** *Applied time series analysis for managerial forecasting*. v. 23, n. 1, p. 181-182, jan./fev. Operations Research , 1975.
- PENIN, C.A.S.** *Combate, prevenção e otimização das perdas comerciais de energia*. Tese. (Doutorado em Engenharia) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2008.
- SLUTZKY, E.** *The summation of random causes as the source of cyclic processes*. n. 5, p. 105 – 146. Econometrica, 1937.
- WOLD, H.O.** *A Study in the analysis of stationary time series*. Uppsala: Almqvist and Wiksell, 1938.
- YULE, G.U.** *Why do we sometimes get nonsense correlations between time-series*. n. 89, p. 1 – 64. Journal of the Royal Statistical Society, 1926.