

## Detecção de Perdas não Técnicas em Sistema de Distribuição Empregando a Transformada Wavelet

Almir Carlos dos Santos Júnior\*  
Tarso Vilela Ferreira\*\*

\*Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Campina Grande, Campina Grande, Paraíba, Brasil (e-mail: almir.carlos.junior@ee.ufcg.edu.br)

\*\*Departamento de Engenharia Elétrica da Universidade Federal de Sergipe, Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Sergipe  
Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Campina Grande  
São Cristóvão, Sergipe, Brasil (e-mail: tarso@academico.ufs.br)

---

**Abstract:** Non-technical losses account for 7,5% of the energy injected into the Brazilian electricity grid. That losses cannot be fully passed on from distributors to consumers, which impacts on the profit of the companies, but also impacts in the responsible consumer, since part of this loss is prorated among regular users. The development of non-technical loss detection methodologies is important since, by discovering fraud, there is the possibility of revenue recovery and mitigation of other consequences. In this paper, two specific algorithmic solutions are proposed, capable of acting as a tool for extracting intrinsic characteristics of consumption patterns, which allow the detection of behavior associated with fraud. To develop both, a database with monthly consumption information in kilowatt-hour was created. The first technique makes use of discrete wavelet transform, so wavelet families were evaluated in order to define which one is the most suitable for implementing the technique. The Haar wavelet proved to be promising for the database in question, allowing a rate of correct answers in the classification of commercial customers around 77% and 86% accuracy in detecting fraud in the same customers. The second technique had a 79% hit rate for commercial customer ratings and a 78% fraud detection rate. As highlighted throughout the text, the complementarity of results for each of the techniques motivated the development of both.

**Resumo:** As perdas não técnicas respondem por 7,5% da energia injetada na rede elétrica brasileira. Tais perdas não podem ser repassadas integralmente das distribuidoras para os consumidores, o que impacta no lucro das empresas, mas também impacta o consumidor responsável, já que parte deste prejuízo é rateado entre os usuários regulares. O desenvolvimento de metodologias de detecção de perdas não técnicas é importante uma vez que, descobrindo-se a fraude, há a possibilidade de recuperação da receita e mitigação das demais consequências. Neste artigo são propostas duas soluções algorítmicas específicas capazes de atuar como ferramenta de extração de características intrínsecas dos padrões de consumo, as quais permitem detectar comportamentos associados a fraudes. Para desenvolvimento de ambas, uma base de dados com informações de consumo mensais em quilowatt-hora foi levantada. A primeira técnica faz uso da transformada wavelet discreta, então, famílias wavelet foram avaliadas a fim de se definir qual delas é a mais adequada à implementação da técnica. A wavelet Haar mostrou-se promissora para a base de dados em questão, permitindo uma taxa de acertos na classificação de clientes comerciais em torno de 77%, e 86% de acurácia na detecção de fraudes nos mesmos clientes. A segunda técnica apresentou 79% de taxa de acertos de classificação de clientes comerciais e 78% de taxa de detecção de fraudes. Conforme destacado ao longo do texto, a complementariedade dos resultados relativamente a cada uma das técnicas motivou o desenvolvimento de ambas.

**Keywords:** commercial losses; wavelet transforms; algorithm; energy distribution; revenue recovery.

**Palavras-chaves:** perdas comerciais; transformada wavelet; algoritmo; distribuição de energia; recuperação de receita.

---

### 1. INTRODUÇÃO

A perda de energia é a diferença entre a energia requerida pela empresa, composta por compras e/ou geração própria, e a vendida aos clientes. As perdas são segregadas quanto a sua origem em técnicas e não técnicas. As perdas técnicas são

aquelas relacionadas à operação dos equipamentos da rede. Já as não técnicas, ou comerciais correspondem a energia entregue ao cliente, mas não computada como vendida (Leite and Mantovani, 2016, Almeida et al. 2006).

Em 2020, as perdas não técnicas (PNT) representaram 7,5% da energia injetada na rede, o que significa 37,9 TWh

(ANEEL, 2021). Com esta quantidade de energia seria possível atender a cidade de São Paulo durante um ano e quatro meses (Secretaria de Infraestrutura de São Paulo, 2020). Dentro de percentuais regulatórios fixados nas Resoluções Homologatórias (REHs), o repasse tarifário dos níveis eficientes de perdas é repassado para a revisão das tarifas em ciclos de 3, 4 ou 5 anos (ANEEL, 2021). Isto significa que as PNT são prejudiciais tanto para o consumidor, que terá sua conta encarecida, quanto para as distribuidoras, que não recebem integralmente por essas perdas. Além disso, as PNT também são prejudiciais ao meio ambiente, uma vez que usuários inadimplentes tendem a não economizar no consumo de energia elétrica.

As perdas comerciais e formas de mitigação destas são assuntos caros nas empresas de distribuição e na literatura relacionada à eficiência energética. A exemplo do trabalho desenvolvido por Almeida et al. (2006) em que se destacam as formas de fraude e furto encontrados pela Coelba e o que a empresa faz para diminuir as perdas. Foram enfatizadas as campanhas de regularização que utilizam como parâmetro de busca o Índice de Perdas em Subestações e de eficiência as Energia Recuperada e Agregada.

Em 2008, Penin ressaltou a inclusão das *smart grids*, redes inteligentes, como um aliado ao combate aos roubos de energia. Ademais, foi destacada a utilização de métodos computacionais como auxílio na localização e sinalização de potenciais fraudadores. Utilizando as técnicas de mineração de dados, este autor percebeu que diversos fatores, inclusive a localização, influenciam nas PNT (Penin, 2008).

Assistentes para a detecção de fraudes baseados em redes neurais artificiais (RNA) foram desenvolvidos em alguns trabalhos. Em 2017, Menezes definiu os graus de influência de variados parâmetros nos resultados das suas redes, a saber: tamanho da saída, configuração de entrada, anos e dimensão da camada oculta, desenvolveu resultados melhores para saídas de tamanho 1, configuração de entrada C4, 3 anos de dados de consumo e dimensão de camada oculta 1, obteve a máxima taxa de acertos de 81,26% (Menezes, 2017). Também em 2017, Silveira utilizou os parâmetros de quantidade de épocas de treinamento, taxa de acerto, taxa de detecção de fraudadores e taxa de indicações indevidas para classificar as arquiteturas de redes utilizadas. Concluiu-se o trabalho com os melhores resultados utilizando-se de 349 épocas de treinamento, taxa de detecção de fraudadores de 64,4% e taxa de indicações indevidas de 13,5%, com uma taxa máxima de acertos de 80,8% (Silveira, 2017).

Ramos, em 2014, fez uma ampla revisão bibliográfica acerca das PNT, a fim de caracterizar-se e detectar-se algoritmicamente uma fraude. O caminho seguido pelo autor passa pelas RNA, máquinas de vetores de suporte e floresta de caminhos ótimos, além de diversas técnicas evolutivas para a otimização. Nos experimentos deste trabalho, dividiram-se diversas bases de dados para receber as metodologias supracitadas. Os parâmetros de eficácia avaliados foram o tempo de treinamento, tempo de testes e taxa de acertos, sendo que a maior taxa de acertos obtida foi de cerca de 98%. Vale ressaltar que a técnica proposta neste trabalho possui alto custo computacional e, devido a composição do banco de dados, o

resultado pode tender a detectar-se características mais marcantes devido a magnitude do consumo, além disso, clientes residenciais não foram analisados (Ramos, 2014).

Em 2018, Avila e outros desenvolveram uma técnica que fazia uso do método da máxima sobreposição discreta na transformada *wavelet packet* como o principal mecanismo para extração de características de clientes fraudulentos e não fraudulentos. Além disso, fez o aumento da subamostragem aleatoriamente e para implementação do resultado, utilizou-se a lógica *fuzzy*. Neste estudo, foi obtida uma acurácia de 93% utilizando dados de consumo apurados dia a dia e 80% para a média mensal de consumo. No entanto, a base de dados possuía 2153 consumidores não fraudulentos e apenas 118 fraudulentos, o que pode indicar uma tendência do algoritmo a detectar não fraudulentos muito facilmente e, mesmo errando, não afetar o desempenho global (Avila et al. 2018).

Punmiya e Choe, em 2019 avaliaram métricas advindas da curva de consumo diário, ressaltando características, para melhorar a caracterização das PNT (Punmiya and Choe, 2019). Raggi e outros em 2020 apresentaram uma metodologia de identificação de PNT baseada em medições de *smart meter* e extrapolação de dados (Raggi et al. 2020). Já em 2021, Barros utilizou técnicas de *advanced analytics* para caracterizar as PNT não só com as características físicas, mas também, em critérios econômicos, a saber: a melhor época para realização de visitas técnicas, campanhas de conscientização e cortes de energia (Barros, 2021).

Portanto, torna-se evidente a importância econômica, ambiental e social da mitigação das PNT. Assim, concessionárias buscam minimizá-las ao máximo e são recomendadas por meio da regulação a manter as PNT dentro de patamares estabelecidos pela ANEEL. Percebe-se, portanto, que se fazem necessárias complementações às técnicas de detecção de perdas comerciais já existentes e desenvolvimento de novas metodologias que auxiliem nesta problemática, as quais permitiriam detectar clientes elegíveis à visita de fiscais das distribuidoras de energia.

Dessa forma, propõe-se, neste artigo, um novo método, com baixo custo computacional e simples implementação para detecção de fraudes em sistema de distribuição utilizando a decomposição por meio da transformada *wavelet* discreta (TWD). A transformada *wavelet* (TW), amplamente utilizada para filtragem de características ruidosas de sinais, foi utilizada para ressaltar características fraudulentas de clientes neste trabalho. Resultando em possível mitigação de fraudes em sistemas e processos de aferição e faturamento de energia elétrica. Somado à decomposição, a ferramenta de avaliação da similaridade entre as famílias *wavelet* e os dados avaliados foi a correlação cruzada. Em seguida, a eficácia deste método foi comparada com outro, análogo, mas que não faz uso da transformada, por meio de conceitos como taxa de acertos e taxa de detecção de fraudes.

Ademais, o artigo está estruturado da seguinte forma: na seção atual, a motivação do tema foi apresentada; na seção 2 é apresentada a fundamentação teórica acerca do trabalho; na seção 3 a metodologia é descrita; na seção 4 são apresentados os resultados e; na seção 5, as considerações finais.

## 2. TRANSFORMADA WAVELET

A transformada *wavelet* é uma operação linear que mapeia um sinal de seu domínio original para o domínio tempo–frequência. Um sinal unidimensional (por ex., tempo) é mapeado em um sinal bidimensional (tempo–frequência) (Castro, 2012). Uma *wavelet* tem sua energia concentrada no tempo para dar uma ferramenta de análise de fenômenos transitórios, não estacionários ou variantes no tempo (Burrus et al. 2013).

A transformada *wavelet* contínua (TWC) é um histórico ao longo do tempo do produto entre um sinal contínuo, uma escala e uma função que o desloca. A TWC é dada por (1):

$$W_f(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \Psi\left(\frac{t-b}{a}\right) dt \quad (1)$$

Em que  $a$  é o coeficiente de escala,  $b$  o de deslocamento,  $W_f(a, b)$  são os coeficientes *wavelet* contínuos,  $f(t)$  é o sinal contínuo que será decomposto e  $\Psi(t)$  é a *wavelet* mãe deslocada e que muda a escala do sinal  $f(t)$ .

A implementação da transformada *wavelet* em computadores digitais se dá por meio da sua versão discreta. De acordo com Walker (1999), todas as transformadas *wavelet* discretas decompõem um sinal discreto em dois subsinais com metade do tamanho do sinal original.

Na TWD, o sinal é decomposto em termos de *wavelet* filhas,  $\Psi_{a,b}(t)$  que são funções obtidas a partir da *wavelet* mãe,  $\Psi(t)$ . Uma discretização típica é apresentada em (2):

$$\Psi_{j,k}(t) = 2^{\frac{j}{2}} \Psi(2^j t - k) \quad (2)$$

Em que  $j, k \in \mathbb{Z}$ .  $\mathbb{Z}$  é o conjunto dos números inteiros e  $2^{\frac{j}{2}}$  mantém uma norma constante, independente da escala de  $j$  (Burrus et al. 2013). O resultado dos coeficientes para análise com um sinal discreto  $h(n)$  ficam conforme (3):

$$\Psi_{j,k}(t) = 2^{\frac{j}{2}} \sum_{-\infty}^{\infty} h(n) \Psi(2^j n - k) \quad (3)$$

Em que  $j, k \in \mathbb{Z}$  e  $h(n)$  é um sinal discreto.

Os subsinais gerados pela decomposição *wavelet* são chamados de aproximação e detalhe. Em aplicações discretas podem ocorrer distorções de borda indesejáveis, os quais, neste trabalho, foram contornados empregando-se a técnica do preenchimento com zeros (Walker, 1999).

Somado à decomposição *wavelet*, a ferramenta de avaliação da similaridade entre as famílias *wavelet* e os dados avaliados foi a correlação cruzada.

### 2.1 Correlação cruzada

A correlação usa dois sinais para produzir um terceiro que é chamado correlação cruzada (Smith, 1999). A correlação é utilizada, principalmente, para medir a similaridade entre dois sinais. Na forma discreta é dada por (4):

$$c_i = \sum_{j=-\infty}^{\infty} f_{j-i} g_i = \sum_{j=-\infty}^{\infty} f_j g_{i+j} \quad (4)$$

Em que  $c_i$  é um vetor de correlação cruzada,  $f$  e  $g$  são sinais a serem correlacionados transladando-se em  $i$  e  $j$  que  $\in \mathbb{Z}$ . O valor da correlação cruzada é maximizado quando um sinal é alinhado com as mesmas características do outro sinal, o que significa que um pico de correlação cruzada ocorrerá sempre que os dois sinais se parecerem muito (Smith, 1999).

## 3. METODOLOGIA

Como material para o desenvolvimento das técnicas, foi adotada uma base de dados constituída pelo registro de consumo de 2566 clientes, de uma distribuidora de energia, registrados durante 60 meses, divididos em duas classes:

- i. Cliente com conexão irregular (CCI), 1666 registros;
- ii. Cliente com conexão regular (CCR), 900 registros.

A situação de regularidade dos clientes é conhecida de antemão e foi validada durante a construção da base de dados.

Visando permitir a classificação automática dos registros presentes na base de dados foram propostas duas metodologias descritas a seguir. Na primeira, a composição se dá em três etapas, a saber: pré-processamento; aplicação da TW e a classificação dos clientes a partir dos subsinais obtidos com a TW. Na segunda, há as etapas de pré-processamento e classificação dos clientes, não passando pela TW. Um diagrama de blocos contendo a representação da primeira etapa das técnicas propostas é apresentado na Fig. 1. Vale ressaltar, também, que foram feitos experimentos de treino e testes, os experimentos de treino fazem uso de 70% da base de dados disponível para cada tipo de cliente, comercial ou residencial e que os testes são feitos com os 30% restantes. Além disso, cada experimento foi repetido 10 vezes.

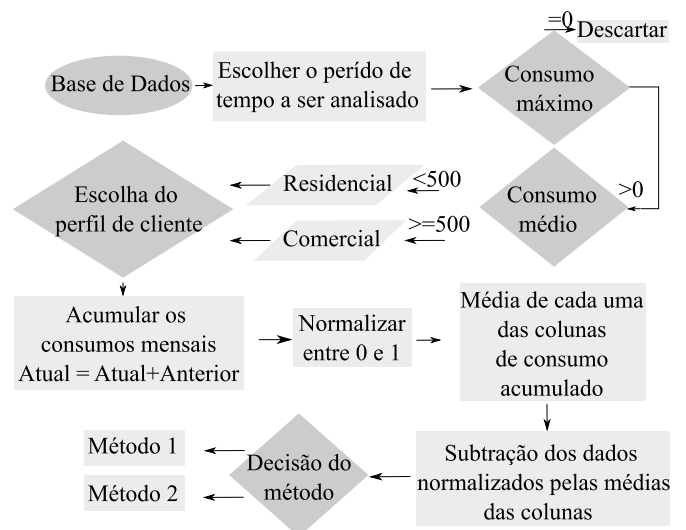


Fig. 1 Metodologia de pré-processamento da base de dados.

As etapas de pré-processamento, na Fig. 1, consistem em:

Classificar os clientes com consumos médios acima ou iguais a 500 kWh/mês ao longo dos 60 meses como comerciais (CC)

e aqueles com consumo médio inferior àquele patamar como residenciais (CR);

Escolher o período em que se deseja analisar os dados, entre 18 e 60 meses, e descartar da base de dados os clientes que apresentam inconsistências, como por exemplo um consumo máximo de 0 kWh/mês no período de observação;

Garantir a igualdade de amostras entre as classes: como o número de clientes divididos entre as duas classes não é originalmente igual na base de dados, para garantir a igualdade entre números de registros por classe nas etapas posteriores, sorteiam-se aleatoriamente números iguais de CCR e CCI;

Acumular os consumos mês a mês de cada cliente, o que resulta em um número de colunas de dados acumulados igual ao período em análise, a soma é dada seguindo (5):

$$CAA = CA + CMA \quad (5)$$

Em que *CAA* é o consumo atual acumulado, *CA* é o consumo acumulado e *CMA* é o consumo do mês atual.

Normalizar os registros no intervalo de 0 a 1;

Remover a sazonalidade periódica: este passo garante a remoção da sazonalidade do consumo, evitando que características devidas ao perfil de consumo sazonal não sejam confundidas com fraudes (a exemplo do aumento do consumo nos meses de verão e redução do consumo no inverno):

- Calcular a média do acúmulo normalizado dos consumos mensais de todos os clientes.
- Subtrair dos dados normalizados as médias de consumos acumulados normalizados.

O resultado obtido após a etapa de pré-processamento foi utilizado como entrada para a aplicação de uma TWD que forneceu como resultado vetores com os coeficientes de aproximação e detalhe que servirão de entradas para o classificador. Contudo, para a realização de uma decomposição que maximize a captura das características de interesse dos sinais, foi preciso escolher adequadamente a *wavelet* mãe.

### 3.1 Escolha da família wavelet

A fim de detectar-se a TW adequada à finalidade de fornecer características importantes do sinal pré-processado ao classificador, fez-se necessário aplicar uma ferramenta capaz de determinar o grau de similaridade entre os sinais. Portanto, calculou-se a correlação cruzada entre o vetor de suporte da *wavelet*, que é uma amostragem da *wavelet* mãe, e o sinal resultante do pré-processamento (Ngui et al. 2013). Neste trabalho, o parâmetro de comparação determinante para a escolha da *wavelet* foi a maior soma das energias resultantes da correlação cruzada. A energia de um sinal é dada por (6).

$$E_{total} = \sum_{i=1}^n c_i^2 \quad (6)$$

Em que  $c_i$  representa uma amostra discreta da operação de correlação cruzada e estas amostras variam de 1 até  $n$ . Em que  $n$  é o número de amostras do sinal.

A *wavelet* que apresentou maior energia de correlação foi a utilizada para decompor os sinais pré-processados. Então, os sinais decompostos pela TW foram para um algoritmo de classificação por limiar, desenvolvido para a finalidade deste trabalho, este procedimento é o método de classificação 1.

### 3.2 Subsinal e nível de decomposição

Após a definição da família *wavelet* a ser utilizada, deve-se escolher qual o subsinal e a escala de decomposição mais adequada ao método. Neste trabalho, o nível de decomposição foi definido a partir do método de classificação, ou seja, o nível de decomposição que sensibilizou melhor o algoritmo de limiarização foi o escolhido.

A medida da sensibilização é dada pela taxa de acertos (8) e de detecção de fraudes (9), ou seja, o nível de decomposição em que estes índices foram os maiores foi o escolhido e é dado por (7):

$$P = TA (\%) \times TDF (\%) \quad (7)$$

Em que o  $P$  é o produto que é o resultado da multiplicação entre a  $TA$  (taxa de acertos) e a  $TDF$  (taxa de detecção de fraudes). Ambas em percentual.

### 3.3 Método de classificação 1

A primeira metodologia de classificação fez uso da TW como mecanismo de extração de características latentes aos sinais pré-processados. As características advêm dos coeficientes resultantes da decomposição *wavelet*. O procedimento é apresentado na Fig. 2.

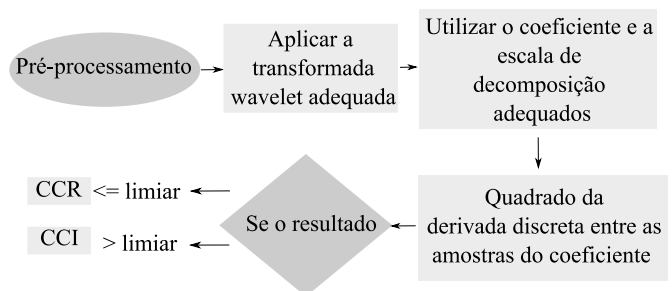


Fig. 2 Método de classificação 1.

Após a dizimação do sinal por dois, ocorrida o número de vezes igual ao nível de decomposição, aplica-se a derivada discreta entre as amostras do sinal resultante da decomposição anterior e eleva-se este resultado ao quadrado, a fim de não haver dados negativos. Então aplica-se uma detecção por limiar em que o quadrado da derivada local que for maior que este limiar classifica todo o sinal como CCI e, em casos em que seja menor ou igual a este limiar classifica o sinal como CCR. Outra metodologia foi proposta com a finalidade de comparação com a metodologia 1, a metodologia 2, sem o uso da TW.

### 3.4 Método de classificação 2

Após o desenvolvimento da primeira técnica, alguns testes foram realizados, a fim de perceber se trechos do algoritmo, sem a TWD, teriam potencial para classificação de clientes.

Esta metodologia não utiliza a TWD como forma de extração de características que diferem os CCI dos CCR. O método é apresentado na Fig. 3.

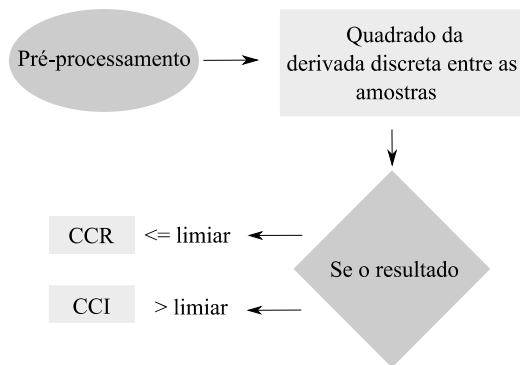


Fig. 3 Método de classificação 2.

Neste método, a derivada foi aplicada diretamente ao sinal pré-processado e, novamente, avaliada quanto ao limiar. Se a derivada local que for maior que este limiar, classifica-se todo o sinal como CCI e em casos em que seja menor ou igual a este limiar, classifica-se o sinal como CCR.

E, por fim, os conceitos de avaliação quantitativa dos métodos avaliados neste artigo foram as taxas de acerto e de detecção de fraudes.

### 3.5 Cálculo da taxa de acertos

A taxa de acertos (TA) foi uma das métricas utilizadas para determinar a eficácia dos métodos e é dada por (8).

$$TA = \frac{NACCR + NACCI}{T} \quad (8)$$

Em que  $NACCR$  é o número de acertos de clientes com conexão regular e  $NACCI$  é o número de acertos de clientes com conexão irregular e  $T$  representa o número total de clientes avaliados.

### 3.6 Cálculo da taxa de detecção de fraudes

A taxa de detecção de fraudes (TDF) é importante para determinar a aptidão do método para detectar verdadeiros clientes com conexões irregulares e é dada por (9).

$$TDF = \frac{NACCI}{TCCI} \quad (9)$$

Em que  $NACCI$  é o número de acertos de clientes com conexão irregular e  $TCCI$  é o total de clientes com conexão irregular que foram postos para análise.

## 4. RESULTADOS E ANÁLISES

### 4.1 Escolha da wavelet

No trabalho de Ngui et al. (2013) apontam-se algumas formas para determinar quais funções *wavelet* mãe são mais adequadas para realização de determinada tarefa. Dentre elas, destaca-se a correlação cruzada entre o sinal a ser decomposto e o vetor de suporte da *wavelet*.

As três famílias *wavelet* mais recorrentes nas referências levantadas para a execução deste trabalho foram Haar (Tsui and Basir 2006, Paul et al. 2021), Doubechies (Santos 2019) e Coiflet (Al-Shayea 2020, Wang et al. 2004).

Dentro das famílias *wavelets*, foram escolhidas algumas *wavelets* mãe para fornecerem vetores de suporte a fim de serem correlacionados com o resultado do pré-processamento. As escolhidas foram Haar, Doubechies 4, 6, 8, 10, 12 e Coiflets 6 e 12. Optou-se pela escolha destas *wavelets* mãe devido a sua ocorrência nas referências consultadas e sua facilidade de implementação.

A *wavelet* escolhida foi a que obteve maior energia no processo de correlação entre os vetores de suporte normalizados e o pré-processamento também normalizado. O resultado é apresentado na Tabela 1.

Tabela 1. Energia das correlações cruzadas para cada família *wavelet*

Wavelet mãe	Energia média da correlação	Soma das energias da correlação
Haar	0,3186	573,4506
Doubechies 4	0,0616	110,9028
Doubechies 6	0,0449	80,7915
Doubechies 8	0,0488	87,7523
Doubechies10	0,0437	78,6942
Doubechie12	0,0387	69,5736
Coiflet6	0,0963	173,3885
Coiflet12	0,0036	6,4851

Observa-se que as energias das correlações cruzadas foram diferentes a depender da *wavelet* mãe. Obteve-se uma maior energia contida na correlação cruzada advinda da TW Haar, com um valor médio de 0,3186 e uma soma de todas as amostras de correlação igual a 573,4506. Inclinando-se, portanto, para a escolha da TW Haar para a análise dos dados, haja vista que além da conservação da energia, esta é a transformada mais simples de ser implementada e suas operações demandam um menor custo computacional.

### 4.2 Subsinal e nível de decomposição

O subsinal utilizado foi definido de acordo com a sensibilização do método de classificação por limiar, neste caso, o coeficiente e o nível de decomposição foram definidos pelo produto entre a taxa de acertos e a taxa de detecção de fraudes.

Tabela 2. Subsinal de aproximação e nível de decomposição.

Nível de decomposição	Produto clientes residenciais	Produto clientes comerciais
1°	955,9	1105,9
2°	4220,6	4320,6
3°	4846,6	4923,4
4°	3750,6	3995,7

Devido aos produtos obtidos, verificou-se que o nível de decomposição do subsinal a ser utilizado é o terceiro, a

exemplo, obteve-se-se 4923,4, que resulta da multiplicação entre 67,2 e 73,3 que são, respectivamente, as taxas de acerto e detecção de fraudes do cliente comercial. O mesmo raciocínio pode ser estendido aos outros produtos.

Os resultados se mostraram melhores empregando-se os subsinais de aproximação. A mesma técnica aplicada aos subsinais de detalhes originou taxas de acerto e detecção de fraudes próximas a 50%. Este resultado alinha-se à física do fenômeno, haja vista que os coeficientes de aproximação tendem à média e os coeficientes de detalhe tendem a flutuações. Dentre os métodos encontrados, o que apresentou-se como mais computacionalmente eficaz (e com boa incidência de acerto) passa pela realização dos coeficientes de aproximação da primeira escala, seguida de uma classificação por limiar. Análises multiresolucionais baseadas na TWD também apresentaram resultados satisfatórios, mas com custo computacional mais elevado.

#### 4.3 Percentuais de acerto

Nesta subseção é apresentada o resultado dos métodos para diversas janelas de tempo a considerar: 18, 24, 30, 36, 42, 48, 54 e 60 meses. A taxa percentual de acertos das metodologias propostas foi dividida em dois gráficos: na Fig. 4 apresenta-se o percentual de acertos considerando o perfil de clientes residenciais, e na Fig. 5 o percentual de acertos considerando o perfil de clientes comerciais. Conforme apresentado na metodologia, os experimentos foram repetidos 10 vezes cada e das Fig 4 à Fig 7 são apresentadas as médias desses experimentos.

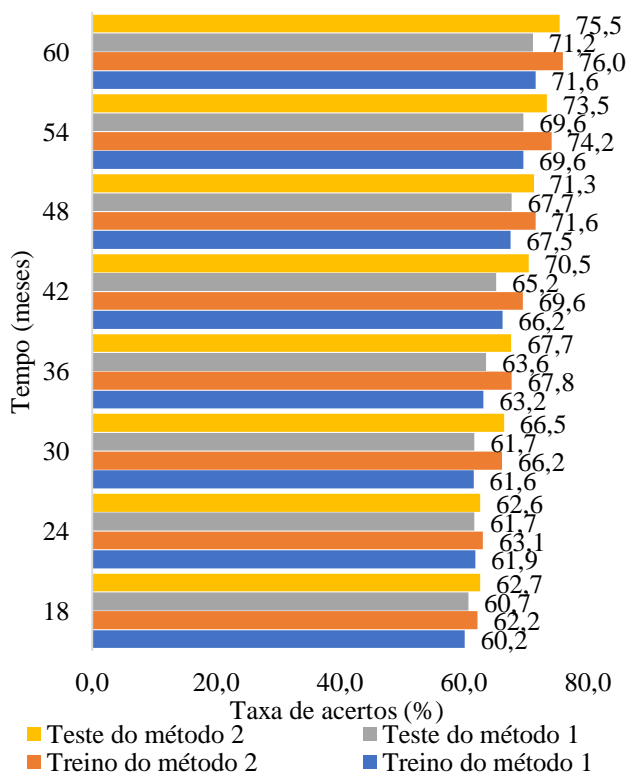


Fig. 4 Percentual de acertos para clientes residenciais.

A partir da Fig. 4, observa-se que a taxa de acertos vai aumentando conforme a janela de tempo avaliada aumenta.

Observa-se também que a taxa de acertos do método 1 em treino varia de 60,2% em 18 meses a 71,6% em 60 meses enquanto em teste os percentuais variam entre 60,7% e 71,2% no mesmo período.

Para o método 2 observam-se taxas de acertos, geralmente, maiores que o primeiro método. A variação durante o treinamento parte de 62,2% em 18 meses até 76% em 60 meses. Durante o teste, as taxas variam entre 62,7% e 75,5% na mesma janela temporal. Para o quesito de taxa de acertos, o método 2 mostrou-se mais assertivo 4,3% em 60 meses em comparação com o método 1.

Na Fig. 5, são apresentadas as taxas de acerto para clientes de perfil comercial.

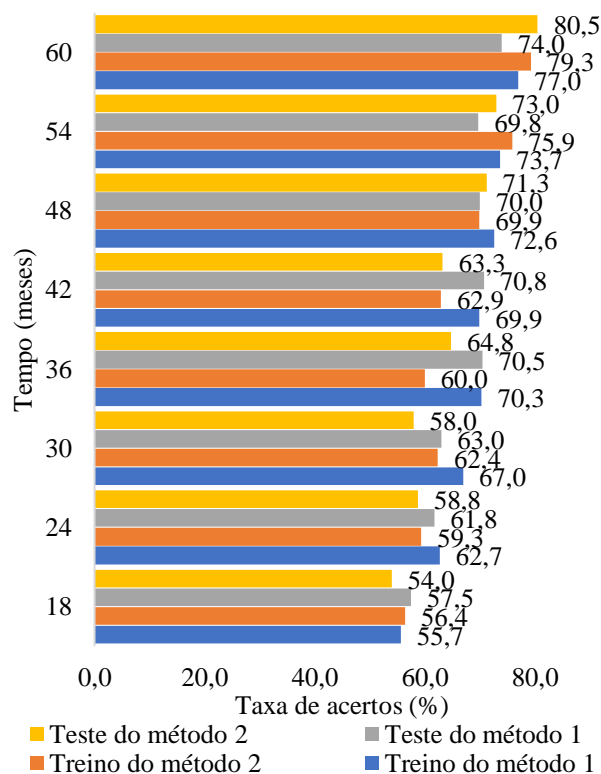


Fig. 5 Percentual de acertos dos clientes comerciais.

A partir da Fig. 5, observa-se que a taxa de acertos vai aumentando conforme a janela de tempo avaliada aumenta. Observa-se também que a taxa de acertos do método 1 em treino varia de 55,7% em 18 meses a 77% em 60 meses enquanto em teste os percentuais variam entre 57,5% e 74% no mesmo período.

Para o método 2 observam-se taxas de acertos, geralmente, maiores que o primeiro método. A variação durante o treinamento parte de 56,4% em 18 meses até 79,3% em 60 meses. No teste, as taxas variam entre 57,5% e 74% na mesma janela temporal. O segundo método mostrou-se 6,5% mais assertivo no período de 60 meses comparado com o método 1. Quando a TW é aplicada, a cada iteração ocorre a dizimação do sinal original por dois. O que significa que há sempre menos dados a serem avaliados no primeiro método que no segundo. Somado a isso, sabendo-se que a taxa de acertos leva em conta o acerto de classificação dos CCR e CCI que realmente o são. Logo, *a priori*, o método 2 é um método

melhor para detectar CCR que CCI, porém não impede que o método 1 seja aplicado sozinho ou em conjunto com ele.

#### 4.4 Taxa de detecção de fraudes

A taxa de detecção de fraudes, assim como a taxa de acertos é representada em duas figuras. Na Fig. 6 é apresentada a taxa de detecção de fraudes em clientes residenciais, e na Fig. 7 apresentam-se os mesmos resultados para clientes comerciais.

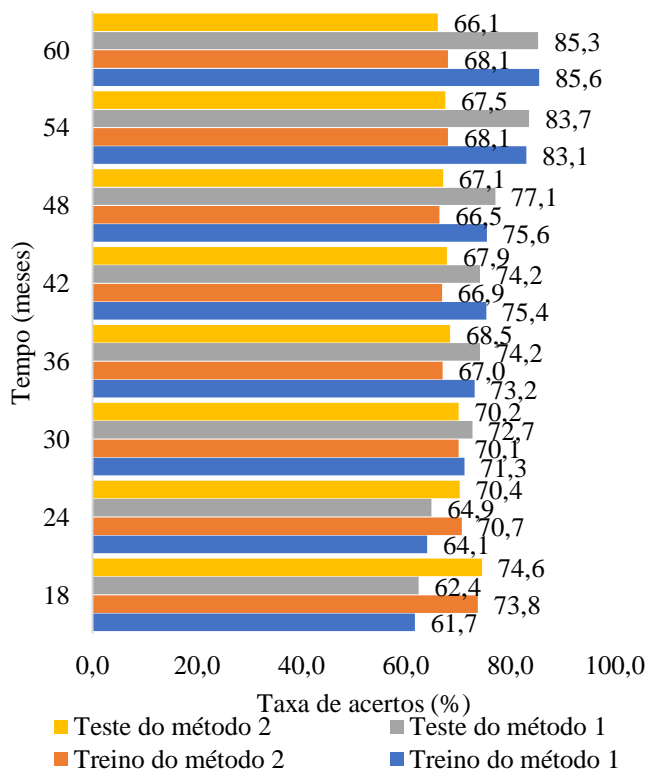


Fig. 6 Taxa de detecção de fraudes para clientes residenciais.

Para os clientes residenciais, observa-se que para períodos maiores o método 1 obteve taxas de detecção de fraudes bem maiores que o método 2. No método 1 em treinamento, as taxas variam de 61,7% a 85,6% no período de 18 a 60 meses. Na fase de testes as taxas variam de 62,4% a 85,3% no mesmo período.

Já para o método 2, o que acontece é uma diminuição das taxas de detecção de fraudes. Para o treinamento do método, a taxa varia de 73,8% a 68,1% de 18 a 60 meses respectivamente. Nos testes, as taxas excursionam entre 74,6% e 66,1% no período de 18 a 60 meses.

Neste caso, ficou claro que as tendências das taxas de detecção de fraudes têm comportamentos inversos, enquanto o método 1 tende a aumentar a eficiência na detecção de fraudes quanto maiores forem as janelas de tempo, o método 2 tende a maior eficiência enquanto a análise é feita em 18 meses e, a partir de 36 meses, as taxas de acerto se mantiveram praticamente constantes em torno de 67%.

Para os clientes com perfil comercial as tendências apresentadas anteriormente para os residenciais são novamente visualizadas, de maneira mais discreta, porém ainda visíveis, o que se verifica na Fig. 7.

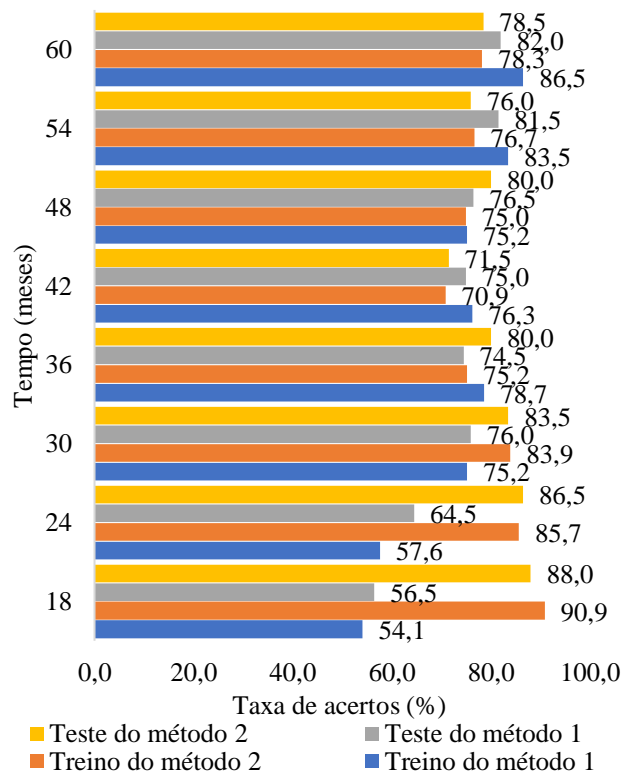


Fig. 7 Taxa de detecção de fraudes para clientes comerciais.

No treinamento, foi obtido com o método 1, 54,1% de taxa de detecção de fraudes em 18 meses e 86,5% em 60 meses. No teste, no mesmo período, as taxas variam entre 56,5% e 82%. O método 2 em treino apresentou a TDF de 90,9% em uma janela de observação de 18 meses e 78,3% em 60 meses. Para os testes da metodologia, 88% em 18 meses e 78,5% em 60 meses.

O coeficiente avaliado na aplicação da TW é o de aproximação e este, no geral, é a soma de duas amostras consecutivas. E, conforme os dados pré-processados, os CCIs são mais destacados por seu comportamento bem mais abrupto, e portanto, a soma dessas amostras são mais energéticas e, carregando mais informação e sensibilizam melhor o limiar de detecção de fraudes. Remontando a boa aplicabilidade do primeiro método, mas não excluindo o uso do segundo em paralelo ou sozinho.

## 5. CONCLUSÕES

A base de dados fornecida foi analisada e os perfis de clientes foram traçados, permitindo-se a rotulação dos dados. Dois métodos de detecção de fraudes em redes de distribuição foram propostos e avaliados. A transformada *wavelet* mostrou forte potencial para a classificação de clientes com conexão irregular. A primeira metodologia proposta deixa claro que há informações contidas nos comportamentos dos clientes irregulares que são independentes da sazonalidade e que são evidenciadas por meio da decomposição *wavelet*. Os melhores resultados obtidos por esta metodologia foram: 71,2% e 74% de percentuais de acertos no período de 60 meses, respectivamente, para clientes residenciais e comerciais. As taxas de detecção de fraudes foram de 85,3% para clientes

residenciais e 82% para clientes comerciais no período de 60 meses.

O segundo método foi promissor, primeiramente por obter taxas de acertos maiores que o primeiro. No entanto, na detecção de fraudes, mostrou-se menos assertivo que o primeiro, apontando na direção da complementariedade entre as metodologias. Cabe, portanto, em análises futuras o uso em paralelo das duas técnicas a fim de obter-se os melhores índices de acerto e detecção de fraudes. Os melhores resultados obtidos pelo segundo método foram 75,5% e 80,5% de acertos no período de 60 meses para, respectivamente, clientes residenciais e comerciais. As taxas de detecção de fraudes foram de 66% para clientes residenciais e 78,5% para clientes comerciais no período de 60 meses.

No geral, os experimentos foram todos condizentes e concluem que, quanto ao método 1, a TW é também uma boa ferramenta para a extração de características latentes aos clientes e conforme os resultados apresentados, observou-se que a família *wavelet* mais adequada à análise dos dados processados é a *wavelet* Haar, devido à maior energia advinda da operação de correlação cruzada, além do baixo custo computacional.

#### AGRADECIMENTOS

Os autores gostariam de agradecer o apoio da Coordenadoria de Pós-Graduação em Engenharia Elétrica (COPELE) da UFCG e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte financeiro.

#### REFERÊNCIAS

- Almeida, M. A. S., Oliveira, W. C. and Dantas, P. R. (2006), Redução de Perdas de Energia Elétrica na COELBA Estratégias e Resultados Pós-Privatização, *CIER*, pp. 19-22.
- Al-Shayea, T. K., Mavromoustakis, C. X., Batalla, J. M., Mastorakis, G., Mukherjee, M. and Pallis, E. (2020). A Novel Gaussian in Denoising Medical Images with Different Wavelets for Internet of Things Devices. Taipei, Taiwan., *IEEE*, pp. 1-6.
- ANEEL (2021), Perdas de Energia Elétrica na Distribuição, *Brasília*: s.n.
- Avila, N. F., Figueroa, G. and Chu, C., (2018). NTL detection in electric distribution systems using the maximal overlap discrete wavelet-packet transform and random undersampling boosting, *IEEE Transactions on Power Systems*, 33(6), pp. 7171-7180.
- Barros, R.M.R., (2021). Advanced analytics aplicado à gestão da perda não técnica de energia em sistemas elétricos de distribuição, Ph. D. Thesis, Campina Grande University, Campina Grande.
- Burrus, C. S., Gopinath, R. and Guo, H., (2013), *Wavelets and Wavelet Transform*, Houston, Texas: *Connexions*.
- Castro, F. N., (2012), Aplicação da Transformada Wavelet na Classificação do Estado Operacional de Câmaras de Disjuntores, Master's dissertation, *Campina Grande Federal University*, Campina Grande.
- Leite, J. B. and Mantovani, J. R. S. (2016). Detecting and Locating Non-technical Losses in Modern Distribution Networks. *IEEE Transactions on Smart Grid*, 9(2), pp. 1023-1032.
- Menezes, D. P.(2017). Assistente Baseado Em Redes Neurais Artificiais Para Indicação De Clientes Suspeitos De Fraude Em Sistema De Distribuição De Energia Elétrica De Baixa Tensão, Bachelor's degree work, *Campina Grande Federal University*, Campina Grande.
- Ngui, W. K., Leong, M. S., Hee, L. M. and Abdelrhman, A. M. (2013). Wavelet Analysis: Mother Wavelet Selection Methods. *Applied Mechanics and Materials*, pp. 953-958.
- Paul, T., Vainio, S. and Roning, J. (2021). Clustering and classification of virus sequence through music communication protocol and wavelet transform, *Elsevier Genomics*, pp. 778-784.
- Penin, C. A. S.(2008). Combate, Prevenção e Otimização das Perdas Comerciais de Energia Elétrica, Ph.D. Thesis, São Paulo University, São Paulo.
- Punmiya, R., Choe, S. (2019). Energy Theft Detection Using Gradient Boosting Theft Detector With Feature Engineering-Based Processing. *IEEE Transaction on Smart Grid*, pp. 2326-2329.
- Raggi, L.M.R. ,Trindade, F.C.L, Cunha, V.C., Freitas, W. (2020). Non-technical Loss Identification by Using Data Analytics and Customer Smart Meters. *IEEE Transaction on Power Delivery*,pp. 2700-2710.
- Ramos, C. C. O. (2014), Caracterização de Perdas Comerciais em Sistemas de Energia Através de Técnicas Inteligentes, Ph.D. Thesis, São Paulo University, São Paulo.
- Santos, I. A. (2019), Disjuntores PVO: Diagnóstico Via Resistência Dinâmica de Contatos Empregando Transformada Wavelet Packet, Bachelor's degree work, Federal University of Sergipe, São Cristóvão.
- Secretaria de Infraestrutura do Estado de São Paulo (2020), Anuário de Energéticos por Municípios do Estado de São Paulo 2020 - ano base 2019, São Paulo.
- Silva, J. M. P. L. E, (2017). Aplicação da Transformada Wavelet Packet e Redes Neurais Artificiais para Monitoramento de Condição de Motores de Indução, Ph.D. Thesis, Pernambuco Federal University, Recife: s.n.
- Silveira, C. G. A. (2017). Identificação de Fraudes em Energia Elétrica, Bachelor's degree work, Federal University of Sergipe, São Cristóvão.
- Smith, S. W., (1999). The Scientist and Engineer's Guide to Digital Signal Processing, 2nd ed., San Diego California, *Technical Publishing*.
- Tsui, P. P. and Basir, O. A., (2006). Wavelet basis selection and feature extraction for shift invariant ultrasound foreign body classification.. *Elsevier Ultrasonics*, June, v.45, pp. 1-14.
- Walker, J. S., (1999). *A primer on Wavelets and their Scientific Applications*. 1st ed. New York, *Chapman & Hall/CRC*.
- WANG, S.Y., Liu, X., Yianni, J., Aziz, T.Z. and Stein F. J. (2004). Extracting burst and tonic components from surface electromyograms in dystonia using adaptive wavelet shrinkage. *Journal of Neuroscience Methods*, April, 139(2), pp. 177-184.