

ESTIMAÇÃO DA DISTRIBUIÇÃO ESPACIAL DAS PERDAS NÃO TÉCNICAS VIA REGRESSÃO PONDERADA GEOGRAFICAMENTE EM SISTEMAS DE DISTRIBUIÇÃO DE ENERGIA ELÉTRICA

GUSTAVO ESTEVO FELIX, NICOLE COSTA DA SILVA, LUCAS TELES FARIA

UNESP – Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus Experimental de Rosana
Av. dos Barrageiros, 1881, 19274-000, Rosana, São Paulo, Brasil

E-mails: gustavo.felix@unesp.br, nicole.costa@unesp.br, lucas.teles@unesp.br

Abstract— Non-technical losses are commonly caused by various types of fraud, clandestine connections, self-links, and faulty meters, for example. They burden power distributors financially and generate various negative impacts on society. Numerous studies have been carried out to combat and prevent these losses with the main focus on detecting irregular consumer units (UCs) without considering their spatial location. However, the study of the characteristics of the place where they occur may bring essential information for a better understanding of the problem. In this context, the geographically weighted regression (GWR) model is adopted to estimate the distribution of non-technical losses (NTL) in the geographical space of the city. After the process of exploratory analysis and the understanding of the spatial patterns of the NTL disposition, it is found that the socioeconomic and technical attributes of the subareas of the city, such as average income, rented consumer unit and load density, can be used as input for a spatial regression model to estimate NTL. The results obtained in this study indicate that inspections are being carried out in subareas with fewer irregular UCs. From the analysis of the subareas identified from the GWR spatial regression, it can be seen that the eastern region of the city is more prone to NTL. These subareas should have more consuming units inspected by the energy distribution inspection teams.

Keywords— Commercial Losses, Geographically Weighted Regression, Non-Technical Losses, Power Distribution Systems, Spatial Data Analysis.

Resumo— As perdas não técnicas originam-se comumente devido aos diversos tipos de fraudes, ligações clandestinas, autorreligiões e medidores avariados, por exemplo. Elas oneram financeiramente as distribuidoras de energia elétrica e geram diversos impactos negativos à sociedade. Inúmeros estudos têm sido realizados para combate e prevenção dessas perdas com foco principal na detecção de unidades consumidoras irregulares sem considerar a localização espacial das mesmas. No entanto, o estudo das características do lugar onde elas ocorrem pode trazer informações para melhor compreensão do problema. Nesse contexto, adota-se neste trabalho, um modelo de regressão espacial ponderada geograficamente ou *geographically weighted regression* (GWR) para estimar a distribuição das perdas não técnicas no espaço da cidade. Os resultados obtidos neste trabalho apontam que as inspeções estão sendo realizadas em subáreas com menos unidades consumidoras irregulares. A partir da análise das subáreas identificadas a partir da regressão espacial GWR constata-se que a região leste da cidade apresenta maior propensão às perdas não técnicas. As subáreas da região leste deveriam ter mais unidade consumidoras inspecionadas pelas equipes de inspeção da distribuidora de energia.

Palavras-chave: Análise Espacial de Dados Geográficos, Perdas Não Técnicas, Perdas Comerciais, Regressão Espacial Ponderada Geograficamente, Sistemas de Distribuição de Energia Elétrica.

1 Introdução

As perdas de energia no sistema de distribuição de energia elétrica (SDEE) representam prejuízo financeiro elevado para muitas distribuidoras brasileiras, principalmente para aquelas localizadas nas regiões norte e nordeste; juntas, representam um montante de perdas sobre energia injetada de 48,6% (ANEEL, 2019).

As perdas de energia ou perdas globais são definidas como sendo a diferença entre a energia comprada e distribuída e a energia efetivamente paga pelos consumidores (ANEEL, 2019). Essas perdas são classificadas em duas categorias: perdas técnicas e perdas não técnicas (PNTs) também conhecidas como perdas comerciais.

As PNTs no SDEE correspondem a uma parcela das perdas globais após serem descontadas as perdas técnicas. As perdas técnicas são inerentes ao transporte de energia elétrica, como as perdas por efeito Joule, correntes de Foucault e perdas por histerese, por

exemplo. Por outro lado, as PNTs são ocasionadas predominantemente por ações ilícitas de consumidores finais como: furtos de energia por adulterações no medidor, desvios no sistema de medição ou *bypass*, autorreligiões, ligações clandestinas e outros (MPF, 2019).

As PNTs no sistema elétrico brasileiro ocorrem principalmente por questões sociais, comportamentais, educativas e socioeconômicas (Penin, 2008).

Um estudo elaborado pela ANEEL (2019) contabilizou que no ano de 2018, o prejuízo das distribuidoras de energia elétrica devido às PNTs superou o montante de 3 bilhões de reais no Brasil. Segundo GESEL (2019) e Smith (2004), esses indicadores demonstram a deterioração econômica e a infraestrutura insuficiente nos âmbitos da educação, renda, emprego, segurança pública e desigualdade social.

Na Fig.1, é apresentado um panorama das PNTs em 32 distribuidoras brasileiras no ano de 2019, referente ao ano de 2018. Observa-se que as distribuidoras localizadas nas regiões norte e nordeste são aquelas que possuem os maiores percentuais de PNTs.

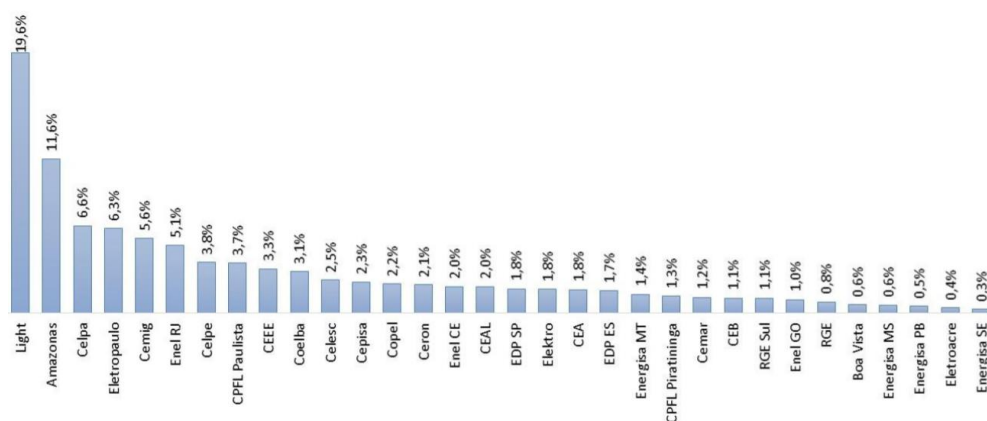


Figura 1. Detalhamento das PNTs em 32 distribuidoras de energia elétrica brasileiras (ANEEL, 2019).

As distribuidoras Amazonas Energia e a CELPA (Centrais Elétricas do Pará) possuem PNTs de 11,6% e 6,6%, respectivamente. A Light é uma exceção, isto é, uma distribuidora com altas PNTs (19,6%) e que está localizada fora das regiões norte e nordeste – localiza-se na região sudeste no Estado do Rio de Janeiro. Nesse caso específico, devido ao alto nível de corrupção e ao relevo acidentado da cidade do Rio de Janeiro com muitos morros, favorece o furto de eletricidade; em muitos casos, até mesmo o poder público tem acesso limitado em busca de ilícitos.

A partir desse cenário supracitado, torna-se importante a obtenção de formas de maximizar a energia disponível para a comercialização. Com esse fim, este estudo emprega técnicas de análise espacial de dados geográficos para localização de áreas vulneráveis às PNTs.

As pesquisas em PNTs, em sua maioria, têm enfoque principal na detecção pontual das Unidades Consumidoras (UCs) irregulares (Angelos *et al.*, 2011; Cabral *et al.*, 2004; Faria, 2012; Jiang, 2002; Nagi *et al.*, 2010; Nagi *et al.*, 2011; Nizar, Dong e Wang, 2008; Nizar, Dong e Zhang, 2008; Ramos *et al.*, 2011; Ramos *et al.*, 2012; Ramos, 2014; Trevizan *et al.*, 2015; Buzau *et al.*, 2019; Monteiro e Maciel, 2018; Glauner *et al.*, 2016; Sengupta *et al.*, 2017; Viegas *et al.*, 2017).

Os trabalhos supracitados empregam técnicas de sistemas inteligentes como: redes neurais, lógica fuzzy, algoritmos de aprendizado de máquina, mineração de dados e de texto e outras. Esses estudos detectam as PNTs pontualmente, ou seja, indicam onde estão as UCs irregulares. No entanto, devido à escassez de informações disponíveis por UCs pertencentes às redes convencionais de distribuição, implica em resultados imprecisos em muitos casos. Acresce que o estudo das características do lugar onde ocorrem as PNTs é uma estratégia complementar às técnicas de detecção pontual de UCs irregulares e que pode trazer informações relevantes para melhor compreensão do problema.

Em Felix, Silva e Faria (2020), a incorporação do estudo do espaço geográfico ao problema das PNTs foi feita através da aplicação de técnicas de exploração de análise espacial de dados geográficos. Para o estudo das características do lugar onde ocorreram as

PNTs foram avaliados indicadores socioeconômicos e técnicos das subáreas da cidade tais como: renda média, número de UCs alugadas e densidade de carga. Esses atributos auxiliaram na explicação da distribuição espacial das PNTs. Nesse estudo constatou-se que as PNTs estão associadas, em sua maioria, a subáreas que possuem uma maior quantidade de UCs alugadas, menor renda média e menor densidade de carga.

Após o processo de análise exploratória realizado em Felix, Silva e Faria (2020), convém a execução de modelos espaciais mais sofisticados para a compreensão da relação entre as variáveis relacionadas às PNTs. Nesse contexto, este trabalho emprega um modelo de regressão espacial ponderada geograficamente ou *Geographically Weighted Regression* (GWR) para estimar a distribuição espacial das PNTs.

Destaca-se que este trabalho é um avanço em relação ao trabalho de Felix, Silva e Faria (2020) cujos resultados dão suporte para explicar a disposição das PNTs no espaço geográfico do município em estudo.

Este trabalho segue na Seção 2 com o estudo da distribuição espacial das PNTs. Na Seção 2.1 abordam-se aspectos relevantes da análise espacial de dados geográficos. Na Seção 2.2 tem-se tópicos relevantes da análise exploratória nos dados espaciais. Na Seção 2.3 aborda-se a regressão espacial ponderada geograficamente. Na seção 3 são apresentados os resultados com o emprego da regressão espacial GWR. Na Seção 3.1 apresenta-se uma breve descrição da base de dados utilizada neste estudo. A Seção 3.2 contém a análise exploratória para a região em estudo. Na Seção 3.3 aplica-se a regressão espacial GWR. Por fim, na Seção 4, é apresentada a conclusão deste trabalho.

2 Distribuição Espacial das Perdas Não Técnicas

Segundo Messinis e Hatzigryriou (2018), há uma necessidade do desenvolvimento de novos métodos para a detecção das PNTs. A maioria dos trabalhos avaliados em PNTs são feitos com técnicas de *soft computing* como: redes neurais, lógica fuzzy, aprendizado de máquina e outras. No entanto, a incorporação das técnicas de análise espacial de dados geográficos contribuem para a identificação das regiões mais vulneráveis às PNTs, bem como auxilia na explicação do por que as PNTs estão mais concentradas em

determinadas localidades da cidade (Faria, Melo e Padilha-Feltrin, 2016; Felix, Silva e Faria, 2020).

2.1 Análise Espacial de Dados Geográficos

A análise espacial permite incorporar o espaço e levantar as características do fenômeno de estudo. É composta por um conjunto de ferramentas para explorar e modelar processos que se expressam através de uma distribuição no espaço, por exemplo – índice de criminalidade, mortalidade infantil, óbitos por doenças transmissíveis e outros.

Na análise espacial mensuram-se as propriedades e relacionamentos, de modo a considerar a localização espacial do fenômeno em estudo de maneira explícita (Druck *et al.*, 2004).

Os problemas em análise espacial de dados geográficos são classificados em três categorias: (1) eventos ou padrões pontuais (ocorrências de furtos, doenças, por exemplo); (2) áreas com contagens agregadas (levantamento populacionais, como censos demográficos e estatísticas de saúde, por exemplo) e (3) superfícies contínuas – levantamento de recursos naturais, por exemplo.

A análise espacial por áreas empregada neste estudo é uma subdivisão da análise espacial de dados geográficos. Essa técnica é útil quando se lidam com eventos agregados por município, bairros ou setores censitários (Druck *et al.*, 2004).

2.2 Análise Exploratória em Dados Espaciais

A análise exploratória em dados espaciais é a primeira etapa no processo de análise espacial de dados geográficos.

Neste estudo, as PNTs são representadas pela *Taxa de Perdas (TXPER)*, que é obtida pela razão entre o número de UCs irregulares e o número de UCs inspecionadas por setor censitário ou por subáreas.

A estrutura de vizinhança entre as subáreas e a matriz de ponderação $W_{(s)}$ são obtidas através da função de *kernel* bi-quadrado em (1).

O ajuste da ponderação local w_{ij} é feito de modo que as subáreas mais próximas exerçam maior influência. Em (1), d_{ij} é a distância Euclidiana entre os centroides das subáreas i e j e b é o parâmetro largura de banda ou *bandwidth* que regula o nível de alisamento da função de *kernel*.

$$w_{ij} = \begin{cases} 1 - \left(\frac{d_{ij}}{b}\right)^4, & \text{se } |d_{ij}| < b \\ 0, & \text{caso contrário} \end{cases} \quad (1)$$

2.3 Regressão Espacial Ponderada Geograficamente

Os modelos de regressão utilizam o relacionamento que há entre duas ou mais variáveis de modo que uma delas possa ser descrita e seu valor possa ser estimado a partir das demais (Faria, Melo e Padilha-Feltrin, 2016).

Os modelos de regressões espaciais locais são os mais adequados para representar a heterogeneidade espacial presentes em dados advindos de censos demográficos, por exemplo. Portanto, modelos de regressões locais são mais adequados em comparação aos modelos de regressões globais (Gollini *et al.*, 2015; Lu *et al.*, 2014).

A regressão espacial ponderada geograficamente ou *Geographically Weighted Regression (GWR)* é um modelo de regressão pertencente à análise espacial por áreas e integra a família dos modelos ponderados geograficamente ou *Geographically Weighted Models (GWMs)*. É uma adaptação do modelo de regressão local por áreas. Esses modelos permitem a incorporação de variáveis que caracterizam as diversas subáreas da cidade.

Neste contexto, a *Taxa de Perdas (TXPER)* por subáreas constitui a variável dependente ou variável de estudo da regressão espacial GWR.

A regressão espacial GWR “baseia-se no ajuste de um modelo de regressão a cada ponto observado, ponderando as demais observações como função da distância a este ponto” (Faria, Melo e Padilha-Feltrin, 2016). Em suma, com os dados geograficamente ponderados, as observações sofrem influência das subáreas vizinhas a ela, resultando em uma melhor estimação do conjunto local dos coeficientes da regressão (Gollini *et al.*, 2015).

Em (2) apresenta-se a regressão GWR, onde mensura-se o relacionamento que há entre duas ou mais variáveis de modo que uma delas (variável y_i) possa ser descrita e seu valor estimado a partir das demais variáveis independentes ou explicativas x_{ik} .

$$y_i = \beta_{i0} + \sum_{k=1}^m \beta_{ik} x_{ik} + \epsilon_i \quad (2)$$

Em (2) y_i é a variável dependente na região i ; x_{ik} são os valores da k -ésimas variáveis independentes na região i ; m é o número de variáveis independentes avaliadas; β_{i0} é o coeficiente de regressão local para a k -ésima variável independente na região i e ϵ_i é o resíduo.

Ressalta-se que o modelo mensura as relações inerentes em torno de cada ponto i , onde os coeficientes são estimados via método dos mínimos quadrados ponderados em (3). Observa-se que em (3), $W_{(s)}$ é a matriz de ponderação geográfica para cada dado observado na coordenada geográfica s e que foi obtida via (1).

$$\beta_{(s)} = (X^T W_{(s)} X)^{-1} X^T W_{(s)} Y \quad (3)$$

3 Resultados e Discussões

As simulações apresentadas nesta seção foram realizadas no *software R* versão 3.6.1 (R Core Team, 2015). O *R* é um *software* livre para a computação estatística. Nele estão implementadas uma série de

bibliotecas conhecidas para análise numérica, assim como o pacote *GWmodel*, empregado neste estudo.

3.1 Descrição da Base de Dados

As características ou variáveis espaciais envolvem atributos socioeconômicos, dados técnicos da rede de distribuição de energia elétrica e dados de inspeções em campo advindos do censo demográfico (IBGE, 2010).

Em Smith (2004), relacionam-se as PNTs com as características socioeconômicas do lugar onde elas ocorrem. Adicionalmente, aproveitam-se as informações públicas disponíveis a partir do censo demográfico (IBGE, 2010). Os dados disponibilizados pelo IBGE estão agregados por áreas; assim o aproveitamento das informações do censo demográfico contribui para o estudo das UCs irregulares.

Neste trabalho, as PNTs são representadas pela *Taxa de Perdas (TXPER)* que é a variável dependente da regressão GWR cujas variáveis independentes ou explicativas são: *Renda Média (REN)*, *Densidade de Carga (DENC)* e *UCs alugadas (ALUG)*.

Em Smith (2004) relaciona-se as PNTs com a vulnerabilidade socioeconômica. Esse fato será representado pela variável *REN*. A variável *DENC* está diretamente relacionada às PNTs, pois há maior possibilidade de encontrar UCs irregulares em áreas com maior número de UCs; portanto, com maior densidade de carga. Em último, relaciona-se a variável *ALUG* com as PNTs, pois indivíduos mal-intencionados que residem em casas alugadas podem implementar fraudes no medidor e serem encobertos por não serem proprietários das mesmas.

A Tabela 1 apresenta as descrições dos dados de entrada para o modelo de regressão espacial GWR.

Tabela 1. Descrição das variáveis analisadas por subáreas (IBGE, 2010).

Variáveis	Siglas	Descrições
<i>Renda Média</i>	<i>REN</i>	Rendimento médio mensal (com e sem renda em reais)
<i>Densidade de Carga</i>	<i>DENC</i>	Densidade de carga por setor censitário em KVA/km ²
<i>UCs Alugadas (%)</i>	<i>ALUG</i>	Percentual de domicílios particulares alugados
<i>Taxa de Perdas</i>	<i>TXPER</i>	Razão entre o número de UCs irregulares e o número de UCs inspecionadas por setor censitário

É importante observar que a utilização de *TXPER* como variável dependente da regressão GWR é justificável, pois são consideradas as parcelas das UCs irregulares e UCs inspecionadas. Ao utilizar o valor bruto de UCs irregulares são contabilizadas apenas subáreas com UCs inspecionadas. O modelo GWR pode indicar subáreas em oculto, ou seja, aquelas com

poucas ou nenhuma UCs inspecionadas, podendo exibir maior propensão às PNTs.

3.2 Análise Exploratória em Perdas Não Técnicas

A Figura 2 apresenta a distribuição espacial das PNTs no município em estudo.



Fig. 2 Distribuição percentual da *Taxa de Perdas (TXPER)* por subáreas (Felix, Silva e Faria, 2020).

Observa-se a partir da Fig. 2 que a distribuição espacial da *TXPER* apresenta uma distribuição mediana crescente na região central. Embora seja importante a avaliação da distribuição espacial da variável de estudo, a aplicação do resumo estatístico ponderado geograficamente e de modelos ponderados geograficamente (GWMs), colaboram para a explicação dos fatores associados a distribuição espacial das PNTs.

Em Felix, Silva e Faria (2020) avaliou-se a correlação das PNTs com a disposição espacial das variáveis explicativas: *Renda Média (REN)*, *Densidade de Carga (DENC)* e *UCs alugadas (ALUG)*. A partir da análise da Tabela 2, observa-se que a correlação local ponderada geograficamente ou *Geographically Weighted (GW)* está disposta de forma heterogênea na cidade. Algumas subáreas têm correlações moderadamente positivas; outras, moderadamente negativas.

Cada subárea da cidade está associada a um valor do coeficiente de correlação local. A Tabela 2 identifica subáreas mais propensas às PNTs via análise exploratória local ponderada geograficamente.

Tabela 2. Coeficientes de correlação local GW contabilizados por subáreas (Felix, Silva e Faria, 2020).

Intervalos de Coeficientes Locais de Correlação GW	<i>TXPER e REN</i>	<i>TXPER e DENC</i>	<i>TXPER e ALUG</i>
0,0 a 0,3	15,6%	25,2%	57,5%
0,3 a 0,6	2,3%	5,6%	9,3%
0,6 a 1,0	0,0%	0,4%	0,4%
-0,3 a 0,0	64,2%	67,8%	29,5%
-0,6 a -0,3	17,9%	1,0%	3,3%
-1,0 a -0,6	0,0%	0,0%	0,0%

A maioria das subáreas apresenta correlação local GW fracamente negativa (-0,3 a 0,0) entre *TXPER* –

REN (64,2%) e *TXPER-DENC* (67,8%) e fracamente positiva (0,0 a 0,3) entre *TXPER-ALUG* (57,5%).

Os coeficientes de correção foram classificados em três níveis em valor absoluto: 0,0 – 0,3 (correlação fraca); 0,3 – 0,6 (correlação moderada) e 0,6 – 1,0 (correlação forte). Em Felix, Silva, Faria (2020) identificou-se através da análise exploratória local GW que *TXPER* e *REN* foi a única correlação que apresentou quantidade relevantes de subáreas no nível *moderado* (17,9%). Portanto, atenção especial deve ser dada às PNTs em subáreas vulneráveis socioeconomicamente.

3.3 Aplicação da Regressão Espacial Ponderada Geograficamente em um Município do Interior Paulista

Nas Figuras 3-5 são apresentados mapas temáticos com os coeficientes estimados para as variáveis independentes do GWR para cada subárea. Observa-se que, caso não existisse a variação dos coeficientes, não se tornaria necessário à aplicação do modelo de regressão local GWR, sendo mais conveniente a aplicação de um modelo de regressão global mais simplificado (Bivand, Pebesma e Gómez-Rúbio, 2008). A Tabela 3 apresenta algumas métricas da estatística descritiva para os coeficientes estimados para as variáveis independentes aplicadas na regressão espacial GWR.

Tabela 3. Métricas da estatística descritiva aplicadas aos coeficientes estimados das variáveis independentes da regressão espacial GWR.

Variáveis	Min.	Max.	Max. – Min.
<i>REN</i>	$-3,16 \times 10^{-4}$	$6,00 \times 10^{-4}$	$9,16 \times 10^{-4}$
<i>DENC</i>	$-1,97 \times 10^{-2}$	$4,20 \times 10^{-2}$	$6,17 \times 10^{-2}$
<i>ALUG</i>	-1,40	1,31	2,71

Max.: Valor máximo. Min.: Valor mínimo.

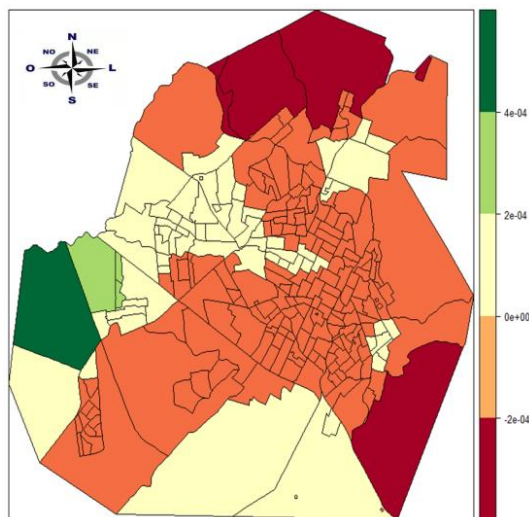


Fig.3 Coeficientes GWR para variável independente Renda Média (*REN*).

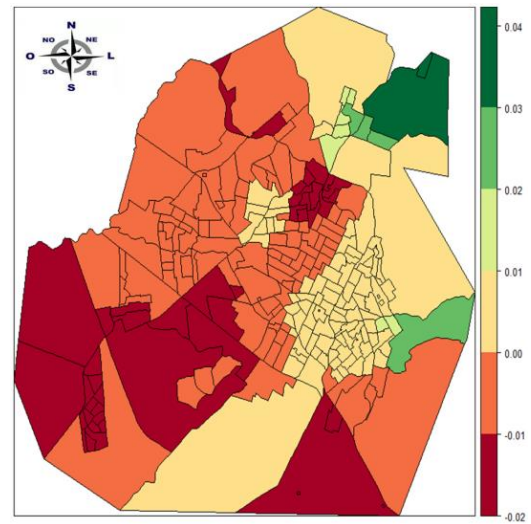


Fig.4 Coeficientes GWR para variável independente Densidade de Carga (*DENC*).

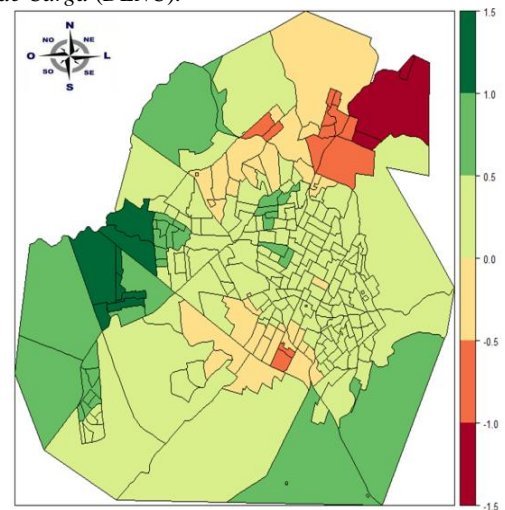


Fig.5 Coeficientes GWR para variável independente %UCS alugadas (*ALUG*).

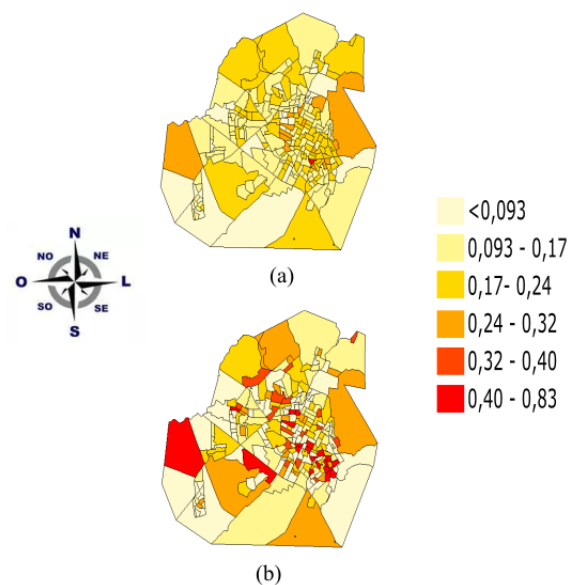


Fig. 6 Comparação da distribuição da Taxa de perdas. (a) Taxa de Perda estimada via regressão espacial GWR. (b) Taxa de Perda real obtida a partir de inspeções em campo.

Tabela 4. Comparação entre as PNTs Estimadas via GWR com as PNTs Reais contabilizadas percentualmente por intervalos de classes por subáreas.

Intervalos	PNTs Estimadas	PNTs Reais	Classes
0,000 a 0,093	11,8%	34,6%	Baixo
0,093 a 0,170	53,5%	23,6%	
0,170 a 0,240	28,0%	18,1%	Médio
0,240 a 0,320	6,3%	10,6%	
0,320 a 0,400	0,0%	6,7%	Alto
0,400 a 0,830	0,4%	6,4%	

Na Fig. 6 (a) apresenta-se o principal resultado deste trabalho. Estimam-se as PNTs via modelo de regressão espacial GWR. São comparadas a *TXPER* estimada (Fig.6 (a)) com a *TXPER* real (Fig.6 (b)) obtida via inspeções em campo. A Tabela 4 compara o valor da *TXPER* estimadas com a *TXPER* real para as subáreas da cidade em intervalos fixos.

Da Tabela 4 depreendem-se a partir da comparação da *TXPER* estimada com a *TXPER* real.

- ❖ Observa-se que 11,8% das subáreas apresentaram uma *TXPER Estimada* entre 0-0,093 e a *TXPER Real* nesse mesmo intervalo é de 34,6%. Isso sugere que as inspeções realizadas no município estão concentradas em subáreas com menos PNTs. Os indicadores das *TXPER Real* próximo de zero refletem essa relação. Em contrapartida, o modelo proposto detectou que o número de subáreas com *TXPER Estimada* próximo de zero é bem menor sendo de 11,8%. Acresce que a maioria das distribuidoras brasileiras não têm uma metodologia padrão para seleção de UCs a serem inspecionadas. Comumente executam a seleção de forma rudimentar baseado em técnicas semiautomáticas aplicadas em planilhas eletrônicas e contando com a experiência de colaboradores no combate às PNTs. Em geral, isso implica em *TXPER* real reduzida em muitas subáreas conforme mostra a Tabela 4.
- ❖ Observa-se que 53,5% das subáreas apresentaram uma *TXPER* estimadas entre 0,093-0,17 e a *TXPER* real nesse mesmo intervalo é de 23,6% das subáreas. Isto evidencia que ocorreram poucas inspeções nessas subáreas e; conforme o modelo proposto, existem UCs irregulares que não foram identificadas.
- ❖ Observa-se que no intervalo 0,4–0,83, a *TXPER* estimada é de 0,4%, enquanto para o mesmo intervalo apresenta um valor de 6,4% das subáreas com *TXPER* real. Infere-se que esse resultado é coerente com a realidade, pois, geralmente as distribuidoras executam a varredura, isto é, a busca por UCs irregulares com grande quantidade de inspeções em áreas pequenas, acarretando uma elevada *TXPER* real em poucas subáreas. Esse fato não é adequadamente capturado pelo

modelo, o qual promove uma elevação suave da *TXPER* estimada.

4 Conclusão

Neste trabalho incorporou-se a análise espacial de dados geográficos por áreas aos problemas das PNTs. Este trabalho foi um avanço de Felix, Silva e Faria (2020) que implementaram a primeira etapa da análise espacial de dados geográficos: a análise exploratória nos dados espaciais. As PNTs foram representadas pela *Taxa de Perdas (TXPER)* cuja distribuição espacial foi estimada através de variáveis técnicas e socioeconômicas que influenciam na distribuição das PNTs no espaço geográfico da cidade em estudo.

A disposição quanto à classificação das perdas PNTs estão segmentadas em três níveis: *baixo* (0,000 a 0,093; 0,093 a 0,170), *médio* (0,170 a 0,240; 0,240 a 0,320) e *alto* (0,320 a 0,400; 0,400 a 0,830). Observa-se da Fig. 6 (a) e (b) que a região central dos mapas apresenta uma distribuição com *TXPER* predominante *média*. Na Fig.6 (a) pode-se verificar que o GWR demonstrou que os valores estimados de perdas (*TXPER* estimada) aumentaram em cerca de 44,1%, evidenciando que ocorreram poucas inspeções nas subáreas em estudo.

No intervalo classificado como *baixo* (0 – 0,093) as PNTs reais (*TXPER* real), advindas das inspeções de campo apresentam percentuais superiores aos encontrados pelo modelo. Isso é uma evidência de que as inspeções estão concentradas em regiões menos problemáticas da cidade em estudo, com menos UCs irregulares. Observa-se que, da análise exploratória, as subáreas problemáticas da cidade estão localizadas na região leste. A partir da correlação local GW 17,9% das subáreas avaliadas apresentaram uma correlação moderadamente negativa para o par de variáveis *TXPER* e *REN*; logo, supõe que essa região apresenta maior propensão as PNTs.

Depreende-se a partir da tabela 4 e da Fig.6 (a) que na região leste da cidade existem UCs irregulares que possivelmente não foram alvo de inspeções; portanto, recomenda-se uma maior atenção das distribuidoras de energia na região leste da cidade em estudo.

Agradecimentos

Os autores agradecem à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pela concessão da bolsa de Iniciação Científica (Processo: 19/04417-0).

Referências Bibliográficas

- Angelos, E.W.S., Saavedra, O.R., Cortés, O.A.C., Souza, A.N. (2011). Detection and Identification of Abnormalities in Customer Consumptions in Power Distribution Systems, *IEEE Trans. on Power Delivery*, v. 26, p. 2436-2442.

- Bivand, R.S., Pebesma, J. Gómez-Rubio, V. (2008). *Applied Spatial Data Analysis with R*. New York: Springer.
- Buzau, M., Tejedor-Aguilera, J., Cruz-Romero, P., Gómez-Expósito, A. (2019). Detection of non-technical losses using smart meter data and supervised learning. *IEEE Transactions on Smart Grid*, v. 10, p. 2661-2670.
- Cabral, J.E., Pinto, J.O., Gontijo, E.M., Filho, J.R. (2004). Fraud detection in electrical energy consumers using rough sets, *Anais de Conferência, IEEE International Conference on Systems, Man and Cybernetics*, 10-13 Out., The Hague, Netherlands.
- Druck, S., Carvalho, M.S., Câmara, G., Monteiro, A.V. (2004). Análise Espacial de Dados Geográficos. Brasília: Embrapa.
- Faria, L.T., Melo, J.D., Padilha-Feltrin, A. (2016). Spatial-Temporal Estimation for Non-Technical Losses. *IEEE Trans. Power Delivery*, v. 9, n. 2, p. 743-799.
- Felix, G.E., Silva, N.C., Faria, L.T. (2020). Análise Espacial Exploratória das Perdas Não Técnicas em Sistemas de Distribuição de Energia Elétrica. *Anais de Conferência, VIII Simpósio Brasileiro de Sistemas Elétricos (VIII SBSE)*, 25-28 Ago., Santo André, São Paulo, Brasil.
- Grupo de Estudos do Setor Elétrico – GESEL (2019). Os desafios da regulação de perdas não técnicas no segmento de distribuição no Brasil. Avaliado em: 5 Abr. 2020. Disponível em: <http://gesel.ie.ufrj.br/app/webroot/files/publications/48_castro233.pdf>.
- Glauner, P., Meira, J. A., Dolberg, L., State, R., Bettinger, F., Rangon, Y. (2016). Neighborhood features help detecting non-technical losses in big data sets. *Anais de Conferência: IEEE/ACM 3rd International Conference on Big Data Computing Applications and Technologies*.
- Gollini, I.; Binbim, L., Charlton, M.; Brunson, C.; Harris, P. (2015). GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software*, v. 63, p. 1-50.
- Instituto Brasileiro de Geografia e Estatística – IBGE (2010). Censo demográfico 2010: Agregado por setores censitários. 2^a Ed. Brasília, DF. Disponível em: <<http://www.ibge.gov.br>>.
- Jiang, R.J. (2002). Wavelet based feature extraction and multiple classifiers for electricity fraud detection. *IEEE/PES Transmission and Distribution Conference and Exhibition, Anais, Asia Pacific*.
- Lopes, S.B.; Brondino, N.C.M.; Silva, A.N.R. (2006). Análise do desempenho de modelos de regressão espacial na previsão de demanda por transportes. Disponível: <<http://www.redpgv.coppe.ufrj.br/index.php/es/produccion/articulos-cientificos/2006-1/254-lobes-brondino-silva-panam2006/file>>.
- Nagi, J., Yap, K.S., Tiong, S.K., Ahmed, S.K., Mohamad, M. (2010). Nontechnical Loss Detection for Metered Customers in Power Utility Using Support Vector Machines. *IEEE Trans. Power Delivery*, v. 25, p. 1162-1171.
- Nagi, J., Yap, K.S., Tiong, S.K., Ahmed, S.K., Nagi, F. (2011). Improving SVM-Based Nontechnical Loss Detection In Power Utility Using the Fuzzy Inference System. *IEEE Trans. Power Delivery*, v. 26, p. 1284-1285.
- Nizar, A.H., Dong, Z.Y., Wang, Y. (2008). Power Utility Nontechnical Loss Analysis with Extreme Learning Machine Method. *IEEE Trans. Power Systems*, v. 23, p. 946-955.
- Nizar, A.H., Dong, Z.Y., Zhang, P. Nizar, A.H., Dong, Z.Y., Zhang, P. (2008). Detection Rules for Non-Technical Losses Analysis in Power Utilities. *IEEE Power and Energy Society General Meeting*, Pittsburgh, PA, USA.
- Messinis, G.M.; Hatziargyriou, N.D. (2018). Review of non-technical loss detection methods. *Electrical Power Systems Research*, v. 158, p. 250-266.
- Monteiro, M. D. and Maciel, R. S. (2018). Detection of commercial losses in electric power distribution systems using data mining techniques. *Simpósio Brasileiro de Sistemas Elétricos (SBSE)*. Niterói, 12-16 Mai 2018.
- Ministério Público Federal (MPF). 3^a Câmara de Coordenação e Revisão (Org.). Perdas não técnicas de energia elétrica, 2019.
- Penin, C.A.S. (2008). Combate, prevenção e otimização das perdas comerciais de energia elétrica. Tese (Doutorado em Engenharia Elétrica) – Escola Politécnica, Universidade de São Paulo.
- Ramos, C.C.O., Souza, A. N., Falcão, A.X., Papa, J. P. (2012). New insights on nontechnical losses characterization through evolutionary-based feature selection. *IEEE Trans. on Power Delivery*, v. 27, p. 140-146.
- Ramos, C.C.O. (2014). Caracterização das Perdas Comerciais em Sistemas de Energia através de Técnicas Inteligentes. Tese (Doutorado em Ciências) – Escola Politécnica, Universidade de São Paulo.
- Ramos, C.C.O., Sousa, A.N., Papa, J.P., Falcão, A.X. (2011). A New Approach for Nontechnical Losses Detection Based on Optimum-Path Forest. *IEEE Trans. Power Systems*, v. 26, p. 181 -189.
- R Core Team (2015). A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Disponível em: <<http://www.R-project.org/>>.
- Sengupta, R., Heeks, R., Chattapadhyay, S. and Foster, C. (2017). Exploring big data for development: an electricity sector case study from India. *Development Informatics Working Paper*, 66, [online] Available at: <<https://ssrn.com/abstract=3431737>> [Accessed 19 May 2020].
- Smith, T.B. (2004). Electricity theft: a comparative analysis. *Energy Policy*, v. 32, p. 2067-2076.

- Viegas, J. L., Esteves, P. R., Melício, R., Mendes, V. M. F. and Vieira, S. M. (2017). Solutions for detection of non-technical losses in the electricity grid: A review. *Renewable and Sustainable Energy Reviews*, 80, pp.1256-1268.
- Trevizzan, R.D., Rossoni, A., Bretas, A.S., Gazzana, D.S., Martin, R.P., Bretas, N.G., Bettiol, A.L; Carniato, A.; Passos, L. F.N. (2015). Non-technical losses identification using optimum-path forest and state estimation. *PowerTech*, Eindhoven.