

Princípios da Detecção de Perdas Não Técnicas em Rede de Distribuição Usando a Análise de Componentes Principais

Rodrigo F. G. Sau^{1,2}, Madson C. de Almeida¹

¹Departamento de Sistemas e Energia, FEEC, Unicamp, SP
Av. Albert Einstein, 400 - Distrito Barão Geraldo, CEP: 13083-852, Campinas-SP

²Instituto Federal de São Paulo, IFSP
R. Maria Cinto de Biaggi, 130 - Jardim Santa Rosália, CEP: 18095-410, Sorocaba - SP,
E-mails: rodrigof@dssee.fee.unicamp.br, madson@dssee.fee.unicamp.br

Abstract: The detection of non-technical losses in the distribution networks is a very important issue for utilities. With the evolution of network monitoring systems and the installation of advanced metering infrastructures, new opportunities for the development of methods to mitigate this problem are available. This work proposes the principle of a method for detecting and identifying suspect customers of metering tampering using principal component analysis. Based on simulations results, by using measurements over time, throughout the day, it is possible to detect intermittent non-technical losses.

Resumo: A detecção de perdas não técnicas nas redes de distribuição é um tema muito importante para as distribuidoras. Com a evolução dos sistemas de monitoramento das redes e a inclusão de infraestrutura avançada de medição, abrem-se novas oportunidades para o desenvolvimento de métodos para mitigar esse problema. Neste trabalho é proposto o princípio de um método de detecção e identificação de clientes suspeitos de fraudar os sistemas de medição usando a análise de componentes principais. Por meio de simulações, verificou-se que é possível aproveitar as medições no tempo, ao longo do dia, para detectar perdas não técnicas intermitentes.

Keywords: non-technical losses; energy meter tampering; distribution systems; principal component analysis; outlier detection.

Palavras-chaves: perdas não técnicas; fraudes em sistemas de medição; sistemas de distribuição; análise de componentes principais; detecção de anomalias.

1. INTRODUÇÃO

Um problema típico de redes de distribuição é a detecção de perdas não técnicas, também conhecidas como perdas comerciais. Essas perdas são devidas a erros de leitura, falhas em medidores, erros no faturamento e, especialmente, a fraudes e furtos de energia elétrica. De acordo com os dados da ANEEL (ANEEL, 2019), as perdas não técnicas no Brasil representam 6,6% da energia injetada, sendo que na região norte do país ultrapassam 20%.

Os sistemas de distribuição têm passado por um processo de modernização, incluindo uma infraestrutura avançada de medição AMI, do inglês *Advanced Metering Infrastructure*. Esses novos dispositivos permitem a aplicação de técnicas avançadas de gerenciamento das redes. Este artigo propõe os princípios de um método para a detecção de perdas não técnicas usando essa infraestrutura mais moderna de medição. O desafio é detectar medições muito fora dos padrões, que levam à identificação de possível perdas comerciais. Cabe ressaltar que, diferentemente da transmissão, as redes de distribuição não possuem um elevado nível de redundância das medições e, mesmo nas redes mais modernas, não é esperado

que haja medidores ao longo das redes de baixa tensão. Neste trabalho é considerado que há medidores de potência e tensão nos consumidores e que diversas amostras de medição são obtidas ao longo do dia.

Diversos trabalhos têm tratado da detecção de perdas não técnicas. Nos últimos anos, diferentes métodos classificadores usando aprendizado de máquina têm sido propostos. Em (Ghori, et al., 2020) vários desses métodos têm seu desempenho comparado. Essas técnicas são muito promissoras à medida em que mais dados estão disponíveis; tem havido um aumento do número de medidores eletrônicos dotados de comunicação nos consumidores e a taxa de amostragem das medidas deve aumentar. Entretanto, esses métodos são propensos a falsos alarmes (Messinis & Hatzargyriou, 2018), o que pode implicar em elevados custos de inspeção.

Outra forma de detectar as perdas não técnicas é por meio de métodos que usam dados da rede elétrica, especialmente os baseados em fluxo de carga e estimação de estado. No caso dos métodos baseados em fluxo de carga, normalmente é realizado o balanço de potência e, através de comparação de tensão, são realizadas a detecção da perda e a identificação (localização) do consumidor suspeito (Ferreira, et al., 2020), (Weckx, et al.,

2012), (Kadurek, et al., 2010). Esses métodos são muito sensíveis a erros nas medições de tensão. Já os trabalhos baseados em estimação de estado fazem uso de técnicas de detecção de erros grosseiro ou análise da variância para detectar as perdas (Huang, et al., 2013). Eles são muito robustos, mas o desafio está na instalação de medidores para garantir a observabilidade e a redundância das medidas. São necessários, em alguns casos, medidores ao longo das linhas (Bretas, et al., 2020) ou até mesmo PMUs (Souza, et al., 2020).

Neste trabalho é apresentado o princípio de um método baseado nas medições de potência e tensão nos consumidores, sem usar técnicas de análise de rede, e que também não faz uso de aprendizado de máquina supervisionado. As medidas são transformadas usando a Análise de Componentes Principais (PCA), do inglês *Principal Component Analysis*, e as perdas não técnicas são tratadas como medidas discrepantes (anômalas). Há trabalhos que usam a PCA como técnica de redução de dimensionalidade, mas que dependem da definição de um limiar empírico (Singh, et al., 2019) ou a detecção da perda é baseada no número de grupos de um algoritmo de agrupamento (Konstantinos & Georgios, 2019). Na última referência, os consumidores suspeitos são detectados por não se adequarem aos padrões pré-estabelecidos. Neste trabalho é proposto o uso de um limiar baseado na função densidade de probabilidade de cada grupo (consumidores normais e com a perda não técnica), sendo que a identificação dos centros de massa das densidades de probabilidades é feita usando a técnica de agrupamento *k-means*, que já foi testada na detecção de perdas não técnicas (Zhang, et al., 2020), (Jindal, et al., 2020). A implementação computacional é simples e não há dependência de análise de dados históricos.

O objetivo deste trabalho é mostrar que é possível detectar perdas não técnicas nas redes de distribuição usando diversas medições ao longo do dia, que permitem a identificação de amostras de medidas discrepantes. Com isso, é possível detectar fraudes intermitentes, as quais ocorrem em períodos específicos do dia. Há casos em que um consumidor reduz drasticamente seu consumo, por exemplo, fazendo uma conexão em paralelo com o medidor de faturamento. Nesses casos, a expressiva redução do consumo é uma sinalização de que houve adulteração no medidor. Entretanto, existem casos mais sofisticados em que os medidores de faturamento são adulterados de modo a subnotificar o consumo em parte do dia (Azevedo, et al., 2018), por exemplo no horário de ponta. É neste segundo caso que o método apresentado é aplicável.

O artigo está organizado da seguinte forma: na segunda seção são apresentados conceitos básicos da PCA. Na sequência é apresentada a metodologia, que inclui resultados numa rede de três barras. Na seção 4 estão resultados de testes do método proposto realizados numa rede com dez barras e, por fim, as conclusões estão na quinta seção.

2. ANÁLISE DE COMPONENTES PRINCIPAIS

A PCA é uma técnica usada para a reduzir a dimensão de grandes sistemas nos quais as variáveis estão relacionadas (Jolliffe, 2002). Para que isso seja possível, as variáveis do sistema estudado são transformadas em outras, as quais são chamadas componentes principais (PC), do inglês *Principal*

Components, que são ortogonais. A transformação também ordena os componentes de modo que, de forma decrescente, cada componente tenha a maior variância.

A Fig. 1 faz uma representação gráfica da modificação feita nas variáveis do sistema a ser estudado de acordo com a PCA. Nesse exemplo, há vinte conjuntos de medidas com duas variáveis, conforme a Tabela 1. Nessa tabela estão apresentadas as variáveis originais centralizadas (variáveis obtidas de medições) e os PCs.

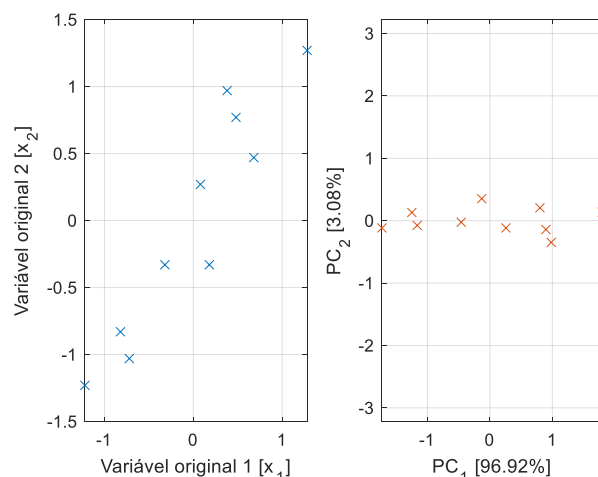


Fig. 1 Representação gráfica da análise de componentes principais [adaptado de (Jolliffe, 2002)].

Tabela 1. Variáveis originais e PCs

Conjunto de Medição	Variáveis Originais		Componentes Principais	
	X_1	X_2	PC_1	PC_2
1	2,50	2,40	0,8006	0,2059
2	0,60	0,70	-1,7285	-0,1160
3	2,20	2,90	0,9818	-0,3483
4	1,90	2,20	0,2564	-0,1164
5	3,10	3,20	1,7981	0,1352
6	2,30	2,70	0,8962	-0,1417
7	2,00	1,60	-0,1314	0,3522
8	1,00	1,10	-1,1643	-0,0758
9	1,50	1,60	-0,4590	-0,0256
10	1,10	0,90	-1,2499	0,1307
11	2,50	2,40	0,8006	0,2059
12	0,60	0,70	-1,7285	-0,1160
13	2,20	2,90	0,9818	-0,3483
14	1,90	2,20	0,2564	-0,1164
15	3,10	3,20	1,7981	0,1352
16	2,30	2,70	0,8962	-0,1417
17	2,00	1,60	-0,1314	0,3522
18	1,00	1,10	-1,1643	-0,0758
19	1,50	1,60	-0,4590	-0,0256
20	1,10	0,90	-1,2499	0,1307

Analisando a Fig. 1, observa-se que a variância dos dados originais é grande em relação às duas variáveis, e também que

as duas variáveis estão fortemente correlacionadas. Ao lado direito da figura estão as pontuações dos PCs, calculadas por meio da multiplicação das variáveis originais por uma matriz de vetores ortogonais que concentra a variância nos primeiros PCs. Como neste exemplo há somente dois PCs, toda a variância das variáveis originais está apresentada no gráfico à direita da Fig. 1. Contudo, ela está mais concentrada no primeiro PC, que tem 96,92% da variância total. Nas medidas originais, ao desprezarmos uma variável, muita informação é perdida. Já nos PCs, ainda que desprezemos o segundo componente, temos uma boa aproximação.

Como a PCA visa a redução da dimensão do problema estudado através de transformações lineares, precisamos de uma matriz que transforme as variáveis originais em PCs. Uma forma de obter essa matriz é calculando os autovetores da matriz de covariância das medições conforme (1), (Brunton & Kutz, 2019).

$$C \cdot V = V \cdot D \quad (1)$$

Em que C representa a matriz de covariância da matriz de medições X , os autovetores são representados por V , e D é uma matriz diagonal com os autovalores.

Os PCs são calculados pela multiplicação dos autovetores, chamados de carregamentos, pela matriz de dados X com a média \bar{X} subtraída, chamada B , conforme equação 2. Essa matriz centralizada B é a que está apresentada na Fig. 1 como variável original.

$$PC = B \cdot V \quad (2)$$

Em que:

$$B = X - \bar{X} \quad (3)$$

Os autovalores da matriz D representam as variâncias de cada PC. A partir desses valores são calculadas as variâncias percentuais apresentadas no gráfico à direita da Fig. 1.

Por fim, cabe ressaltar que, no exemplo dado nesta seção, não há nenhuma medida discrepante. Entretanto, o aspecto interessante do uso de PCA para detecção de perdas não técnicas é o fato de essas medidas se destacarem das restantes. Quando existe uma medida muito discrepante, ela pode ser detectada pela análise dos PCs, ao passo que a identificação da variável relacionada à medida atípica é feita pela análise dos carregamentos, os autovetores. Essas ideias estão apresentadas na próxima seção.

3. METODOLOGIA

Nesta seção o método proposto é descrito. A detecção e a identificação de perdas comerciais por meio da PCA serão apresentadas por meio de um exemplo numa rede de três barras.

2.1 Método de detecção de perdas

As redes de distribuição têm se modernizado e, através da AMI, será possível obter medidas dos consumidores ao longo

do dia. Neste trabalho são consideradas disponíveis as medidas de tensão e injeção de potência ativa e reativa em todos os consumidores da rede. Para a detecção de perdas comerciais, são necessárias diversas amostras de medidas obtidas ao longo do dia. O procedimento de teste do método segue a Fig. 2, sendo que os dados da rede têm origem no modelo barra-ramo e o cálculo do fluxo de carga foi realizado usando o método de Newton, implementado em Matlab.

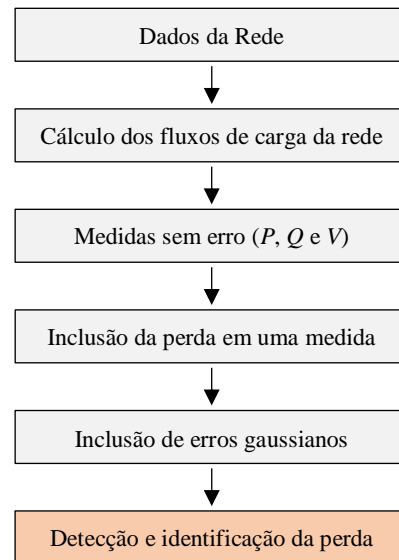


Fig. 2 Fluxograma de teste do método de detecção de perdas.

Seguindo a Fig. 2, inicialmente os dados da rede são carregados. Eles representam o modelo da rede e incluem as impedâncias das linhas e as injeções de potência em cada barra de carga. Na rotina de fluxo de carga são calculadas as tensões nas barras para cada amostra de medição. Por exemplo, se for usada uma medida por hora, são calculadas 24 amostras de medição para um dia. Os valores das magnitudes de tensão e das injeções de potência são usados como medidas perfeitas (sem erros). Então, a uma medida de injeção de potência é incluída a perda não técnica, aplicando-se um fator à medida sem erro. Na sequência, são incluídos erros gaussianos a todas as medidas de acordo com a precisão dos medidores.

Para a detecção, o conjunto de medições é organizado numa matriz (X). Nessa matriz, cada linha representa uma amostra do conjunto de medições e cada coluna representa uma grandeza medida. As variáveis medidas são a magnitude de tensão e as potências ativa e reativa. Com esses dados, são calculadas as pontuações dos PCs e seus carregamentos. A amostra que se destaca é considerada suspeita de conter a perda não técnica e o maior carregamento dessa amostra identifica a medida contendo a perda.

O exemplo a seguir ilustra o conceito. Nele é considerado que apenas uma amostra de medidas tem erro, ou seja, uma medida de injeção de potência possui perda não técnica em uma única amostra do dia. Na seção de resultados, os dados apresentados consideram mais de uma amostra das medidas com erro; é considerado que um consumidor adulterou seu consumo no horário de ponta.

2.2 Exemplo de aplicação da PCA numa rede de 3 barras

A Fig. 3 apresenta a rede de teste usada nesta seção. Há dois tipos de linhas, conforme a Tabela 2. As medições disponíveis estão indicadas na figura, sendo que as cargas nominais estão na Tabela 3. Para gerar uma curva de carga, as potências nominais são multiplicadas pelos fatores da Tabela 4 em função do horário.

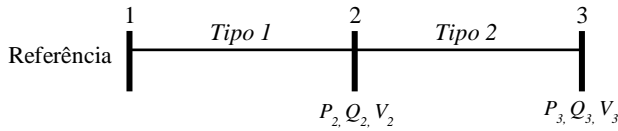


Fig. 3 Esquema de um sistema de potência de três barras.

Tabela 2. Impedância [pu] (base: 1 kVA e 220 V)

Linha	r	x
Tipo 1	0,70	0,80
Tipo 2	0,40	0,40

Tabela 3. Potências nominais nas barras de consumo [pu] (base: 1 kVA)

Tipo	Barra 2	Barra 3
P	0,50	1,00
Q	0,20	0,50

Tabela 4. Fatores aplicados às potências nominais em função do horário

Horário	1-6 h	7-9 h	10-17 h	18-21h	22-24 h
Fator	0,2	0,8	0,4	1,1	0,6

Com os dados das tabelas anteriores, a rotina de fluxo de carga calcula os conjuntos de medições para um dia (24 h). Neste exemplo é considerada uma amostra por hora, ou seja, 24 amostras de medições. Para os testes, foram usadas como medidas as magnitudes de tensão e as injeções de potência ativa e reativa nas barras dos consumidores (2 e 3). A essas medidas foram incluídos erros gaussianos relativos à exatidão dos medidores. Dessas, apenas a medida de injeção de potência ativa da barra 3 (medida 3), às 18 h, contém uma perda não técnica; foi considerado que a medida indica apenas 10% de sua potência real. A Tabela 5 apresenta esse conjunto de medidas.

Como foi calculada uma amostra de medição por hora, a perda está na 18ª medida. Aplicando a PCA nos valores de medição, foram obtidas as pontuações dos PCs e seus carregamentos. Esses valores estão apresentados no gráfico da Fig. 4. Para evitar a sobreposição de números na figura, somente o componente e carregamento relativos à perda foram numerados. Como há seis variáveis de medição, podem ser calculados até seis PCs. Contudo, somente os dois primeiros estão sendo analisados; a principal ideia da PCA é concentrar a variância nos primeiros componentes. Portanto, a Fig. 4 possui dois eixos que representam os dois PCs. Notar que a variância realmente se concentra nos primeiros PCs,

especialmente no primeiro. Neste exemplo o PC₁ possui aproximadamente 94% da variância total. Na figura também é mostrada uma elipse representando três desvios padrões (admitindo que as medidas seguem uma distribuição normal). Nota-se que, com exceção da amostra de medição 18, todas estão dentro do elipsoide. Em outras palavras, a medida fora do limite de três desvio padrões é inesperada, e, neste caso, indica a perda não técnica. Dessa forma, é possível detectar a perda, analisando as pontuações dos PCs.

Tabela 5. Medições para análise via PCA [pu]

Hora	V ₂	V ₃	P ₂	Q ₂	P ₃	Q ₃
1	1,0008	1,0031	-0,1003	-0,0399	-0,2027	-0,1000
2	1,0027	1,0040	-0,1011	-0,0399	-0,1980	-0,1010
3	0,9967	1,0045	-0,1004	-0,0401	-0,2003	-0,0995
4	0,9992	0,9995	-0,0995	-0,0402	-0,2017	-0,1015
5	0,9959	0,9974	-0,0996	-0,0398	-0,2005	-0,0993
6	0,9998	1,0024	-0,1011	-0,0400	-0,1998	-0,1011
7	0,9978	0,9973	-0,3971	-0,1608	-0,7899	-0,3981
8	1,0002	1,0022	-0,3969	-0,1602	-0,7925	-0,4019
9	1,0010	1,0037	-0,3952	-0,1582	-0,7912	-0,4020
10	1,0004	0,9934	-0,1988	-0,0793	-0,4023	-0,1993
11	1,0033	1,0017	-0,2004	-0,0802	-0,4001	-0,1989
12	0,9955	1,0032	-0,2012	-0,0792	-0,3996	-0,2003
13	0,9992	0,9975	-0,2008	-0,0804	-0,4011	-0,1997
14	1,0054	0,9990	-0,1990	-0,0809	-0,3994	-0,2005
15	0,9999	1,0011	-0,2014	-0,0793	-0,4005	-0,1976
16	0,9998	0,9984	-0,2006	-0,0805	-0,4000	-0,2019
17	0,9990	0,9983	-0,1990	-0,0806	-0,3961	-0,1994
18	1,0015	0,9952	-0,5473	-0,2201	-0,3321	-0,5439
19	1,0038	1,0046	-0,5483	-0,2200	-1,0989	-0,5480
20	1,0001	1,0100	-0,5500	-0,2209	-1,0969	-0,5483
21	0,9922	0,9995	-0,5486	-0,2204	-1,1006	-0,5524
22	1,0018	1,0009	-0,2486	-0,1007	-0,5001	-0,2502
23	0,9988	0,9928	-0,2515	-0,0994	-0,5037	-0,2491
24	1,0036	0,9942	-0,2510	-0,1003	-0,4996	-0,2495

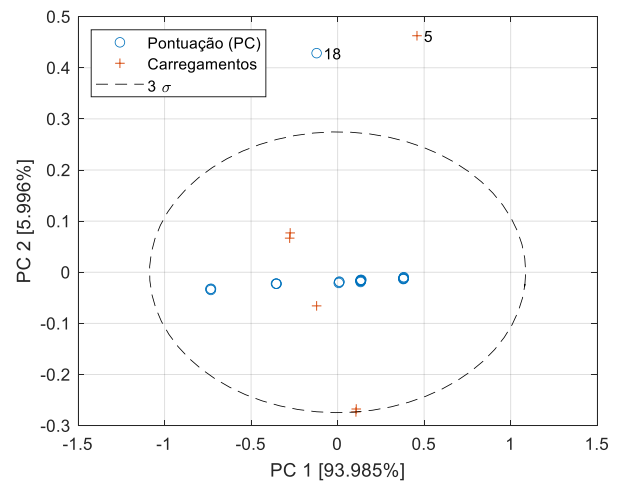


Fig. 4 PCs e carregamentos da rede de 3 barras.

Ainda em relação à Fig. 4, estão presentes os carregamentos relativos às grandezas medidas. No caso testado, foi incluída uma perda na medição 5 das 18 h. Observa-se que a amostra das 18 h é fortemente influenciada pela medida 5 (estão bem próximas). Além disso, o carregamento da medição 5 é o maior na direção do segundo PC. De fato, o carregamento 5 está levando a amostra 18 a um valor muito grande na direção do segundo PC. Por meio desta análise é possível identificar a medida em que está a perda.

2.3 Detecção de perdas no horário de ponta

O resultado apresentado na seção anterior funciona muito bem quando há apenas uma medida com erro. A PC com mais de três desvios padrões é suspeita de conter a perda. Entretanto, para detectar um consumidor com perda não técnica intermitente, o método deve ser capaz de detectar um conjunto de amostras de medidas com erro. A dificuldade está no fato de que, quando há diversas medidas com erro, elas não se distanciam da origem como aparece na Fig. 4. Contudo, ainda que as pontuações das amostras de medidas com a perda não técnica não ultrapassem o limiar de três desvios padrões, elas se concentram, permitindo a detecção. A ideia proposta é detectar o centroide que se destaca.

A Fig. 5 apresenta o fluxograma do método de detecção e identificação de clientes suspeitos. Inicialmente, a matriz com o conjunto de medição é formada. Então a PCA é realizada, calculando-se os PCs e os carregamentos. É inicializado um contador com o número de PCs a serem analisadas. Nos testes realizados foram analisadas três PCs. Para cada PC, os centroides são calculados usando a técnica de agrupamentos *k-means*, na qual cada ponto pertence ao grupo mais próximo da média, ou seja, é o centro da distância euclidiana dos pontos. São calculados dois centroides para cada PC, relativos às amostras com e sem perda. Caso seja detectado que algum centroide excede o limiar pré-estabelecido, em qualquer dos PCs, a medida relativa ao maior carregamento é tomada como suspeita.

A Fig. 6 apresenta um teste realizado com dez medições por hora na rede da Fig. 7, cujos dados de potência estão na Tabela 6. Para cada PC, na vertical de 1 a 3, são apresentadas as pontuações em círculos azuis, sendo seus dois centroides destacados com o símbolo de uma estrela. Os carregamentos são representados pelo símbolo (+) em laranja. Neste exemplo, a detecção foi realizada pelo segundo PC. O centroide mais distante da origem, à direita nesse caso, tem seu valor aproximadamente igual a 1,9. É por meio dele que a detecção é realizada. Quando o centroide de um dos três PCs tem sua distância em relação à origem maior do que um limiar pré-definido, ele é considerado suspeito. As linhas verticais a $\pm 1,62$ representam o limiar adotado. Portanto, os conjuntos de amostras de medição associados ao centroide da direita do segundo PC são suspeitos de conter a perda não técnica. A medida causadora da perda está destacada em um quadrado sobre o carregamento. É o maior carregamento padronizado que, nesse caso, é aproximadamente 3,6.

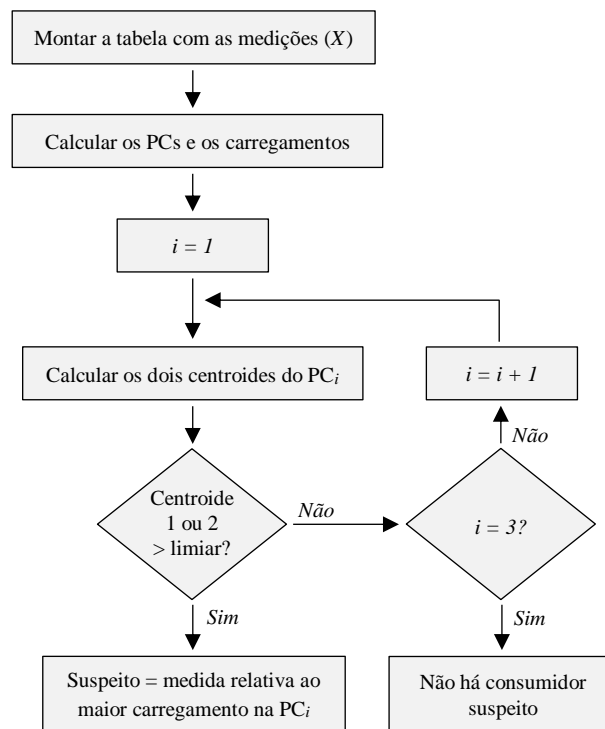


Fig. 5 Fluxograma do método de detecção de perdas.

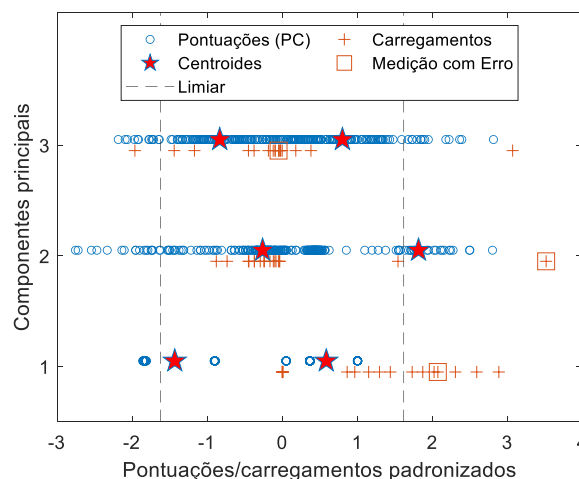


Fig. 6 Detecção e identificação da perda não técnica.

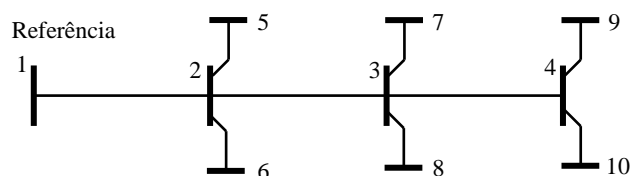


Fig. 7 Esquema de uma rede de distribuição de 10 barras.

Tabela 6. Potências nominais nas barras de consumo [pu] (base: 1 kVA)

Tipo	5	6	7	8	9	10
P	0,50	0,80	0,75	0,60	1,00	0,90
Q	0,30	0,45	0,35	0,40	0,70	0,65

2.4 Definição do limiar de detecção

Na seção anterior, o exemplo apresentado indicava um limiar igual a 1,62 para detecção de perda não técnica usando o maior centroide. Ele foi definido considerando-se que os centroides têm distribuição normal com média $\sqrt{2/\pi}$ e variância igual a 0,5. Assim, é muito provável que os dois centroides estejam a $\pm\sqrt{2/\pi} \cong 0,8$ da origem. O terceiro PC da Fig. 6 é um bom exemplo das posições em que os centroides são esperados.

Para se chegar ao valor esperado da posição dos centroides, considerou-se que a distribuição das pontuações dos PCs é normal e que, como há dois centroides, cada um tem metade dessa distribuição. A Fig. 8 ilustra a ideia. Nessa figura, a curva central, normal com média igual a zero e variância unitária, representa a distribuição das pontuações dos PCs. As duas curvas a direita e a esquerda indicam densidade de probabilidade das posições dos dois centroides. O limiar de detecção de 1,62 foi calculado considerando o intervalo de confiança igual a 95 %, ou seja, caso o centroide esteja sob as áreas destacadas da Fig. 8, é muito provável que haja medidas com valores atípicos, devidos a perdas não técnicas.

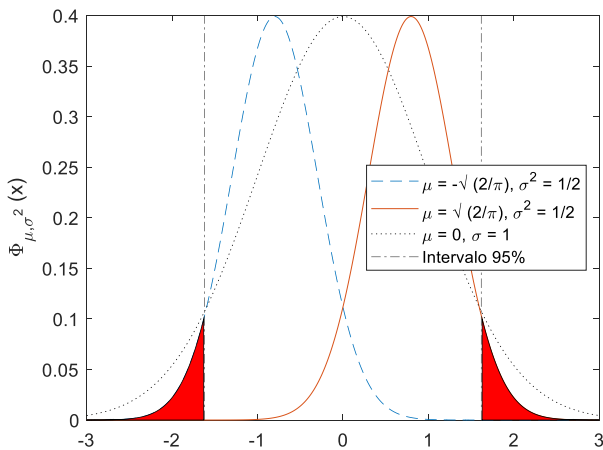


Fig. 8 Distribuições das probabilidades dos centroides.

4. RESULTADOS

O método proposto foi testado na rede de baixa tensão apresentada na Fig. 6. Os ramos do tronco principal (1-2, 2-3 e 3-4) têm a impedância do tipo 1 da Tabela 1 e os demais são do tipo 2. As potências nominais nas barras dos consumidores estão apresentadas na Tabela 6. Assim como nos testes da seção anterior, a análise foi feita nos três PCs. Foram consideradas como disponíveis as medidas de potência ativa e reativa nos consumidores (barras 5 a 10) e também suas, magnitudes de tensão. A barra de referência representa o secundário de um transformador de distribuição.

Neste teste foi considerado que o consumidor na barra 7 tem perda não técnica no horário de ponta, entre 18 h e 21 h. Nesse período, o medidor reporta apenas 30% da demanda real. Para chegar aos valores percentuais de detecção e identificação foram realizadas mil simulações para cada quantidade de medições por hora. Os resultados estão apresentados Fig. 9. As barras mais longas, representam a detecção percentual correta,

ou seja, considerando que há fraude no medidor no horário de ponta, o método deve detectá-la. Nota-se que, com poucas medidas por hora, o método tem dificuldade de detectar a perda. Contudo, a partir de três medições por hora o método consegue detectar cerca de 90% das perdas. As barras mais curtas, apresentam os percentuais de detecção indevida. Nesse caso, a rede foi simulada sem a perda e o método não deveria, portanto, detectá-la. Com poucas medições por hora, o método não tem bom desempenho. Entretanto, a partir de 5 medições por hora a detecção indevida é praticamente eliminada. Por fim, a linha tracejada com marcadores redondos apresenta o percentual de correta identificação. Este índice é calculado a partir das detecções corretas. Após detectar que há perda não técnica em conjuntos de medição, o método deve identificar a qual barra se refere essa perda. A partir de 2 medições por hora o método identifica corretamente mais de 90% dos casos detectados.

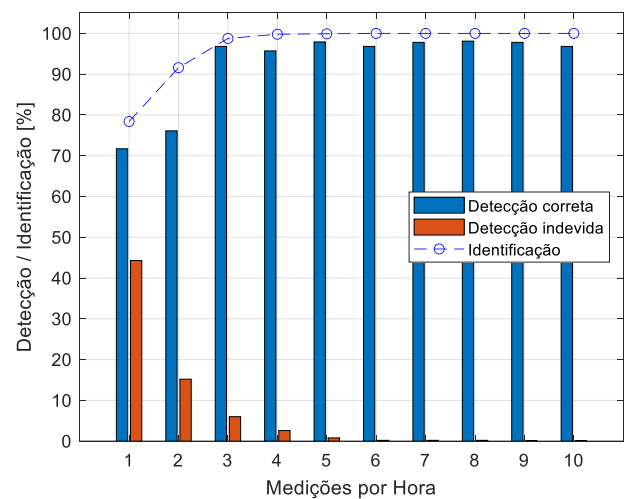


Fig. 9 Detecção em função da quantidade de medidas por hora.

Outro teste foi realizado com o intuito de verificar a sensibilidade do método ao tamanho da perda. Os resultados estão na Fig. 10. Verifica-se na figura que, quando a perda é muito pequena (10% da potência da carga), o percentual de detecção cai significativamente. Contudo, a partir de 20% de perda não técnica o percentual de detecção se mantém estável na ordem de 97%. Neste teste foram consideradas 10 medidas por hora.

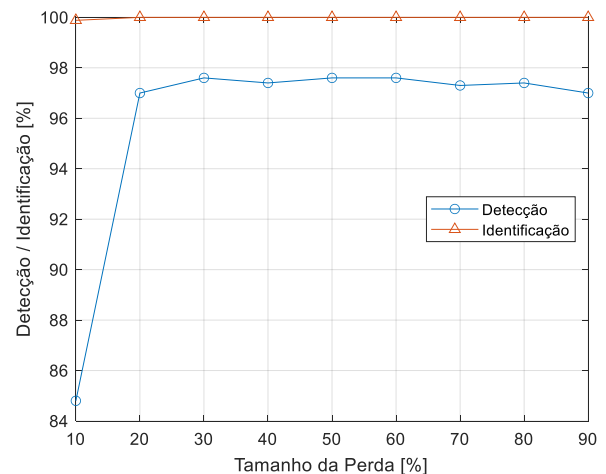


Fig. 10 Detecção em função do tamanho da perda.

5. DISCUSSÃO

O método apresentado tem como pontos fortes a fácil implementação e a não dependência de dados históricos. E, apesar de ainda ser um estudo inicial, o uso da PCA no tratamento das medidas para a detecção de perdas não técnicas se mostra promissor.

Os testes foram realizados numa rede sintética com parâmetros realistas. Entretanto, o método ainda carece de validação em uma rede real de grande porte. Além disso, está sendo avaliada a condição de o consumidor ter perdas não técnicas em diversas horas do dia e de haver mais de um consumidor com perda não técnica simultaneamente.

Em relação às medições usadas, são necessárias medidas de potência (ativa e reativa) e de magnitude de tensão em todos os consumidores; não foram consideradas medidas ao longo das linhas. Apesar de normalmente estarem disponíveis somente medidas de potência neste tipo de análise, verificou-se experimentalmente que a inclusão das medidas de magnitude de tensão melhora o desempenho do método. Outro ponto importante a destacar é que pequenos erros de medição, inerentes à precisão dos medidores, não afeta de forma significativa o método. Todavia, seu desempenho cai quando há grandes variações no perfil de consumo.

6. CONCLUSÕES

Este trabalho apresenta um princípio de detecção de perdas não técnicas em redes de distribuição usando a PCA. Essa técnica permite concentrar as variâncias em poucas PCs, diminuindo a dimensão do problema. O agrupamento dos consumidores suspeitos é realizado pelo método *k-means*. Apesar de a maioria das redes de distribuição ainda não terem sistemas de monitoramento avançado, existe a perspectiva do aumento da quantidade de dispositivos de medição e uma aplicação dessa infraestrutura é a detecção de perdas não técnicas.

O método se mostrou eficiente na detecção e identificação de perdas não técnicas na rede de testes quando há uma taxa de amostragem de ao menos três medidas por hora, admitindo que haja fraude no medidor de um único consumidor durante parte do dia.

REFERÊNCIAS

- ANEEL, 2019. Perdas de Energia Elétrica na Distribuição, Agência Nacional de Energia Elétrica.
- Azevedo, A. L., Dias, D. & Alves, J., 2018. Light faz operação contra 'gatos de luz' em condomínios de classe média. [Online] Available at: <https://g1.globo.com/rj/rio-de-janeiro/noticia/2018/08/30/light-faz-operacao-contragatos-de-luz-em-condominios-de-classe-media.ghtml> [Acesso em 17 06 2020].
- Bretas, A. S., Rossoni, A., Trevizan, R. D. & Bretas, N. G., 2020. Distribution networks nontechnical power loss estimation: A hybrid datadriven physics model-based framework. *Electric Power Systems Research*, 05, Issue 186, p. 10.
- Brunton, S. L. & Kutz, J. N., 2019. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. 1 ed. Cambridge University Press.
- Ferreira, T. S. D., Trindade, F. C. L. & Vieira, J. C. M., 2020. Load flow-based method for nontechnical electrical loss detection and location in distribution systems using smart meters. *IEEE - Transactions on Power Systems*, 03.p. 10.
- Ghori, K. M. et al., 2020. Performance Analysis of Different Types of Machine Learning Classifiers for Non-Technical Loss Detection. *IEEE - Special section on data mining for internet of things*, Volume 8, pp. 16033-16048.
- Huang, S.-C., Lo, Y.-L. & Lu, C.-N., 2013. Non-technical loss detection using state estimation and analysis of variance. *IEEE Transactions on power systems*, 08, 8(3), pp. 2959-2966.
- Jindal, A. et al., 2020. Tackling energy theft in smart grids through data-driven analysis. *Big Island, IEEE*, pp. 410-414.
- Jolliffe, I. T., 2002. *Principal component analysis*. 2 ed. Nova York: Springer-Verlag.
- Kadurek, P., Blom, J., Cobben, J. F. G. & Kling, W. L., 2010. Theft detection and smart metering practices and expectations in the Netherlands. *Gotemburgo, Suécia, IEEE*, p. 6.
- Konstantinos, B. & Georgios, S., 2019. Efficient power theft detection for residential consumers using mean shift data mining knowledge discovery process. *International Journal of Artificial Intelligence and Applications*, Janeiro, pp. 69-85, Vol. 10, No. 1.
- Messinis, G. M. & Hatzigargyriou, N. D., 2018. Review of non-technical loss detection methods. *Electric Power Systems Research*, Volume 158, pp. 250-266.
- Singh, S. K., Bose, R. & Joshi, A., 2019. Energy theft detection for AMI using principal component analysis based reconstructed data. *IET Cyber-Physical Systems: Theory & Applications*, 02, 4(2), pp. 179-185.
- Souza, M. A. d. et al., 2020. Detection and identification of energy theft in advanced metering infrastructures. *Electric Power Systems Research*, 02, Issue 182, p. 12.
- Weckx, S. et al., 2012. Parameter Identification of Unknown Radial Grids for Theft Detection. *Belin, IEEE*, p. 6.
- Zhang, Y. et al., 2020. Energy theft detection in an edge data center using threshold-based abnormality detector. *Electrical Power and Energy Systems*, 10, Issue 121, p. 13.