

# DETECÇÃO DE FRAUDES E ESTIMAÇÃO DE CONSUMO EM REDES DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA UTILIZANDO ANÁLISE DE SÉRIES TEMPORAIS

ALEXANDRE DE A. BITENCOURT, VITOR H. FERREIRA,

*Departamento de Engenharia Elétrica, Universidade Federal Fluminense  
Rua Passo da Pátria, 152-470 – São Domingos, Niterói-RJ, Bloco D, sala 403  
alexandrebiten@hotmail.com, vhferreira@id.uff.br*

**Abstract**—Non-technical losses due to fraud in the distribution networks represent a significant loss for concessionaires in Brazil and worldwide. Several studies have attempted to detect these losses. This paper seeks a way to detect non-technical losses through models based in Holt-Winters and Linear Regression predictions. The first objective is to compare which model would best describe a customer's behavior. By using a hypothesis test we have as a result that Holt-Winters is the best model. The second goal is to identify fraudulent customers, and the Holt-Winters model do it quickly. The third objective is to compare the result of the account payable in the fraudulent month in relation to the indication of norm and the value of the forecast of the month, and results in a difference of 2,4%.

**Keywords**—Non-technical Losses, Fraud, Forecasting, Holt-Winters, Linear Regression

**Resumo**—As perdas não-técnicas devido a fraudes nas redes de distribuição representam um prejuízo significativo para as concessionárias no Brasil e no mundo. Diversos estudos tem tentado detectar estas perdas. Este artigo busca uma forma de detectar as perdas não-técnicas através dos modelos de previsão baseados em Holt-Winters e Regressão Linear. O primeiro objetivo é comparar qual seria o modelo que melhor descreve o comportamento de um cliente. Utilizando um teste de hipóteses tem-se como resultado que o Holt-Winters é o melhor modelo. O segundo objetivo é identificar os clientes fraudadores, e que o modelo Holt-Winters executa de forma rápida. O terceiro objetivo é comparar o resultado da conta a pagar no mês fraudado em relação a indicação de norma e o valor da previsão do mês, e tem como resultado uma diferença de 2,4%.

**Palavras-chave**—Perdas Não-Técnicas, Fraudes, Previsão, Holt-Winters, Regressão Linear

## 1 Introdução

Perdas na rede de distribuição podem ser classificadas como técnicas e não-técnicas, conforme a causa. As perdas técnicas são inerentes ao transporte de energia elétrica referente à transformação da energia elétrica em energia térmica devido ao efeito Joule, perdas nos núcleos dos transformadores, perdas dielétricas, dentre outras. Já as perdas não-técnicas correspondem à diferença entre as perdas totais e as perdas técnicas, considerando então todas as demais perdas associadas à distribuição de energia elétrica, tais como furto de energia, erro de medição, erros no processo de faturamento, dentre outras causas.

As perdas não técnicas podem alcançar até 40% da energia elétrica distribuída em países como Índia, Malásia, Líbano e Brasil (Glauner et al. 2017; Kumar V., Prasad, e Samikannu 2017). No Rio de Janeiro, segundo os dados da distribuidora Enel, as perdas por furto são os maiores de toda a área de atuação da empresa, 20,4% (O Globo 2017). Em um ano as perdas geraram um prejuízo a empresa de R\$ 95 milhões, que também deixam de ser arrecadados como impostos, causando prejuízo a toda população.

A identificação destas fraudes tem sido estudada, e inclui diversas formas de detecção, como métodos estatísticos, árvores de decisão, redes neurais, máquina de vetor suporte, métodos de agrupamento, modelos regressivos e outros (Ahmad 2017; Messinis e Hatzigiorgiou 2018; Guerrero et al. 2010). Essa identificação pode ser interpretada como um problema de previsão, onde o consumo do cliente diverge demasiadamente do valor previsto.

Boa parte das propostas encontradas na literatura utiliza a curva de consumo de energia para previsão e consequente identificação da fraude (Messinis e Hatzigiorgiou 2018), em que o consumo pode ser visto como uma série temporal (Kumar e Jain 2010; Islam et al. 2013; Seunghyoung Ryu, Jaekoo Noh, e Hongs-eok Kim 2016). Existem diversas técnicas para previsão de séries temporais, desde técnicas de inteligência artificial até modelos clássicos como regressão linear e alisamento exponencial (de Oliveira e Cyrino Oliveira 2018; Tao Hong et al. 2010).

No Brasil a Resolução Normativa Nº 414 da ANEEL estabelece as condições gerais de fornecimento de energia elétrica de forma consolidada, e teve sua última atualização em 2010 (ANEEL 2010). No capítulo XI, artigo 130, a resolução estabelece que a distribuidora ao comprovar o procedimento irregular deve apurar as diferenças entre os valores efetivamente faturados e aqueles apurados por meio de um dos critérios, que são: valor proporcional aos últimos 30 dias, aplicação de um fator de correção, média dos três maiores valores disponíveis de consumo nos últimos 30 dias e por fim determinação do consumo de energia excedente por meio da carga desviada ou por meio de carga instalada verificada no momento da constatação da irregularidade. Em todos os casos o valor cobrado pode não ser suficiente caso a irregularidade esteja sendo feita por um longo período.

Este artigo tem como objetivo desenvolver uma metodologia para detecção automática de fraudes e estimação do consumo fraudado. Para tanto, o sistema deve inicialmente selecionar o modelo mais adequado para estimação do consumo de um dado cliente, pos-

sibilitando a realização de previsões. Desvios em relação ao valor previsto são indicativos de possíveis fraudes, com a previsão podendo ser utilizada para estimação do valor mais provável para o consumo do cliente. Em função da característica dos dados disponíveis, relativos a clientes residenciais de uma distribuidora brasileira, são considerados modelos clássicos (regressão linear e alisamento exponencial Holt-Winters) para realização das previsões e construção dos intervalos de confiança para detecção de fraudes.

O artigo segue dividido na seção 2, aspectos teóricos, que contém uma breve descrição dos artigos relacionados ao trabalho, regressão linear, Holt-Winters e medidas de erro. A seção 3 descreve a metodologia proposta e os dados utilizados, a seção 4 os resultados encontrados e por fim a seção 5 a conclusão do trabalho.

## 2 Aspectos Teóricos

Como mencionado anteriormente a identificação de fraudes já tem sido amplamente estudada e pode ser vista como um problema de previsão. A seguir são descritos dois artigos referentes ao tema.

De Oliveira e Oliveira (de Oliveira e Cyrino Oliveira 2018) geram a previsão de dois anos a frente da demanda de energia elétrica mensal para diferentes países combinando diferentes técnicas de decomposição e agregação *bootstrap* (*bagging*) para simular as séries. Então, com as séries geradas, é estimada a previsão com diferentes modelos e comparadas com métricas estatísticas. É utilizada uma decomposição que tem como objetivo obter a tendência, sazonalidade e outros componentes restantes que compoem a série temporal, e então é aplicado uma variante proposta do *bagging* (*Remainder Sieve Bootstrap* - RSB) a estas componentes restantes a fim de aumentar a precisão da previsão. Os modelos utilizados para a previsão da série temporal são, Holt-Winters aditivo e multiplicativo, alisamento exponencial do espaço de estados e ARIMA com fator sazonal (SARIMA), e que são comparados com diversos modelos utilizados na literatura, como redes neurais e máquina de vetor suporte. Os resultados obtidos mostram que a técnica proposta tem precisão consistente para médio e longo prazo, com por exemplo um MAPE de 1,3% para a previsão com horizonte de 2 anos na demanda de energia da Itália.

Yip *et al.* (Yip *et al.* 2017) propoem dois algoritmos baseados em regressão linear para localizar medidores inteligentes com defeito ou furto em uma vizinhança. A ideia principal é adotar múltiplas regressões lineares (*Multiple Linear Regression* – MLP) para estimar e avaliar o coeficiente de anormalidade dos consumidores, baseado nos dados de consumo. O primeiro modelo (LR-ETDM) considera que o defeito ou furto não é interrompido a partir do seu início, sendo então incapaz de detectar defeitos e furtos intermitentes. Já o segundo modelo (CVLR-ETDM) é uma aprimoração do primeiro, sendo então apto a detecção desse comportamento. Foram utilizados sete cenários,

em que variam a duração e horário do furto ou defeito. Os resultados apontam que o modelo CVLR-ETDM foi mais apto na detecção de fraudes e defeitos devido a sua capacidade de distinguir a intermitência destes.

### 2.1 Regressão Linear

Em muitos problemas, duas ou mais variáveis estão relacionadas e é interessante modelar e explorar essa relação. Em geral, existe apenas uma única variável dependente ou resposta  $y$  que depende de  $k$  variáveis independentes, por exemplo  $x_1, x_2, \dots, x_k$ . A relação entre essas variáveis é caracterizada pelo modelo matemático de regressão. O modelo de regressão é adequado a um conjunto de dados de amostra. Em alguns casos, o desenvolvedor conhece a forma exata da relação funcional verdadeira entre  $y$  e  $x_1, x_2, \dots, x_k$ . No entanto, na maioria dos casos, o verdadeiro relacionamento funcional é desconhecido, e o desenvolvedor escolhe uma função apropriada para se aproximar do modelo verdadeiro. Os modelos polinomiais de baixa ordem são amplamente utilizados como funções aproximadas em que um modelo de regressão linear é o modelo de menor ordem. A Eq. (1) mostra um modelo de regressão linear múltipla, onde  $y$  é variável dependente,  $x_{1,2,\dots,k}$  são as variáveis independentes,  $\beta_{1,2,\dots,k}$  são os coeficientes de regressão,  $\beta_0$  a intercessão do plano e  $\varepsilon$  é o erro (Everitt e Skronidal 2010; Montgomery 2009).

$$y(x) = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \beta_0 + \varepsilon \quad (1)$$

A Eq. (1) pode ser reduzida a forma matricial, vista na Eq. (2), em que  $\underline{y}$  é o vetor das observações,  $\underline{X}$  é a matriz de nível das variáveis independentes,  $\underline{\beta}$  é o vetor dos coeficientes de regressão e  $\underline{\varepsilon}$  é o vetor de erro (Montgomery 2009).

$$\underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon} \quad (2)$$

Muitas vezes, é necessário construir estimativas do intervalo de confiança (IC) para os coeficientes de regressão e para outras quantidades de interesse do modelo de regressão. O desenvolvimento de um procedimento para a obtenção desses IC exige que assumamos que os erros sejam distribuídos de forma normal e independente com média zero e variância  $\sigma^2$  (Montgomery 2009).

Intervalos de confiança para preditores lineares consideram dois componentes para estimar a variância do erro de previsão. O primeiro é baseado na incerteza da estimação dos parâmetros. O segundo componente depende das propriedades estatísticas dos resíduos. A estimativa linear de mínimos quadrados garante que, se os resíduos forem pareados não correlacionados, com distribuição gaussiana, média zero e variância comum, então os parâmetros estimados também tem distribuição normal. Além disso, quando a amostra de dados é suficientemente grande, mesmo quando os resíduos não tem distribuição gaussiana, as

estimativas de mínimos quadrados têm uma distribuição aproximadamente gaussiana. Portanto, o erro de previsão do modelo linear pode ser caracterizado estimando-se a distribuição normal dos parâmetros (da Silva e Moulin 2000). O IC dos coeficientes é dado pela Eq. (3) (Montgomery 2009).

$$\begin{aligned} y_k - t_{\frac{\alpha}{2}, n-p} \sqrt{\sigma^2 C_{kk}} &\leq y_k \\ &\leq y_k + t_{\frac{\alpha}{2}, n-p} \sqrt{\sigma^2 C_{kk}} \end{aligned} \quad (3)$$

Em que  $t$  é uma distribuição normal,  $n - p$  o grau de liberdade,  $C_{kk}$  é o elemento da matriz  $(X'X)^{-1}$ ,  $y_k$  é o valor da previsão e  $\sigma^2$  é valor estimado da variância dada pela Eq. (4) em que  $SSe$  é a soma dos quadrados dos erros.

$$\sigma^2 = \frac{SSe}{n - p} \quad (4)$$

## 2.2 Holt-Winters

Uma série temporal é um conjunto de observações feitas sequencialmente através do tempo (Chatfield 2004). A previsão dos valores futuros de uma série temporal observada é um problema importante em muitas áreas. A atualização das equações que calculam a previsão recebe o nome de alisamento exponencial (Chatfield 2004). Esse método pode ser generalizado para tratar séries temporais que contem variações de tendência e sazonalidade, chamado de Holt-Winters (Chatfield 2004). As equações de atualizações do método são descritas pelas Eq. (5) a (7):

$$L_t = \alpha \left( \frac{x_t}{I_{t-S}} \right) + (1 - \alpha)(L_{t-1} + T_{t-1}) \quad (5)$$

$$T_t = \gamma(L_t - L_{t-1}) + (1 - \gamma)T_{t-1} \quad (6)$$

$$I_t = \delta(x_t/L_t) + (1 - \delta)I_{t-S} \quad (7)$$

Onde  $L_t$  é o nível,  $T_t$  a tendência,  $I_t$  a sazonalidade,  $S$  período sazonal e  $\alpha$ ,  $\gamma$  e  $\delta$  são os coeficientes de alisamento que variam entre 0 e 1.

Assim a Eq. (8) descreve a previsão:

$$x_t(h) = (L_t + hT_t)I_{t-S+h} \quad (8)$$

Onde  $x_t$  é a previsão no tempo  $t$  e  $h$  o passo de previsão.

Analogamente ao verificado para o modelo de regressão, também podem ser construídos intervalos de confiança para as previsões feitas pelo modelo de Holt-Winters. Este intervalo é da forma *previsão*  $\pm k\sqrt{\text{variância do erro da previsão}}$ , em que  $k$  de-

pende da distribuição de probabilidade do erro da previsão (Hyndman et al. 2008).

As equações para o cálculo da variância são dadas por (Koehler, Snyder, e Ord 2001):

$$h = 1$$

$$\text{Var}(e) = (L_t + T_t)^2 \sigma^2 \quad (9)$$

$$2 \leq h \leq m$$

$$\begin{aligned} \text{Var}(e) &\approx \frac{[(\alpha + (h-1)\gamma)^2(L_t + T_t)]}{I_{t+h-S}^2} \\ &+ \frac{(\alpha + \gamma)^2(L_t + (h-1)T_t)^2}{I_{t+(h-1)-S}^2} \\ &+ \frac{(L_t + hT_t)^2}{I_{t+h-S}^2} I_{t+h-S}^2 \sigma^2 \end{aligned} \quad (10)$$

Onde  $e$  é o erro da previsão,  $h$  são os passos da previsão e  $\sigma^2$  é a variância dada pela Eq. (11):

$$\sigma^2 = \sum_{t=1}^n \left( \frac{e_t}{L_{t-1} + T_{t-1}} \right)^2 / n \quad (11)$$

## 2.3 Medidas de erro

É necessária uma forma de mensurar o erro em modelos preditivos para que seja possível avaliar o resultado (Asare-Bediako, Kling, e Ribeiro 2013). A forma mais comum é através do erro percentual absoluto médio (em inglês, *mean absolute percentage error* - MAPE), apresentado na Eq. (12).

$$\text{MAPE} = \frac{100}{n} \cdot \sum_{t=1}^n \frac{|A_t - F_t|}{A_t} \quad (12)$$

Em que  $A_t$  é o valor atual e  $F_t$  é o valor da previsão.

## 2.4 Teste de Hipóteses

Um teste de hipóteses é um método de inferência estatística. É um estudo baseado na análise de uma amostra, usado para avaliar determinados parâmetros que são desconhecidos numa população.

Os testes de hipótese são constituídos de alternativas que são testadas através de uma estatística de teste. Uma população tem uma amostra retirada e através da aplicação de teoria de probabilidades é possível tirar conclusões em relação a essa amostra, como por exemplo, determinar sua veracidade em relação à composição da população ou auxiliar na comprovação de uma teoria. Para esse estudo será utilizado o teste  $t$ .

O teste t, dentre outras formas de uso, pode ser utilizado a fim de comparar duas médias de diferentes conjuntos de dados distintos com variâncias desconhecidas. Para mais informações sobre métodos de inferência, (Cox 2006) e (L. Loveland 2011).

### 3 Metodologia Proposta

Os dados de entrada considerados para previsão são os dados de consumo de energia elétrica de Clientes da cidade inteligente de Búzios, no Rio de Janeiro. A base de dados possui medições diárias de energia ao longo de 609 dias. Dentre os diversos medidores na base de dados, foram selecionados 14 clientes com a menor taxa de falha (ausência de leitura). O método de Holt-Winters foi utilizado para preencher as medições faltantes.

A Figura 1 mostra o fluxograma de funcionamento do modelo, que se inicia selecionando os dados do medidor correspondente e os separa em treinamento e medição futura. Os primeiros 579 dias são os dados de validação e os 30 últimos representam a medição de 1 mês que será feita. Em seguida os modelos de previsão Holt-Winters (HW) e Regressão Linear (RL) são aplicados e se registra a previsão e o erro da série estimada. São calculados então os limites da previsão a partir dos intervalos de confiança de cada modelo. É feito em seguida um teste de hipóteses (teste t) para descobrir o melhor modelo a partir do erro da série estimada com um grau de confiança de 99%. Então a fraude é definida, existindo dois tipos possíveis. O primeiro tipo, tipo 1, é uma fraude diária e que se aplica um percentual de redução do consumo baseado nos dados da medição futura. O segundo tipo, tipo 2, é uma fraude que só ocorre nos fins de semana, e que também aplica um percentual de redução de consumo quando a fraude ocorre. Em seguida, usando as características dos modelos tenta-se detectar as fraudes e o primeiro dia em que ocorrem utilizando a condição de que caso a medição seja menor do que o limite mínimo do modelo há fraude. A programação foi feita no MATLAB.

O modelo de Holt Winters utilizado é o multiplicativo para nível e tendência, com passo de previsão de 1 dia e período sazonal 1 dia. Os coeficientes de alisamento são definidos por código a fim de minimizar o erro quadrático da previsão. O  $k$  para um nível de confiança de 99% para os limites de previsão tem valor de 2,3263.

Foram considerados para análise quatro estudos de caso, em que se varia o tipo de fraude e a quantidade de consumo reduzido. O primeiro caso é uma fraude do tipo 1 com redução de 50%, o segundo caso é do tipo 1 com 100% de redução, o terceiro caso é do tipo 2 com redução de 50% e o quarto caso do tipo 2 com 100% de redução.

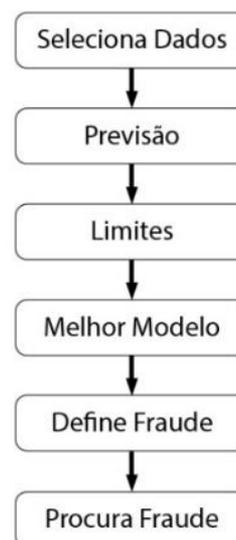


Figura 1. Fluxograma da metodologia proposta

### 4 Resultados

Para todos os medidores o modelo de HW foi selecionado como o melhor modelo baseado no teste hipóteses para o erro da série estimada, em que os resultados do MAPE deste erro são apresentados na Tabela 1.

Tabela 1. MAPE do erro da série estimada

Medidor	MAPE Série Estimada	
	RL	HW
1	0,003	0,002
2	0,007	0,001
3	0,071	0,000
4	0,003	0,002
5	0,020	0,000
6	0,012	0,001
7	0,001	0,002
8	0,011	0,002
9	0,018	0,002
10	0,080	0,000
11	0,016	0,000
12	0,085	0,004
13	0,020	0,000
14	0,011	0,001

Entretanto o modelo de HW não foi capaz de detectar todas as fraudes, como visto na Tabela 2, em que para o valor 0 o modelo não encontrou a fraude e 1 encontrou. A Tabela 2 também indica a partir de qual dia a fraude foi detectada. Para o medidor 4, por exemplo, a fraude foi detectada logo no primeiro dia considerando o modelo HW. Caso fosse utilizado o modelo RL, a fraude seria detectada somente no quarto dia.

É possível observar que, na maioria dos casos em que o HW detectou a fraude a RL também o fez. Porém como o modelo de HW foi selecionado como o melhor modelo sua resposta é mais confiável, já que a previsão da RL não capturou o comportamento do

cliente tanto quanto o HW. Tal resultado pode ser visto na Figura 2, que apresenta os resultados do medidor 4 e que apesar de ambos os modelos terem encontrado a fraude, o HW foi capaz de encontrar primeiro.

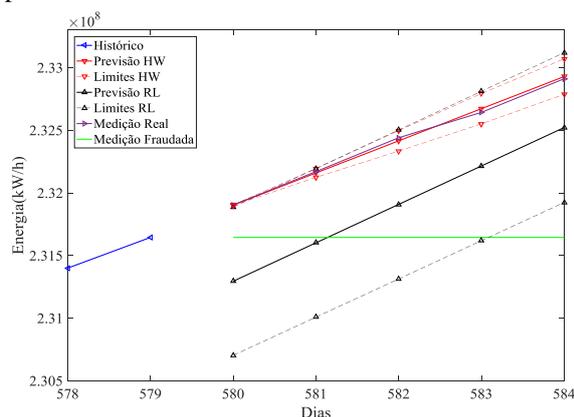


Figura 2. Detecção de fraude para medidor 4 no caso 2

Também é possível notar que houve 4 medidores (3, 5, 11 e 13) em que o HW não detectou fraude e a RL detectou. Analisando esses medidores é possível observar através do histórico um comportamento suspeito, e que o modelo HW assimilou, tendo assim um resultado tendencioso na previsão. Para o medidor 10 não foi detectada a fraude em nenhum modelo, porém novamente, analisando o histórico é possível observar comportamento suspeito, visto na Figura 3. Por fim destaca-se que para os caso 1 e 2 as fraudes foram detectadas da mesma forma, porém no caso 2 elas foram detectadas com mais rapidez.

Tabela 2. Caso 2 Fraude tipo 1 com redução de consumo de 100%

Medidor	Fraude diária - 100%			
	Fraudou		Dia Fraude	
	HW	RL	HW	RL
1	1	1	1	1
2	1	1	1	1
3	0	1	0	1
4	1	1	1	4
5	0	1	0	1
6	1	1	1	1
7	1	1	1	2
8	1	1	1	15
9	1	1	1	1
10	0	0	0	0
11	0	1	0	1
12	1	0	1	0
13	0	1	0	1
14	1	1	1	20

O caso 3 é o mais ameno, em que o cliente fraudava apenas 50% do consumo nos fins de semana, logo pode-se concluir que este seria o furto mais difícil de se detectar dentre os casos estudados. O resultado pode ser visto na Tabela 3.

Já para os medidores 8, 12 e 14 não foi possível detectar a fraude para o caso 3. Para o caso 4 os resultados podem ser vistos na Tabela 4, em que o medidor 1 e 8 passam a ter suas fraudes detectadas pelo HW. Pode-se observar também na Tabela 4 que as fraudes

foram detectadas logo no primeiro dia pelo HW, com exceção do medidor 8.

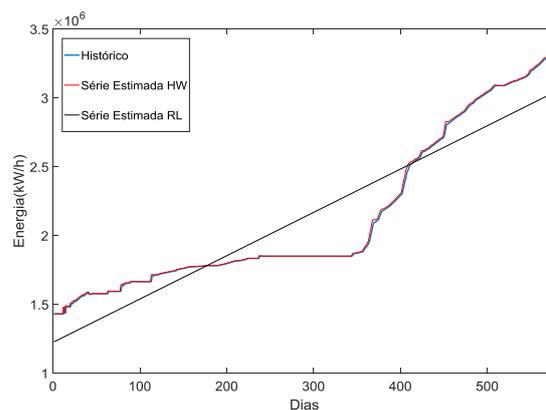


Figura 3. Histórico e séries estimadas para o medidor 10

Os medidores 3, 5, 11 e 13 possuem as mesmas conclusões que o caso 1 e 2, assim como o medidor 10. Entretanto no medidor 1 o modelo de HW não acusa fraude enquanto a RL acusa. Porém pode-se observar na Figura 4 que até mesmo para um caso de medição comum o modelo de RL teria acusado como fraude.

Tabela 3. Caso 3 Fraude tipo 2 com redução de consumo de 50%

Medidor	Fraude fim de semana - 50%			
	Fraudou		Dia Fraude	
	HW	RL	HW	RL
1	0	1	0	1
2	1	1	7	1
3	0	1	0	1
4	1	1	6	14
5	0	1	0	1
6	1	1	13	1
7	1	1	6	6
8	0	0	0	0
9	0	1	0	1
10	0	0	0	0
11	0	1	0	1
12	0	0	0	0
13	0	1	0	1
14	0	0	0	0

Os resultados do caso 1 não foram explorados por apresentarem o mesmo comportamento que o caso 2, tendo diferenças apenas no dia da detecção da fraude no modelo RL.

É importante destacar que, para o caso 3 e 4, apesar da RL encontrar fraudes no dia 1 isso não é possível, já que as primeiras fraudes acontecem no dia 6 e 7. Esse resultado fortalece a ideia de que a RL estima a previsão com um valor acima do real. Isso se dá ao fato da RL ser um modelo mais simples que o HW e não captura o comportamento do cliente com a mesma precisão.

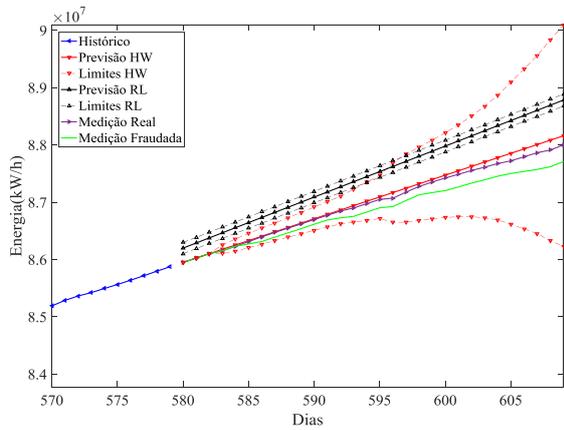


Figura 4. Detecção de fraude para medidor 1 no caso 3

O tempo para se encontrar a fraude é importante para determinar o dia de atualização do histórico da base de treinamento. Considerando o caso 3 como o mais ameno, para o medidor 4 foram necessários 14 dias para se confirmar a fraude pela RL, sendo então recomendada a atualização no dia seguinte, ou seja, a cada 15 dias.

Tabela 4. Caso 4 Fraude tipo 2 com redução de consumo de 100%

Medidor	Fraude fim de semana - 100%			
	Fraudou		Dia Fraude	
	HW	RL	HW	RL
1	1	1	6	1
2	1	1	6	1
3	0	1	0	1
4	1	1	6	13
5	0	1	0	1
6	1	1	6	1
7	1	1	6	6
8	1	0	7	0
9	0	1	0	1
10	0	0	0	0
11	0	1	0	1
12	0	0	0	0
13	0	1	0	1
14	0	0	0	0

Para o cálculo do valor a ser pago ao detectar a fraude foi considerado o caso 2, em que o cliente reduziu o consumo em 100% ao longo de todo o mês e a indicação da Resolução Normativa N° 414 da ANEEL em considerar valor proporcional aos 30 dias anteriores. A Tabela 5 apresenta o resultado pago [R\$] referente aos 30 dias anteriores a fraude recomendado pela norma N° 414 da ANEEL, contra o valor calculado a ser pago baseado na diferença da previsão e a medição fraudada para uma tarifa de 0,70 [R\$/kWh]. Nota-se que os medidores 3, 5, 10, 11 e 13 não possuem valor a ser pago pois não foi detectada a fraude. A diferença entre a média do valor arrecadado considerando os medidores que pagariam multa nos dois métodos é de 2,4%.

Tabela 5. Valor a pagar no mês fraudado

Medidor	R\$ faturado por norma	R\$ faturado proposto
1	1.678,26	1.599,27
2	154,66	150,26
3	14,52	0,00
4	5.501,52	5.386,92
5	14,13	0,00
6	218,92	214,06
7	5.956,88	5.811,56
8	60,76	58,61
9	1.425,40	1.423,29
10	89,70	0,00
11	0,00	0,00
12	68,74	73,87
13	5,49	0,00
14	81,27	67,68

## 5 Conclusão

O artigo apresenta uma estratégia para selecionar o modelo que descreve o comportamento de um cliente com maior precisão. Utilizando um teste de hipóteses conclui-se que, com 99% de confiança, o modelo Holt-Winters tem um erro menor do que o modelo de Regressão Linear. Logo o Holt-Winters é considerado o melhor modelo.

São analisados quatro estudos de caso, em que cada caso representa um tipo de fraude. Os resultados apontam que o modelo Holt-Winters é capaz de detectar fraudes de forma rápida, ou seja, pode ser encontrada no mesmo dia em que ocorre. Entretanto existem casos em que o modelo não é capaz de detectar a fraude, por exemplo quando há no histórico dados suspeitos. Porém é possível utilizar outro modelo para sinalizar os casos que o Holt-Winters não foi capaz de detectar, como a Regressão Linear.

Para o caso mais ameno de fraude, 3 clientes não foram detectados e, a partir deste mesmo caso, foi determinado um prazo de atualização dos dados do histórico de 15 dias.

Por fim, em uma comparação da tarifação do mês fraudado entre a recomendação da norma e a previsão do modelo, obteve-se uma diferença de 2,4%, ou seja, o valor é próximo ao estabelecido pela legislação.

## Agradecimentos

Os autores agradecem à Enel Distribuição Rio pela disponibilização dos dados, fruto da participação da UFF no projeto de P&D ANEEL Cidade Inteligente Búzios (PD-0383-0055/2011).

## Referências Bibliográficas

Ahmad, Tanveer. 2017. "Non-Technical Loss Analysis and Prevention Using Smart Meters". *Renewable and Sustainable Energy Reviews* 72 (maio): 573–89. <https://doi.org/10.1016/j.rser.2017.01.100>.

- ANEEL. 2010. *RESOLUÇÃO NORMATIVA Nº 414*. <http://www.aneel.gov.br/documents/656877/14486448/bren2010414.pdf/3bd33297-26f9-4ddf-94c3-f01d76d6f14a?Version=1.0>.
- Asare-Bediako, B., W. L. Kling, e P. F. Ribeiro. 2013. "Day-ahead residential load forecasting with artificial neural networks using smart meter data". In , 1–6. IEEE. <https://doi.org/10.1109/PTC.2013.6652093>.
- Chatfield, Christopher. 2004. *The analysis of time series: an introduction*. 6th ed. Texts in statistical science. Boca Raton, FL: Chapman & Hall/CRC.
- Cox, D. R. 2006. *Principles of statistical inference*. Cambridge ; New York: Cambridge University Press.
- Everitt, Brian, e Anders Skrondal. 2010. *The Cambridge Dictionary of Statistics*. <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=329356>.
- Glauner, Patrick, Angelo Migliosi, Jorge Augusto Meira, Petko Valtchev, Radu State, e Franck Bettinger. 2017. "Is big data sufficient for a reliable detection of non-technical losses?" In , 1–6. IEEE. <https://doi.org/10.1109/ISAP.2017.8071420>.
- Guerrero, Juan I., Carlos Leon, Felix Biscarri, Inigo Monedero, Jesus Biscarri, e Rocio Millan. 2010. "Increasing the efficiency in Non-Technical Losses detection in utility companies". In , 136–41. IEEE. <https://doi.org/10.1109/MELCON.2010.5476320>.
- Hyndman, Rob J., Anne B. Koehler, J. Keith Ord, e Ralph D. Snyder. 2008. *Forecasting with Exponential Smoothing - The State Space Approach*. Springer.
- Islam, Faridul, Muhammad Shahbaz, Ashraf U. Ahmed, e Md. Mahmudul Alam. 2013. "Financial Development and Energy Consumption Nexus in Malaysia: A Multivariate Time Series Analysis". *Economic Modelling* 30 (janeiro): 435–41. <https://doi.org/10.1016/j.econmod.2012.09.033>.
- Koehler, Anne B., Ralph D. Snyder, e J.Keith Ord. 2001. "Forecasting Models and Prediction Intervals for the Multiplicative Holt–Winters Method". *International Journal of Forecasting* 17 (2): 269–86. [https://doi.org/10.1016/S0169-2070\(01\)00081-4](https://doi.org/10.1016/S0169-2070(01)00081-4).
- Kumar, Ujjwal, e V.K. Jain. 2010. "Time Series Models (Grey-Markov, Grey Model with Rolling Mechanism and Singular Spectrum Analysis) to Forecast Energy Consumption in India". *Energy* 35 (4): 1709–16. <https://doi.org/10.1016/j.energy.2009.12.021>.
- Kumar V., Sampath, Jagdish Prasad, e Ravi Samikannu. 2017. "Overview, Issues and Prevention of Energy Theft in Smart Grids and Virtual Power Plants in Indian Context". *Energy Policy* 110 (novembro): 365–74. <https://doi.org/10.1016/j.enpol.2017.08.032>.
- L. Loveland, Jennifer. 2011. *Mathematical Justification of Introductory Hypothesis Tests and Development of Reference Materials*. Utah State University. <http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1014&context=gradreports>.
- Messinis, George M., e Nikos D. Hatzigiargyriou. 2018. "Review of Non-Technical Loss Detection Methods". *Electric Power Systems Research* 158 (maio): 250–66. <https://doi.org/10.1016/j.epsr.2018.01.005>.
- Montgomery, Douglas C. 2009. *Introduction to statistical quality control*. 6th ed. Hoboken, N.J: Wiley.
- O Globo. 2017. "Metade dos furtos de energia elétrica no Rio está nas classes A, B e C, segundo a Light", 30 de novembro de 2017. <http://www.info4.com.br/ver/ver.asp?Yw=MTY4NQ&YQ=MTY4NQ&bA=Njg2NTEw&b3JpZ2Vt=ZW1haWw>.
- Oliveira, Erick Meira de, e Fernando Luiz Cyrino Oliveira. 2018. "Forecasting Mid-Long Term Electric Energy Consumption through Bagging ARIMA and Exponential Smoothing Methods". *Energy* 144 (fevereiro): 776–88. <https://doi.org/10.1016/j.energy.2017.12.049>.
- Seunghyoung Ryu, Jaekoo Noh, e Hongseok Kim. 2016. "Deep neural network based demand side short term load forecasting". In , 308–13. IEEE. <https://doi.org/10.1109/SmartGridComm.2016.7778779>.
- Silva, A.P.A. da, e L.S. Moulin. 2000. "Confidence intervals for neural network based short-term load forecasting". *IEEE Transactions on Power Systems* 15 (4): 1191–96. <https://doi.org/10.1109/59.898089>.
- Tao Hong, Min Gui, Mesut E Baran, e H Lee Willis. 2010. "Modeling and forecasting hourly electric load by multiple linear regression with interactions". In , 1–8. IEEE. <https://doi.org/10.1109/PES.2010.5589959>.
- Yip, Sook-Chin, KokSheik Wong, Wooi-Ping Hew, Ming-Tao Gan, Raphael C.-W. Phan, e Su-Wei Tan. 2017. "Detection of Energy Theft and Defective Smart Meters in Smart Grids Using Linear Regression". *International Journal of Electrical Power & Energy Systems* 91 (outubro): 230–40. <https://doi.org/10.1016/j.ijepes.2017.04.005>.